Natural Resources for Human Health

eISSN: 2583-1194

Original Research

Received 18 March 2026
Revised 30 April 2026
Accepted 20 May 2026

Natural Resources for Human Health 2026, Vol. 6 (2s): 722–735

KEYWORDS:
Facial Emotion Recognition; Deep Learning; Convolutional Neural Network (CNN); Transfer Learning; VGG16; VGG19; InceptionV3; Affective Healthcare.

A Comparative Evaluation of CNN and Deep Learning Models for Facial Emotion Recognition in Affective Healthcare

Asha M S[1], Anoop G L[2]*, Deepa Y[3], Shruti Japalur[4], Babu Kumar S[5], Santhrupth B C[6]

[1,2,3,4] Department of CSE, Christ University, Bangalore, Karnataka, India.
[5,6] Department of ADSE, Christ University, Bangalore, Karnataka, India.
[1]ashagowda05@gmail.com, [2]*gl.anoop1@gmail.com, [3]deepa.yogish@christuniversity.in,
[4]shruti.jalapur@christuniversity.in, [5]babukumar.sbk@gmail.com,
[6]sa thrupth23@gmail.com *Corresponding Author

ABSTRACT: Facial expression prediction has gained considerable attention in recent years, particularly because of its healthcare related applications in human-computer interaction. This paper presents a comparative evaluation of machine learning and deep learning models for the prediction of face emotion using the CK+ dataset proposed by Cohn-Kanade. The dataset is characterized by seven classes of emotions rep-resented by the labels, namely surprise, happiness, disgust, anger, sadness, fear, and contempt, on 784 training, 98 validation, and 99 testing images. To further improve model performance, preprocessing techniques were employed that enhanced data efficiency. To increase variability in the data and reduce overfitting, all images were scaled to a 48*48-pixel resolution, pixel values were scaled to be between 0 and 1 for uniformity and the following data augmentation techniques were implemented: 10-degree rotation, horizontal flip, 0.15 zoom. The five models that were tested were CNN, SVM, VGG16, InceptionV3 and VGG19. The results demonstrate the high accuracy achieved by the CNN model which showed an accuracy of 98.98%, 99% and 99% in training, validation and test respectively. The SVM classifier achieved a testing accuracy of 99%. Both InceptionV3 and VGG19 on the other hand achieved competitive testing accuracy values of 90.91% and 97.98% respectively, while VGG16 got tested and reached an accuracy of 85.86%.

1 INTRODUCTION

Emotions are the initial indications of how someone feels on the inside when communicating with others. Through their tone of voice, physiological cues, or facial expressions, these emotions allow individuals to communicate with their surroundings and with each other, which has been transforming how people use technology [1]. Human emotional states can be inferred from both verbal and nonverbal cues recorded by different sensors, such as physiological signals, tone of voice, and facial changes [2]. Mehrabian [3] demonstrated in 1967 that 38% of emotional input was audible, 7% was verbal, and 55% was visual. The majority of academics are particularly interested in this modality since facial expressions during a conversation are the first indicators that convey an emotional state.

Facial emotion detection has become one of the most sought-after research areas in artificial intelligence and machine learning in the past few years due to its significance in smarter health management. Affective computing, mental illness diagnosis, security systems, and human-computer interaction are just a few of the numerous areas that can greatly benefit from the automatic detection and interpretation of human emotions from faces. Accurate emotion recognition makes user-interaction systems more reactive and responsive and provides informative information regarding user behavior and emotion-al states. Applications like these and others show how important it is to develop good and sound models for predicting emotions.

Facial expression recognition remains a grand challenge because human emotions are complex and the appearance of the face changes with each different emotion. However, high-quality labeled datasets remain extremely rare. The conventional methods fail in the capture of minute variations in expressions and further failed to generalize across diverse datasets. Choosing an appropriate model for accurate emotion classification is very important since all these models have their own strengths and weaknesses. In the case of a choice among various deep learning and machine learning techniques, such as CNN [4] [5], SVM [6] [7] and advanced architectures like VGG16, InceptionV3 and VGG19, there is definitely a need for comprehensive analysis to decide on the best option.

This work will therefore analyze the performance of these models on the CK+ dataset and try to shed light on which model can be useful for real-world robust facial emotion prediction. In recent times, facial expression recognition has gained large-scale attention, mainly in healthcare applications for human-computer interaction. The current study presents a comparison of different deep learning and ma-chine learning models for the prediction of facial emotion using the CK+ dataset. This dataset is characterized by the following seven classes of emotions represented by their corresponding labels: surprise, happiness, dis-gust, anger, sadness, fear and contempt, thus comprising 784 training images, 98 validation images and 99 testing images. In order to further enhance model performance, preprocessing techniques were resorted to that helped in properly improving data efficiency. All images were resized to provide more variability to data for preventing overfitting and rescaled so that pixel values were uniformly between 0 and 1: rotation through 10 degrees, horizontal flip and 0.15 zoom.

This paper discusses five models, namely CNN, SVM, VGG16, InceptionV3 and VGG19. The paper provides sufficient detail on how each model fared in terms of its strengths and weaknesses in using the CK+ dataset for emotion prediction.

2 LITERATURE SURVEY

Singh et al. [8] present a CNN-based facial emotion detection model, demonstrating superior accuracy over traditional methods and highlighting its applicability in healthcare, education, and interactive systems. Kasar et al. [9] propose "EmoSense," an optimized deep learning approach that enhances emotion recognition accuracy through advanced model tuning and constraint integration, vali-dated on benchmark datasets for real-world applications. Kanna et al. [10] devel-op a CNN-based emotion recognition system tailored for healthcare, achieving high accuracy and robustness, and emphasizing its role in improving patient care, communication, and monitoring.

In [11] A. Jaiswal et. al designed a model for detection of facial emotion using deep learning. The model involves three major steps: Face Detection, Feature Extraction, and Emotion Classification. The input samples were taken from two datasets: Facial Emotion Recognition Challenge (FERC-2013) and Japanese Fe-male Facial Emotion (JAFFE). Convolution Neural Network is a deep learning architecture, was used for detecting the emotion from facial images. Then, performance of the method was evaluated on both FERC-2013 and JAFFE dataset and obtained the accuracy of 70.14 and 98.65 respectively.

