



Facial Expression Recognition Model Depending on Optimized Support Vector Machine

Amel Ali Alhussan¹, Fatma M. Talaat², El-Sayed M. El-kenawy³, Abdelaziz A. Abdelhamid^{4,5},
Abdelhameed Ibrahim⁶, Doaa Sami Khafaga^{1,*} and Mona Alnagar⁷

¹Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh, 11671, Saudi Arabia

²Machine Learning & Information Retrieval Department, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh, 33516, Egypt

³Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura, 35111, Egypt

⁴Department of Computer Science, College of Computing and Information Technology, Shaqra University, 11961, Saudi Arabia

⁵Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, 11566, Egypt

⁶Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Mansoura, 35516, Egypt

⁷Robotics and Intelligent Machines Department, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh, 33516, Egypt

*Corresponding Author: Doaa Sami Khafaga. Email: dskhafga@pnu.edu.sa

Received: 25 January 2023; Accepted: 18 April 2023; Published: 09 June 2023

Abstract: In computer vision, emotion recognition using facial expression images is considered an important research issue. Deep learning advances in recent years have aided in attaining improved results in this issue. According to recent studies, multiple facial expressions may be included in facial photographs representing a particular type of emotion. It is feasible and useful to convert face photos into collections of visual words and carry out global expression recognition. The main contribution of this paper is to propose a facial expression recognition model (FERM) depending on an optimized Support Vector Machine (SVM). To test the performance of the proposed model (FERM), AffectNet is used. AffectNet uses 1250 emotion-related keywords in six different languages to search three major search engines and get over 1,000,000 facial photos online. The FERM is composed of three main phases: (i) the Data preparation phase, (ii) Applying grid search for optimization, and (iii) the categorization phase. Linear discriminant analysis (LDA) is used to categorize the data into eight labels (neutral, happy, sad, surprised, fear, disgust, angry, and contempt). Due to using LDA, the performance of categorization via SVM has been obviously enhanced. Grid search is used to find the optimal values for hyperparameters of SVM (C and gamma). The proposed optimized SVM algorithm has achieved an accuracy of 99% and a 98% F1 score.



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Keywords: Facial expression recognition; machine learning; linear discriminant analysis (LDA); support vector machine (SVM); grid search

1 Introduction

One of the most potent, inherent, and common ways for people to express their emotions and intentions is through facial expression. Due to its practical significance in social robotics [1,2], medical treatment, driver tiredness detection, and many other human-computer interface systems, automated facial expression analysis has been the subject of several types of research. In addition, to identify a student's learning level in real time using emotion detection based on facial expressions in distant learning. As a result, adapting teaching tactics in virtual learning environments to the feelings of the students might be helpful for teachers. So Different facial expression recognition (FER) techniques [3] have been investigated in the fields of computer vision and machine learning to encode expression information from face representations. Automated facial expression analysis has been the focus of several studies due to its practical importance. To encode expression information from face representations, many FER approaches have been studied in the domains of computer vision and machine learning.

Human emotion recognition, which includes facial expressions [4], voice [5], and texts [6], is an emerging area in the field of Cognitive Computing. The key to the next era of digital evolution is to understand human feelings. In sectors including mental health [7], intelligent cars [8], and music [9], recent advancements in the discipline have realized their promise. Recognizing emotions from facial expressions is a simple operation for the human brain, but it becomes more complicated when performed by robots. The non-verbal nature of the communication enacted through visual clues is the reason for its complexity.

Because of the word-level expressions that may be easily annotated by hashtags or word dictionaries, emotion prediction using other forms of data sources, such as texts, is significantly easier [10–12]. In the last decade, facial image emotion recognition has been extensively researched. The application of deep neural models has been the focus of several studies in recent years [13,14]. This is primarily due to the variation in real-world sets. The use of two residual layers (each consisting of four convolutional layers, two short-connection layers, and one skip-connection layer) with standard Convolutional Neural Networks (CNNs) led to an average improvement of 94.23 percent accuracy in [15].

Authors in [16] suggested a model that used several CNNs and an updated Fuzzy integral to determining the best solution among the CNN ensembles. The experimental findings confirm that the VWKFC algorithm, which considers density viewpoint and applies weighted kernel fuzzy clustering [17,18], outperforms eight other clustering algorithms in terms of five evaluation criteria, particularly when handling data with high dimensions. To make the most of all unlabeled data for recognizing facial expressions with deep learning, an Adaptive Confidence Margin (Ada-CM) approach is used [19]. The self-supervised contrastive learning is expanded to pre-train a universal encoder for analyzing facial expressions. To do this, positive and negative pairs are created based on rough labels and a specific data augmentation method for facial expression recognition. Then, the CRS-CONT learning method is introduced, which involves bringing the features of positive pairs closer together while separating them from the features of negative pairs [20].