The authors in [12] propose a very deep CNN (DCNN) approach for facial emotion recognition using Transfer Learning (TL). A pre-trained model of DCNN is retrained by replacing and fine-tuning its dense layers, then applying a new pipe-line strategy that successively fine-tunes every pre-trained block. A model was tested on eight well-known DCNN architectures like VGG-16/19, ResNet-18/34/50/152, Inception-v3, DenseNet-161 using KDEF and JAFFE datasets. Results show high accuracies-96.51% (KDEF) and 99.52% (JAFFE) with Dense-Net-161-outperforming existing methods, including for profile views, mak-ing it promising for real-world applications.

An author [13] addresses the problem of emotion recognition for students in e-learning, classifying seven emotions from facial images in the FER-2013 dataset. Deep learning architectures VGG16, ResNet, and MobileNetV2 were experimented with, and a proposed model named TTNet-a variant of ResNet and VGG-Face2-was introduced, achieving more than 70% accuracy. A software application was created to automatically capture facial information and interpret learners' emotions in real-time, allow-ing instructors to track engagement. Comparative outcomes against six other approaches demonstrated TTNet performed superior to them, qualifying it as a viable emotion management tool for web-based courses and conferences.

Gupta et.al [14] addresses the difficulty of keeping students engaged in online learning during the COVID-19 pandemic by developing a deep learning-based real-time engagement monitoring system that use facial emotion identification. It scans pupils' facial expressions to classify emotions, computes an Engagement Index (EI), and predicts whether they are "engaged" or "dis-engaged." The system was tested using Inception-V3, VGG19, and ResNet-50 models on datasets such as FER-2013, CK+, RAF-DB, and a custom dataset. The results reveal that ResNet-50 outperforms other models for real-time facial emotion categorization in online learning, with an accuracy of 92.3 percent.

A study in [15] explores the field of automatic emotion recognition from facial expressions, which is gaining popularity due to its numerous applications in do-mains such as safety systems, healthcare, and human-machine interfaces. The primary goal of this study was to create strategies for interpreting and coding facial expressions, as well as extracting relevant elements, so that computers can accurately predict human emotion. With recent advances in deep learning, re-searchers have been able to significantly enhance performance by leveraging diverse designs such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models. This study focuses on recent research in Facial Emotion Recognition (FER) using deep learning algorithms.

Chahak Gautam and K.R Seeja [16] proposed a model to recognize Facial emotion using Handcrafted features and CNN. The image samples are taken from Jaffe dataset contains 213 labelled images and CK+ dataset, which is available on Kaggle and it contains 981 grayscale images. The feature descriptor like HOG, which is a technique that extracts characteristics from the input image. This feature descriptor concerned about structure of an object. SIFT was applied to extract the local features from the images. These features involve essential information which mains regards to the emotion detection process. These key features include information about the jawline structure, eyes are winked and mouth open etc. these hand-crafted features are fed into the convolutional and dense layer neural network for classification. Model is evaluated on both datasets, obtained 98.48% for HOG-CNN and 97.96% for SIFT-CNN for CK+ dataset and 91.43% for HOG-CNN while 82.85% for SIFT-CNN for Jaffe dataset.

Kassab et al. [21] introduced a Graph Convolutional Framework (GCF) that combines CNN-based feature extraction with Graph Convolutional Networks to improve facial expression recognition. The model was evaluated on CK+, JAFFE, and FERG datasets and achieved improved performance over conventional CNN models. On the CK+ dataset, the proposed framework increased the accuracy of ResNet18 from 92% to 98% and VGG16 from 89% to 97%, demonstrating the benefits of graph-based feature learning.

Dada et al. [22] proposed a CNN-10 model for facial expression recognition and compared its performance with VGG19, InceptionV3, and Vision Transformer (ViT). The CNN-10 model achieved superior performance, reporting 99.9% accuracy on the CK+ dataset, demonstrating its effectiveness in extracting discriminative facial features for emotion recognition. Przybyła-Kasperek and Zieliński [23] compared five CNN architectures for facial emotion recognition across FERPlus, CK+, and AffectNet datasets. The study highlighted the importance of cross-dataset evaluation and reported that ResNet34v2 achieved the highest accuracy of 83.6% on the CK+ dataset, while emphasizing that dataset quality and class imbalance strongly affect model performance.

Sarvakar and Rana [24] presented a hybrid CNN-LSTM framework with transfer learning for facial emotion recognition. Their experimental results showed that ResNet-50 achieved 76.85% accuracy on the FER2013 dataset, whereas the hybrid CNN-LSTM model obtained 92.30% accuracy on the CK+ dataset, demonstrating the effectiveness of transfer learning and temporal modeling for improving recognition performance.

Aaniba et al. [25] conducted a comparative study of K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Convolutional Neural Network (CNN) for facial emotion recognition using the CK+ dataset. To ensure a fair comparison, all models were evaluated under a unified experimental framework using facial regions extracted with the Dlib library and resized to 48 × 48 pixels. The experimental results showed that the CNN model achieved the highest recognition accuracy of 98.74%, followed by SVM with 98.53% and KNN with 96.85%. Although CNN provided the best classification performance, the study highlighted that SVM offers an effective trade-off between recognition accuracy and computational efficiency, making it suitable for real-time facial emotion recognition applications. Chourasia et al. [26] proposed a hybrid facial expression recognition framework that combines Convolutional Neural Networks (CNN) with handcrafted feature extraction techniques, including HOG, LBP, SIFT, and Gabor filters. The study systematically compared these feature representations on the CK+48 dataset and reported that the Gabor-CNN model achieved the highest accuracy of 98.47%, demonstrating that integrating handcrafted texture descriptors with deep learning significantly improves facial emotion recognition performance.

Natural Resources for Human Health
eISSN: 2583-1194

Table 1. Comparison of various techniques for facial recognition

Title & Year	Techniques	Dataset	Results
[4], 2024	DCNN	FER2013	82.56%
	EfficientNet		62.15%
	ResNet		59.41%
	VGGNet		50.31%
[11], 2020	CNN	Facial Emotion Recognition Challenge (FERC-2013)	70.14
		Japanese Female Facial Emotion (JAFFE)	98.65
[12], 2021	DCNN	KDEF dataset	96.51%
		JAFFE dataset	99.52%
[13], 2024	TTNet	FER-2013	-
[14], 2023	Inception-V3, VGG19, and ResNet-50	FER-2013, CK+, RAF-DB, and a custom dataset	92.3%
[15], 2023	CNN	FER-2013	95.6%
[16], 2023	HOG-CNN	CHplus Dataset	98.48%
	SIFT-CNN	CHplus Dataset	97.96%
	HOG-CNN	JAFFE Dataset	91.43%
	SIFT-CNN	JAFFE Dataset	82.85%
[17], 2015	SVM	JAFFE database	87%
		Mevlana University Facial Expression (MUFE) database.	77%
[18], 2023	CNN-SVM	JAFEE Dataset	98%
[19], 2022	CNN+SVM	BU4D	99%
		CK+	94%
[20], 2022	CNN-based VGG16	FER-2013 dataset	64.52%

Although several studies have reported high recognition accuracy on the CK+ dataset, most focus on a single deep learning architecture or compare a limited number of classifiers. Furthermore, relatively few studies perform a comprehensive comparison between traditional machine learning models, custom CNN architectures, and transfer learning models [27] under identical preprocessing and evaluation settings. Motivated by this gap, the present study evaluates CNN, SVM, VGG16, VGG19, and InceptionV3 using a unified experimental framework.

3 DEEP LEARNING AND MACHINE LEARNING TECHNIQUES

This paper utilizes a blend of machine learning and deep learning techniques for the recognition of facial emotions into seven categories: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. The adopted models are VGG16, VGG19, In-ceptionV3, CNN and SVM classifier. Preprocessed facial image data were used to train the aforementioned models, where images were resized, normalized, and augmented to gain higher generalization and prevent overfitting.

Although several studies have reported high recognition accuracy on the CK+ dataset, most focus on a single deep learning architecture or compare a limited number of classifiers. Furthermore, relatively few studies perform a comprehensive comparison between traditional machine learning models, custom CNN architectures, and transfer learning models under identical preprocessing and evaluation settings. Motivated by this gap, the present study evaluates CNN, SVM, VGG16, VGG19, and InceptionV3 using a unified experimental framework.

3.1 VGG16 and VGG19

The VGG19 and VGG16 models were employed with transfer learning, where pre-trained feature extraction weights were borrowed from the ImageNet data-base. The architectures are a series of convolutional layers featuring small-sized kernels of 3×3. These are succeeded by max pooling layers which progressively shrink the spatial dimensions but retain important facial features. These learned features are then fed through a Global Average Pooling (GAP) layer once more to compress the feature maps prior to dense layer inputs. For VGG16, the densely connected layers use 512 and 256 units with a drop of 0.3 not to overfit. VGG19 is also the same but uses dropout layers of 0.5 and 0.3 for a better regularization scheme. Both models employ the Adam optimizer and categorical cross-entropy loss for maximum learning, topped off with a softmax activation layer at the out-put to classify the emotions into seven classes.

3.2 InceptionV3

The InceptionV3 model, pre-trained on ImageNet similar to the other one, employs inception modules that make use of multiple kernel sizes in parallel so that the network can extract fine-grained as well as abstract facial features. Fine-tuning was implemented in the uppermost layers of the model in order to get it task-specific for the identification of facial emotion. Just like in VGG networks, InceptionV3 utilizes Global Average Pooling, followed by dense layers having 512 and 256 neurons respectively with a dropout of 0.3 for overfitting avoidance. Final classification is done with the use of softmax activation with seven neurons. The model is assisted by an adaptive Adam optimizer, in which learning rate dynamically adapts to make the network more efficient in feature learning and emotion classification.

3.3 SVM Classifier

SVM classifier, a machine learning approach, was utilized for performance com-parison with deep learning techniques. The model was trained with high-dimensional feature vectors learned from pre-trained CNN models rather than raw image data. By using a linear kernel, the SVM classifier determines the best decision boundaries for discrimination between various emotion classes. SVMs are less susceptible to overfitting than deep learning models, especially when dealing with small datasets, and thus are an appropriate option in scenarios where computational resources are limited or dataset size is a constraint.

3.4 CNN

The CNN network was built as a tailored model with a specific configuration for multi-class emotion recognition to identify intricate facial features with hierarchical convolutional layers. The architecture starts with the use of a 512-filter convolutional layer with a kernel size of 5×5 followed by batch normalization to regularize the learning and enhance the gradient flow. The next layer is a 256-filter convolutional layer that enhances feature extraction and identifies deeper patterns in the facial images. As an added layer, another third convolutional layer with 128 filters and a kernel size of 3×3 is included to read fine details. There are MaxPooling layers placed at various places for downscaling spatial dimensions without sacrificing important features. Dropout layers are placed strategically at places in the network in order to avoid overfitting. It also employs a dense layer that is fully connected with 256 neurons and has the ability to learn the deeper facial expression representations prior to the final softmax activation layer classifying the emotions into seven classes. CNN architecture is constructed with the Adam optimizer and categorical cross-entropy loss, which provides stability and efficacy to training. The CNN architecture in detail is illustrated in Figure 1.