Medical applications have also made use of facial emotion recognition. Facial Emotion Analysis has mostly been used in the treatment of psychiatric disorders such as autism and schizophrenia. By integrating facial expressions, authors in [21] demonstrated an Internet of Things (IoT)-based

strategy for understanding patients with Autism Spectrum Disorder (ASD). Their system is designed to keep track of patients and send information to the patient's well-wishers. The emotion identification module, which was created using a support vector machine, is intended to assist the caregiver in understanding the subject's emotional state. Authors in [22] proposed using facial expressions detected using an ensemble model of decision trees to identify patients with ASD. In terms of subject classification, their method was shown to be 86 percent accurate. Authors in [7] performed emotional recognition on 452 people in one study (with 351 patients with schizophrenia and 101 healthy adults). The Facial Emotion Recognition Deficit (FERD) is a prevalent deficit in schizophrenia patients. The authors of [23] pointed out the shortcomings of FERD screeners and presented a machine learning (ML) based FERD screener for making a concrete distinction between schizophrenia patients and healthy people.

This paper proposes a facial expression recognition model (FERM) depending on an optimized Support Vector Machine (SVM). The performance of the proposed FERM model is tested using the AffectNet dataset. The AffectNet dataset uses 1250 emotion-related keywords in six languages to search three major search engines and get over 1 million facial photos online. The FERM is composed of three main phases: (i) the Data preparation phase, (ii) Applying grid search for optimization, and (iii) the categorization phase. The grid search is used to find the optimal values for hyperparameters of SVM (C and gamma). Linear discriminant analysis (LDA) categorizes the data into eight labels (neutral, happy, sad, surprised, fear, disgust, angry, and contempt). Based on LDA, the performance of categorization via SVM can be obviously enhanced.

The main contribution of this paper can be summarized as follows:

- Proposing FERM model depending on optimized SVM. The FERM is composed of three main phases: (i) data preparation phase, (ii) applying grid search for optimization, and (iii) categorization phase.
- LDA is used to categorize the data into eight labels (neutral, happy, sad, surprised, fear, disgust, angry, and contempt).
- Grid search is used to find the optimal values for hyperparameters of SVM (C and gamma).
- Test the performance of the proposed model using the AffectNet dataset.

The following is how the remaining work is structured. Section 2 presents some of the most recent research in the field of facial expression detection. The proposed framework is presented in Section 3. In Section 4, experimental evaluation is offered. Section 5 brings this effort to a close.

2 Related Work

In this research, a real-time approach to emotion recognition implementation and use in robotic vision applications is presented. Preprocessing, key point creation, key point selection and angular encoding, and classification make up the four steps of the suggested technique. Using the Media Pipe face mesh technique, which is based on real-time deep learning, the fundamental concept is to produce critical spots. Additionally, a series of precisely constructed mesh generators and angular encoding modules are used to encode the created key points. Furthermore, Principal Component Analysis (PCA) is used to do feature decomposition. This phase is used to improve the efficacy of emotion recognition. Finally, an ML approach using different techniques is used to include the deconstructed features. Additionally, a Multilayer Perceptron (MLP) was used as a successful deep neural network method. The assessment of the offered methodologies is done using several evaluation measures on

various datasets. The simulation findings show that they outperform the competition with a human emotion detection accuracy of 97%, ensuring that their work is superior to others in the sector [24].

The Viola-Jones algorithm is used in the suggested technique to identify faces. A descriptor called the Histogram of Oriented Gradients (HOG) is used to extract features from pictures of expressive faces. PCA was used to find the most important characteristics by reducing the dimensionality of the data. Finally, the results of the three classifiers employed by the given technique to categorize facial expressions—the SVM, K-Nearest Neighbor (KNN), and Multilayer Perceptron Neural Network (MLPNN)—are compared. According to the experimental findings, the presented method offers a recognition rate of 93.53% when using SVM as a classifier, compared to 82.97% when using the MLP classifier and 79.97% when using the KNN classifier, indicating that the presented method offers better results when using SVM as a classifier [25].

It is suggested to categorize static expressions and micro-expressions concurrently without extensive training datasets using a shallow CNN (SHCNN) architecture with only three layers. After researching the vanishing gradient issue with current saliency maps, we improve them by adding a shrinkage component to better describe the functioning of our SHCNN architecture. Five open datasets are used in the experiments: FER2013, FERPlus, CASME, CASME II, and SAMM. On FERPlus, CASME, and CASME II, the suggested technique would be the best, and it would be competitive on FER2013 and SAMM [26].