Natural Resources for
Human Health
eISSN: 2583-1194

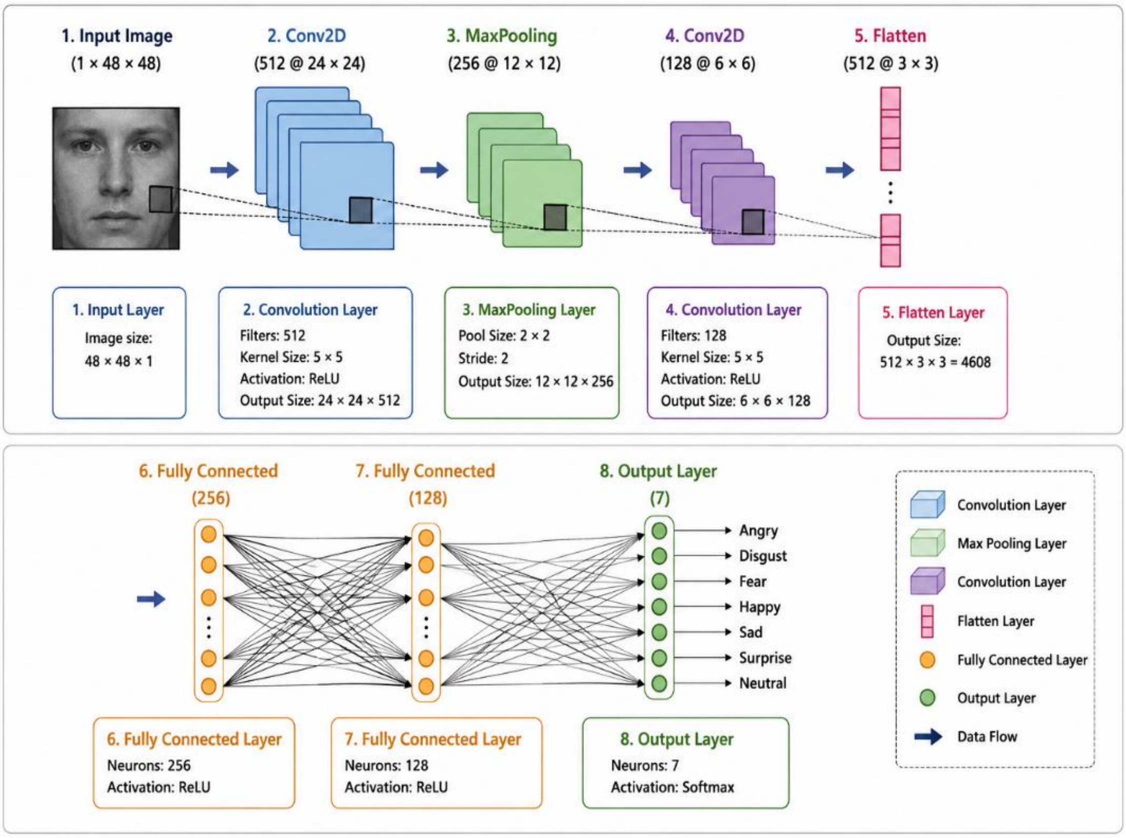

Fig. 1. CNN Architecture.

All of these models have separate strengths in face emotion detection. VGG models are good at feature extraction through their deep convolutional architecture, while InceptionV3 contributes computational effectiveness through inception modules for extracting multi-scale spatial features. SVM classifier offers a consistent option if dataset size is a problem, and the personalized CNN model offers the ability to learn directly from features without depending on pre- trained networks. Through the comparison and examination of these models, this research will conclude with the best method of facial emotion categorization under different conditions.

4 METHODOLOGY

The methodology focuses on the processing and preparation of the CK+ dataset, data preprocessing, data augmentation, and then examining several deep learning models to find out the most effective way of facial emotion prediction. End-to-End Workflow for Facial Emotion Detection Using CNN is given in the figure 2.

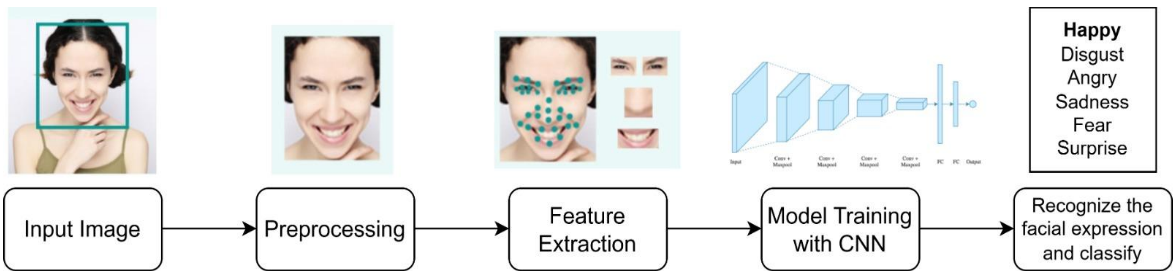

Fig. 2. End-to-End Workflow for Facial Emotion Detection Using CNN.

Dataset: The Extended Cohn-Kanade (CK+) dataset was used for training and evaluating the proposed models. The dataset contains labelled grayscale facial images corresponding to seven emotion categories: anger, disgust, fear, happiness, sadness, surprise, and neutral. It is one of the most widely used benchmark datasets for facial expression recognition due to its high-quality annotations and diverse facial expression samples. To facilitate model development and evaluation, the dataset was divided into training, validation, and testing subsets. Approximately 80% of the data was allocated for training and 20% for testing, while 10% of the training data was reserved for validation to monitor model performance and reduce overfitting during training.

Preprocessing and Data Augmentation: Prior to model training, all facial images were resized to 48 × 48 pixels to maintain a uniform input dimension across all models. Pixel values were normalized to the range [0,1] to improve training stability and convergence. To enhance dataset diversity and reduce overfitting, data augmentation techniques including horizontal flipping, rotation (10°), and zooming (0.15) were applied during training. These preprocessing steps improved the robustness and generalization capability of the learning models.

Model Development and Evaluation: Five classification models, namely CNN, SVM, VGG16, VGG19, and InceptionV3, were implemented and evaluated for facial emotion recognition. The models were trained using the training dataset and validated during the learning process to monitor performance and minimize overfitting. After training, the models were evaluated on the test dataset to assess their generalization capability. Performance was measured using training, validation, and testing accuracies, enabling a comparative analysis of the effectiveness of each model for emotion classification.

5 RESULTS AND DISCUSSION

The performance of the proposed facial emotion recognition framework was evaluated using five classification models, namely Convolutional Neural Network (CNN), Support Vector Machine (SVM), VGG16, VGG19, and InceptionV3. All models were trained and evaluated on the CK+ dataset using identical preprocessing techniques and experimental settings to ensure a fair comparison. The deep learning models (CNN, VGG16, VGG19, and InceptionV3) were analyzed using training and validation accuracy and loss curves to examine their learning behaviour during training. Since the SVM classifier does not follow an epoch-based training process, its performance was evaluated using classification accuracy, the classification report heatmap, and the confusion matrix. Furthermore, confusion matrices were used to investigate the class-wise classification capability of all evaluated models.

Figures 3 and 4 present the training and validation accuracy and loss curves of the deep learning models. These learning curves provide valuable insight into the convergence characteristics, learning stability, and generalization capability of each architecture.

As illustrated in Figure 3, the CNN model converged rapidly during the initial training epochs and consistently maintained high training and validation accuracies throughout the learning process. The small gap between the training and validation curves indicates excellent generalization capability with minimal overfitting. The learning process stabilized after approximately 20 epochs, demonstrating efficient optimization and robust feature extraction.

Among the transfer learning models, VGG19 exhibited stable convergence with steadily increasing training and validation accuracies. Although its convergence rate was slightly slower than CNN, the model maintained a relatively small difference between the training and validation curves, indicating effective utilization of pretrained features for facial emotion recognition.

In contrast, InceptionV3 and VGG16 exhibited comparatively slower convergence during training. Both models showed larger differences between training and validation accuracies, indicating comparatively weaker generalization capability. VGG16 displayed the largest performance gap, suggesting that the model was less effective in learning discriminative facial representations from the CK+ dataset under the adopted experimental settings.

The corresponding training and validation loss curves shown in Figure 4 further support these observations. CNN achieved the lowest training and validation loss with smooth convergence throughout the training process. VGG19 also demonstrated a continuous reduction in validation loss, confirming stable optimization and improved learning behaviour. Conversely, InceptionV3 and VGG16 maintained comparatively higher validation loss values, indicating greater prediction uncertainty during classification. Overall, the learning curves demonstrate that CNN and VGG19 exhibit superior convergence behaviour and learning stability compared with the remaining deep learning models.

Table 2 summarizes the quantitative performance of all evaluated models using training, validation, and testing accuracies.

Among all evaluated models, the SVM classifier achieved the highest overall performance, recording an accuracy of 99.00% across the training, validation, and testing datasets. Unlike the deep learning models, SVM was evaluated based on its final

Natural Resources for
Human Health
eISSN: 2583-1194

classification performance rather than epoch-wise learning behaviour. The confusion matrix (Figure 6) further demonstrate that the SVM classifier accurately distinguished the majority of facial expression classes with minimal misclassification, confirming its effectiveness for facial emotion recognition.

The CNN model achieved comparable performance with training, validation, and testing accuracies of 98.98%, 99.00%, and 99.00%, respectively. These results indicate that the proposed CNN architecture effectively learned discriminative facial representations while maintaining excellent generalization capability.

Table 2. Accuracies for emotion prediction for all the models.

Model	Training accuracy	Validation accuracy	Testing accuracy
CNN	98.98%	99%	99%
SVM	99%	99%	99%
VGG16	88.90%	83.67%	85.86%
InceptionV3	95.41%	92.86%	90.91%
VGG19	98.21%	95.92%	97.98%

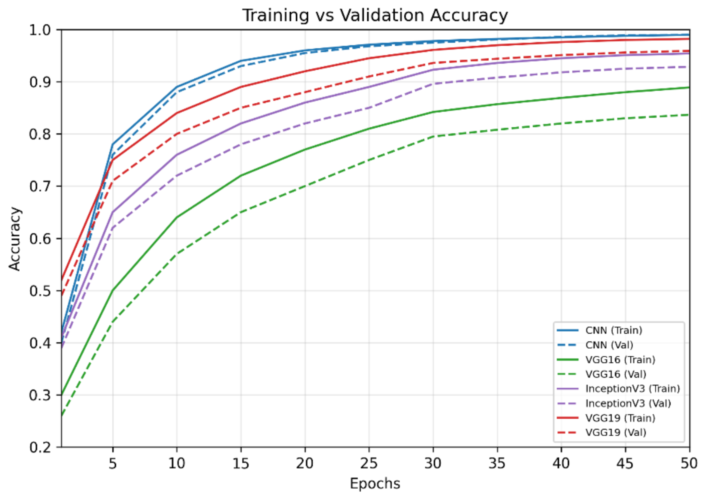

Fig. 3. Training and validation accuracy curves of CNN, VGG16, VGG19, and InceptionV3 models.

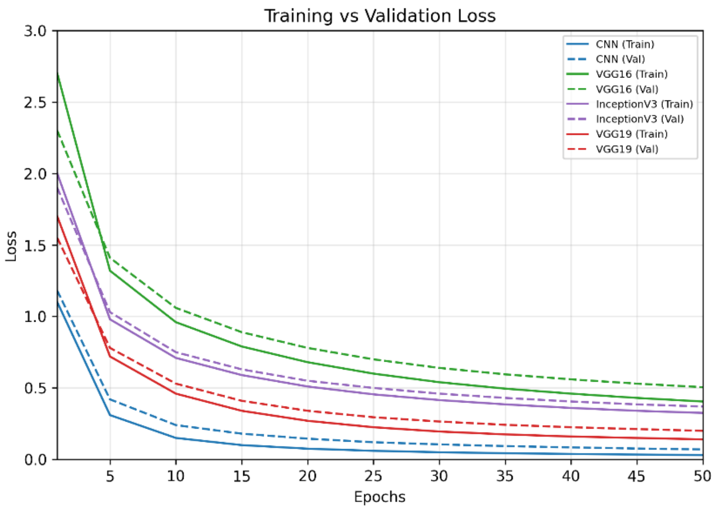

Fig. 4. Training and validation loss curves of CNN, VGG16, VGG19, and InceptionV3 models.

eISSN: 2583-1194

Among the transfer learning models, VGG19 achieved the highest testing accuracy (97.98%), confirming its effectiveness in extracting hierarchical facial features for emotion recognition. InceptionV3 obtained a testing accuracy of 90.91%, demonstrating satisfactory performance but comparatively lower classification capability than CNN, SVM, and VGG19. VGG16 produced the lowest testing accuracy (85.86%), indicating that increased model complexity alone does not necessarily guarantee improved recognition performance on relatively small facial expression datasets such as CK+.

The comparative evaluation clearly indicates that CNN and SVM outperform the remaining evaluated models in terms of classification accuracy and generalization capability. While CNN demonstrated superior learning behaviour among the deep learning models, the SVM classifier achieved the highest overall recognition accuracy with comparatively lower computational complexity. These findings suggest that both traditional machine learning and deep learning approaches can achieve highly reliable facial emotion recognition when appropriate feature extraction and model optimization strategies are employed.

The confusion matrices shown in Figures 5–9 provide a detailed class-wise evaluation of the prediction capability of each model. Unlike overall accuracy, confusion matrices reveal the correctly classified samples as well as the specific emotion categories that are prone to misclassification.

The VGG16 confusion matrix (Figure 5) exhibits comparatively higher off-diagonal values, indicating increased confusion among certain emotion categories. Misclassifications are primarily observed between fear, disgust, happiness, and sadness, suggesting that VGG16 encounters difficulty in distinguishing subtle facial variations among these expressions. Nevertheless, the model consistently classified dominant expressions such as happiness and surprise, indicating its ability to recognize highly distinguishable facial patterns.