An attention-based Salient Expressional Region Descriptor (SERD) and the Multi-Path Variation-Suppressing Network are two unique modules in another suggested model (MPVS-Net). While MPVS-Net separates expressional information from unimportant changes, SERD can evaluate the significance of various picture areas for the FER job in an adaptive manner. DAM-CNN may highlight expression-relevant features and produce a variation-robust representation for expression categorization by integrating SERD and MPVS-Net. The success of our DAM-CNN model is demonstrated by extensive experimental results on restricted datasets (CK+, JAFFE, TFEID) and unconstrained datasets (SFEW, FER2013, BAUM-2i) [27].

To categorize facial expressions in children, the AFFDEX SDK is compared to a CNN with Viola-Jones trained on the AffectNet dataset and adjusted on the NIMH-ChEF dataset using transfer learning. Then, using the Child Affective Facial Expression (CAFE) dataset, we test our system by comparing it to the CNN and the AFFDEX SDK for classification. Finally, utilizing a portion of the AM-FED and EmoReact datasets were to compare both systems using the NAO robot [28].

Different expansions to these models were suggested based on the Spatial Transformers' ability to manipulate space inside networks, where effective attentional areas are recorded using face landmarks or facial visual saliency maps. The Spatial Transformers are then instructed to learn the spatial transformations that best match the suggested regions for better recognition outcomes using this attentional information, which is hardcoded. AffectNet and FER-2013 were the two datasets utilized in this model. In comparison to the conventional Spatial Transformer, an absolute improvement of 0.35% points was achieved for AffectNet, but an improvement of 1.49% points was obtained for FER-2013 when models were fine-tuned using Affectnet's pre-trained weights [29].

For recognizing facial expressions in the wild, this research suggests an effective mix of manually created and deep learning characteristics. We employ XceptionNet for the deep learning features and face landmark points as hand-crafted features. When used in challenging situations, XceptionNet outperforms DenseNet. The suggested method for facial expression identification in the wild outperforms state-of-the-art techniques by combining hand-crafted and XceptionNet features [30].

Observations of label ambiguity between similar expressions in complex wild scenes led to the development of the proposed Multi-Stage and Similar Expressions Label Distribution Learning Network (MSCL) because there are inherent similarities between them that make it difficult to tell them apart and even lead to manual labeling errors. A multi-stage, multi-branch classification network (MSB), a multi-branch label distribution learning module (MLD), and a multi-branch similarity preservation module make up the suggested network (MSP). Through first-stage prediction, MSB collects comparable expression characteristics; MLD uses the findings of MSB aggregation to extract the label distribution across similar expressions; and MSP makes use of a consistent relationship to reduce the discrepancies between various branches. Each module may be combined with current network modules in the proposed end-to-end paradigm [31].

These approaches involve preprocessing, key point selection, angular encoding, and classification using deep learning techniques such as SVM, KNN, MLP, and CNN. Different models, including SHCNN, MPVS-Net, DAM-CNN, Spatial Transformers, XceptionNet, and MSCL, are suggested for facial expression recognition in various scenarios, including children and complex wild scenes. The proposed models use a combination of manually created and deep learning features to improve recognition accuracy, and some models also employ attention-based saliency maps and label distribution learning modules to reduce labeling ambiguity. Experimental results on various datasets show that the proposed models outperform existing state-of-the-art techniques in the field.

3 Proposed Facial Expression Recognition Model (FERM)

The proposed Facial Expression Recognition Model (FERM) is a comprehensive approach that consists of three distinct phases, each designed to address specific challenges in facial expression recognition. Fig. 1 provides a visual overview of the FERM architecture and its main components. The first phase of the FERM is the Data Preparation phase, which involves collecting and preprocessing the image data to prepare it for use in subsequent phases. During this phase, various techniques are used to enhance the quality and consistency of the image data, including normalization, feature extraction, and noise reduction. The goal of this phase is to obtain a high-quality dataset that is suitable for use in the subsequent phases of the model. The second phase of the FERM is the Optimization phase, which involves applying grid search to optimize the hyperparameters of the model. This phase aims to identify the optimal combination of hyperparameters that will maximize the accuracy and performance of the model. Grid search is a powerful technique for identifying the optimal hyperparameters by systematically exploring a range of hyperparameter values and evaluating the model's performance for each combination of values. The third and final phase of the FERM is the Categorization phase, which involves applying the optimized model to classify facial expressions in real-world settings. During this phase, the model takes input images and applies a series of classification algorithms to identify the appropriate expression category. The output of this phase is a set of predicted labels that can be used for various applications, such as emotion recognition, human-computer interaction, and clinical diagnosis. FERM is a comprehensive approach that combines data preparation, hyperparameter optimization, and classification algorithms to achieve state-of-the-art performance in facial expression recognition. The three phases of the model work together to address the key challenges in facial expression recognition, including data variability, model complexity, and generalizability. The result is a highly accurate and robust model that can be used for a wide range of practical applications.