Fig. 5. VGG16 confusion matrix

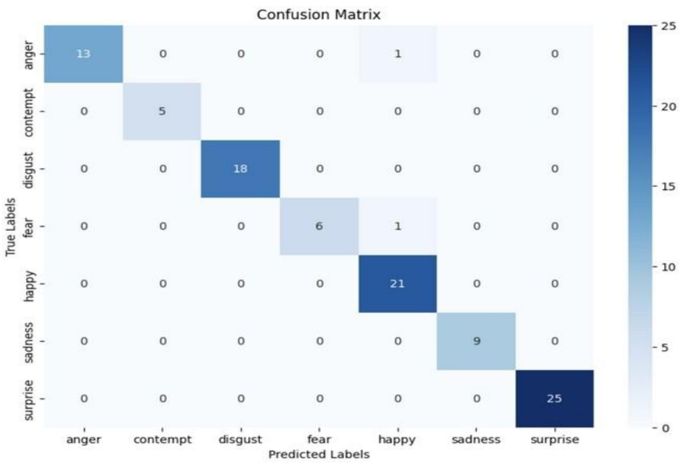

Fig. 6. VGG19 Confusion Matrix

Natural Resources for
Human Health
eISSN: 2583-1194

Figure 6 presents the confusion matrix of VGG19. Compared with VGG16, the model exhibits considerably stronger diagonal dominance, indicating improved classification capability across all emotion classes. Only a few isolated misclassifications are observed, demonstrating the effectiveness of deep hierarchical feature extraction for facial emotion recognition.

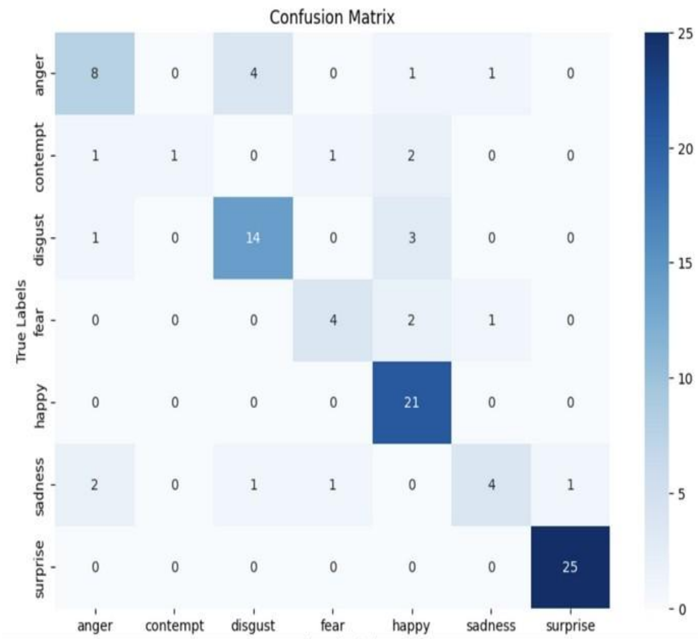

Fig. 7. Inception v3 confusion matrix

The InceptionV3 confusion matrix (Figure 7) reveals moderate classification performance. Although the model accurately classified the majority of happy, sad, and surprise expressions, relatively higher confusion was observed for anger, contempt, and fear, indicating comparatively weaker feature discrimination for subtle facial expressions.

Figure 8 illustrates the confusion matrix of the SVM classifier. The matrix exhibits strong diagonal dominance with very few off-diagonal elements, confirming that the classifier correctly recognized almost all facial expression categories. The excellent class separation achieved by SVM demonstrates the effectiveness of discriminative feature representation for facial emotion recognition.

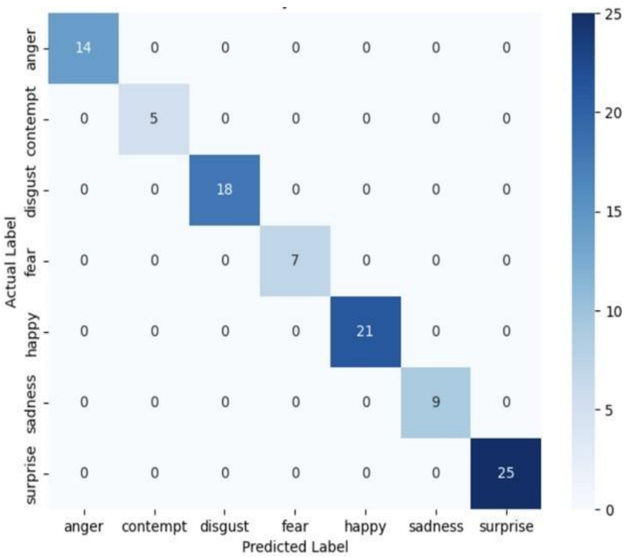

Fig. 8. SVM Classifier confusion matrix

eISSN: 2583-1194

Similarly, the CNN confusion matrix shown in Figure 9 demonstrates nearly perfect classification performance across all seven emotion categories. Almost every sample is correctly classified, resulting in minimal confusion between emotion classes. The strong diagonal structure confirms the ability of the proposed CNN architecture to learn highly discriminative facial features while maintaining excellent generalization capability.

Overall, the confusion matrix analysis supports the quantitative performance reported in Table 2 and further demonstrates the superior class-wise prediction capability of CNN and SVM compared with the remaining evaluated models.

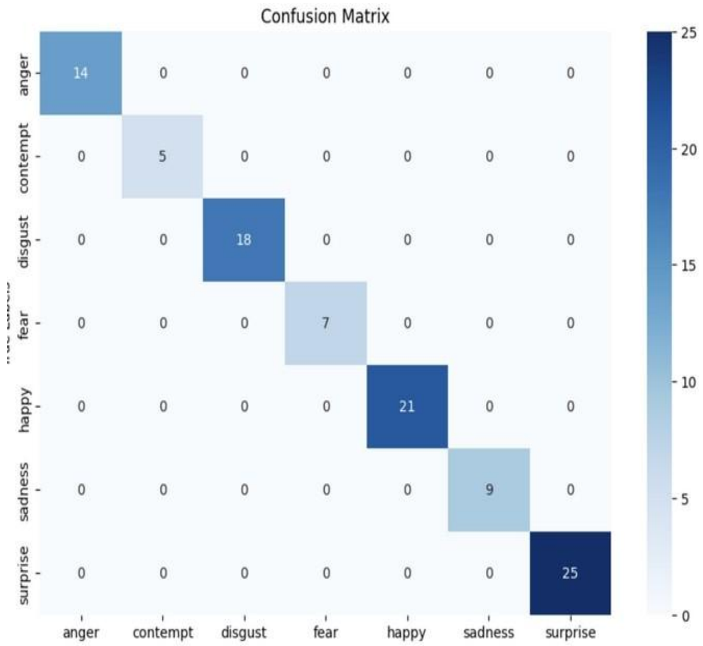

Fig. 9. CNN confusion matrix

Besides classification performance, computational efficiency is an important consideration for practical facial emotion recognition systems. Deep transfer learning models such as VGG16 and VGG19 contain a substantially larger number of trainable parameters, resulting in increased computational and memory requirements during training and inference. Although these models provide competitive recognition performance, their complexity makes them less suitable for resource-constrained environments.

In contrast, the custom CNN architecture achieves comparable recognition accuracy with significantly lower computational complexity, making it suitable for real-time facial emotion recognition applications. The SVM classifier also exhibits low inference complexity once feature extraction has been completed, making it an efficient alternative for practical deployment. InceptionV3 provides a balance between computational complexity and recognition performance but does not outperform CNN or SVM under the adopted experimental settings.

Table 3. Comparison of the Present Study with Recent Facial Emotion Recognition Approaches.

Reference	Method	Dataset	Accuracy (%)
Dada et al. [22]	CNN-10	CK+	99.95
Kassab et al. [21]	GCF (CNN + GCN)	CK+	98.00
Sarvakar and Rana [24]	Hybrid CNN-LSTM	CK+	92.30
Aaniba et al. [25]	CNN, SVM, KNN	CK+	98.74 (CNN) 98.53 (SVM) 96.85 (KNN)
Hybrid CNN + Gabor Features [26]	CNN + Gabor Features	CK+	98.47

Present Study	**CNN, SVM, VGG16, VGG19, InceptionV3**	**CK+**	**99.00 (SVM), 99.00 (CNN), 97.98 (VGG19), 90.91 (InceptionV3), 85.86 (VGG16)**

Table 3 compares the performance of the present study with recently published facial emotion recognition approaches evaluated on the CK+ dataset. The comparison demonstrates that the evaluated models achieve competitive recognition performance under identical preprocessing and experimental settings. Among the evaluated models, both the SVM classifier and the custom CNN architecture achieved the highest testing accuracy of 99.00%, followed by VGG19 (97.98%), InceptionV3 (90.91%), and VGG16 (85.86%). Although the CNN-10 model proposed by Dada et al. reported a slightly higher accuracy of 99.95%, the current work provides a comprehensive comparative analysis of traditional machine learning, deep learning, and transfer learning models using a unified experimental framework. Unlike many existing studies that evaluate only a single architecture or compare a limited number of models, the present study systematically investigates five widely adopted classification approaches under identical preprocessing, augmentation, training, and evaluation conditions. This comprehensive evaluation offers valuable insights into the strengths and limitations of different facial emotion recognition models, enabling a fair comparison of their classification performance, convergence behaviour, and generalization capability.

6 CONCLUSION

Emotion prediction has wide applications in human- computer interaction, monitoring of mental health and personalized learning systems. However, recognizing facial expressions has some innate difficulties: human emotions have multiple dimensions and subtleties, variation in faces is large and datasets are limitedly diverse. Emotion detection requires computationally robust models with generalization from different environmental or situational conditions. The selection of an appropriate model remains critical because this will, in turn, have a great im-pact on the accuracy of classification and computational efficiency. Values for the CNN and SVM classifiers with scores close to or at 99%, for training, validation and testing, embody their strong generalization ability across varying splits of data. This shows that they are best suited for emotion prediction. On the other hand, competitive deep learning VGG16 and InceptionV3 perform slightly worse in accuracy. The VGG19 model was really strong, with accuracy scores of almost 98%, showing quite clearly its capacity for extraction of relevant features from the data to classify the different emotions. The results point out CNN and SVM as superior classifiers since they possess the capability to gather facial information effectively with relatively lower computational cost. Future directions could be toward hybrid models, representing the strengths of both SVM and CNN, where traditional classifiers make the predictions after feature extractions by CNNs. Again, expanding the dataset to broader variations in facial expressions across age groups and ethnicities, and under different conditions, would lead to further model robustness. Other regularization techniques and advanced augmentation methods can also reduce overfitting and improve adaptability for models to real-world applications such as the assessment and improvement of mental health in human-computer interaction systems.

REFERENCES

[1] R. A. Elsheikh, M. A. Mohamed, A. M. Abou-Taleb, and M. M. Ata, “Improved facial emotion recognition model based on a novel deep convolutional structure,” Sci Rep, vol. 14, no. 1, p. 29050, 2024, doi: 10.1038/s41598-024-79167-8.

[2] L. Shu et al., « A Review of Emotion Recognition Using Physiological Signals », Sensors, vol. 18, no 7, p. 2074, juill. 2018, doi: 10.3390/s18072074.

[3] D. and T. K. and P. P. and B. L. and A. C. and W.-W. K. Marechal Catherine and Mikołajewski, “Survey on AI-Based Multimodal Methods for Emotion Detection,” in High-Performance Modelling and Simulation for Big Data Applications: Selected Results of the COST Action IC1406 cHiPSet, H. Kołodziej Joanna and González-Vélez, Ed., Cham: Springer International Publishing, 2019, pp. 307–324. doi: 10.1007/978-3-030-16272-6_11.

[4] D. Bhagat, A. Vakil, R. K. Gupta, and A. Kumar, "Facial Emotion Recognition (FER) using Convolutional Neural Network (CNN)," Procedia Comput Sci, vol. 235, pp. 2079–2089, 2024, doi: https://doi.org/10.1016/j.procs.2024.04.197.

[5] M. Mehrotra, K. P. Singh, and Y. B. Singh, "Facial Emotion Recognition and Detection Using Convolutional Neural Networks with Low Computation Cost," in 2024 2nd International Conference on Disruptive Technologies (ICDT), 2024, pp. 1349–1354. doi: 10.1109/ICDT61202.2024.10489678.

[6] L. Yao, Y. Wan, H. Ni, and B. Xu, "Action unit classification for facial expression recognition using active learning and SVM," Multimed Tools Appl, vol. 80, no. 16, pp. 24287–24301, 2021, doi: 10.1007/s11042-021-10836-w.

[7] K. R. Radhika, H. N. Shenoy, T. R. Vinay, H. Pooja, D. Sharma, R. Priyanka, and S. Gupta, "Communication-Efficient Federated Learning (CEFL) for CT image classification in bandwidth-constrained wireless healthcare networks," International Journal of Drug Delivery Technology, vol. 16, no. 13S, pp. 163–172, 2026, doi: 10.25258/IJDDT.16.13S.17.

[8] S. Singh, M. Nazmul Alam, and S. Lata, "Facial Emotion Detection Using CNN-Based Neural Network," INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT, vol. 07, p. 7, Aug. 2023, doi: 10.55041/IJSREM26437.

[9] M. Kasar, P. Kavimandan, T. Suryawanshi, and B. Garg, "EmoSense: Pioneering Facial Emotion Recognition with Precision Through Model Optimization and Face Emotion Constraints," International Journal of Engineering, vol. 38, no. 1, pp. 35–45, 2025, doi: 10.5829/ije.2025.38.01a.04.

[10] Kishore Kanna, R., et al. (2024). CNN based face emotion recognition system for healthcare application. EAI. doi: 10.4108/eetpht.10.5458.

[11] A. Jaiswal, A. Raju, and S. Deb, "Facial Emotion Detection Using Deep Learning," Aug. 2020, pp. 1–5. doi: 10.1109/INCET49848.2020.9154121.

[12] 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 (Basel), vol. 10, no. 9, 2021, doi: 10.3390/electronics10091036.

[13] T.-D. Tran, "TTNet: A novel machine learning model for facial emotion detection in online learning systems," SoftwareX, vol. 27, p. 101787, 2024, doi: https://doi.org/10.1016/j.softx.2024.101787.

[14] S. Gupta, P. Kumar, and R. K. Tekchandani, "Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models," Multimed Tools Appl, vol. 82, no. 8, pp. 11365–11394, 2023, doi: 10.1007/s11042-022-13558-9.

[15] W. Mellouk and W. Handouzi, "Facial emotion recognition using deep learning: review and insights," Procedia Comput Sci, vol. 175, pp. 689–694, 2020, doi: https://doi.org/10.1016/j.procs.2020.07.101.

[16] C. Gautam and K. R. Seeja, "Facial emotion recognition using Handcrafted features and CNN," Procedia Comput Sci, vol. 218, pp. 1295–1303, 2023, doi: https://doi.org/10.1016/j.procs.2023.01.108.

[17] M. Abdulrahman and A. Eleyan, "Facial expression recognition using Support Vector Machines," 2015 23rd Signal Processing and Communications Applications Conference, SIU 2015 - Proceedings, pp. 276–279, Aug. 2015, doi: 10.1109/SIU.2015.7129813.

[18] Rishu, V. Kukreja, and A. Kumar, "Facial Expression Recognition with CNN-SVM for Emotional State Classification," Aug. 2023, pp. 1–5. doi: 10.1109/ICCUBEA58933.2023.10392007.

[19] J.-C. Kim, M.-H. Kim, H.-E. Suh, M. T. Naseem, and C.-S. Lee, "Hybrid Approach for Facial Expression Recognition Using Convolutional Neural Networks and SVM," Applied Sciences, vol. 12, no. 11, 2022, doi: 10.3390/app12115493.

[20] P. Kanagaraju, M. Ranjith, and K. Vijayasarathy, "Emotion detection from facial expression using image processing," Int J Health Sci (Qassim), pp. 1368–1379, Aug. 2022, doi: 10.53730/ijhs.v6nS6.9748.

eISSN: 2583-1194

[21] Kassab, H., Bahaa, M., & Hamdi, A. (2024). GCF: Graph Convolutional Networks for Facial Expression Recognition. ArXiv. https://arxiv.org/abs/2407.02361

[22] Dada, Emmanuel Gbenga, Oyewola, David Opeoluwa, Joseph, Stephen Bassi, Emebo, Onyeka, Oluwagbemi, Olugbenga Oluseun, Facial Emotion Recognition and Classification Using the Convolutional Neural Network-10 (CNN-10), Applied Computational Intelligence and Soft Computing, 2023, 2457898, 19 pages, 2023. https://doi.org/10.1155/2023/2457898

[23] Przybyła-Kasperek, M., Zieliński, K. (2025). Comparative Analysis of Convolutional Neural Network Architectures for Emotion Recognition in Facial Images. In: Nguyen, N.T., et al. Recent Challenges in Intelligent Information and Database Systems. ACIIDS 2025. Communications in Computer and Information Science, vol 2495. Springer, Singapore. https://doi.org/10.1007/978-981-96-5887-9_18

[24] Sarvakar, K., & Rana, K. (2025). A Hybrid Framework Combining CNN, LSTM, and Transfer Learning for Emotion Recognition. ICCK Transactions on Machine Intelligence, 1(2), 103–116. https://doi.org/10.62762/TMI.2025.572412.

[25] I. Aaniba, H. Chouhayebi, N. Alioua, O. A. Fdili and A. Bendarag, "A Comparative Study Of Classical And Deep Learning Models For Facial Emotion Recognition On The CK+ Dataset," 2025 IEEE International Conference on Advances in Data-Driven Analytics And Intelligent Systems (ADACIS), Sousse, Tunisia, 2025, pp. 1-6, doi: 10.1109/ADACIS65663.2025.11437333.

[26] Chourasia, N., Lamba, C.S. & Gupta, A.K. Hybrid deep learning and feature-based approaches for facial expression recognition. Discov Computing 29, 251 (2026). https://doi.org/10.1007/s10791-026-10146-5.

[27] S. Gupta, I. S. Rajesh, C. Singh, U. N. Ranjitha, K. Siddesha, and P. K. Sekharamantry, “A Hybrid CEFL Approach Using Gradient Sparsification and Quantization for Medical Imaging,” International Journal of Computational Intelligence in Engineering (IJCIE), vol. 1, no. 1, pp. 1–15, 2026.