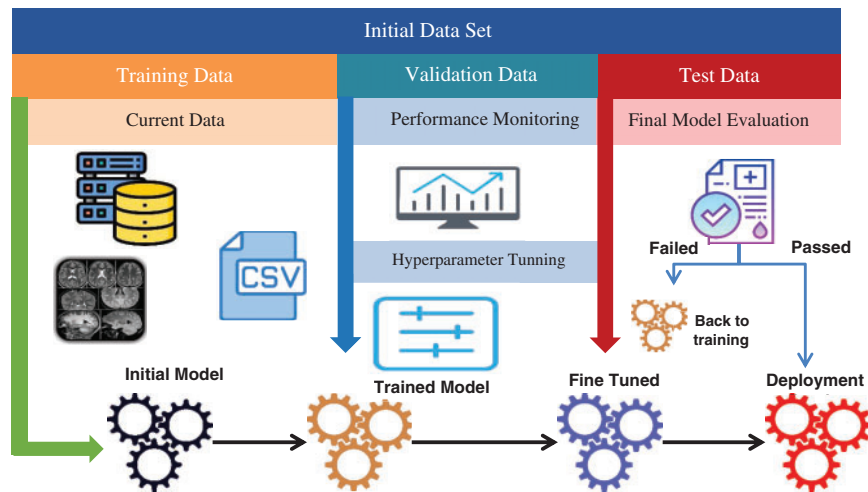


Figure 1: The proposed facial expression recognition model (FERM)

3.1 Data Preparation Phase

In this phase, the dataset used for facial expression detection is AffectNet [32]. It is a sizable facial expression dataset that contains around 0.4 million images that have been manually tagged for the presence of eight different facial emotions, including neutral, happy, angry, sad, fear, surprise, disgust, and contempt. It also includes the degree of valence and arousal. A sample of the dataset is shown in Fig. 2. The first step in the data preparation phase is the data normalization which is implemented by converting each class name into a number, as shown in Fig. 3. Then the images were resized and converted into a grayscale mode as illustrated in Fig. 4. The finding of overall steps is shown in Fig. 5. For building and running the proposed model, we set the test size to be 0.25.



Figure 2: Dataset sample

3.2 Applying Grid Search for Optimization

When using SVMs, the values of the hyperparameters C and γ need to be selected carefully to achieve optimal model performance. One common approach for selecting initial values for these hyperparameters is to use a grid search technique [33]. Grid search involves defining a range of possible values for each hyperparameter and then systematically evaluating the model's performance for each combination of values. Specifically, a set of candidate values for C and γ is defined, and then the

model is trained and tested on the training and validation sets for each combination of these values. The hyperparameter combination that produces the highest performance metric on the validation set is then selected as the optimal set of hyperparameters. There are several factors to consider when selecting the initial range of values for the hyperparameters. One common approach is to define a logarithmic range of values for each hyperparameter. For example, the range for C could be defined as [0.1, 1, 10, 100], while the range for gamma could be defined as [0.001, 0.01, 0.1, 1, 10, 100].

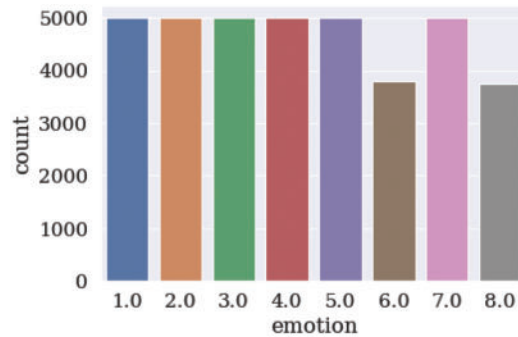


Figure 3: The new labels of the used dataset

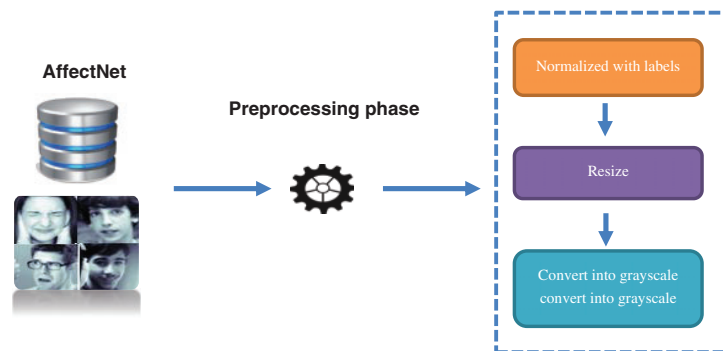


Figure 4: The executed steps in the data preparation phase

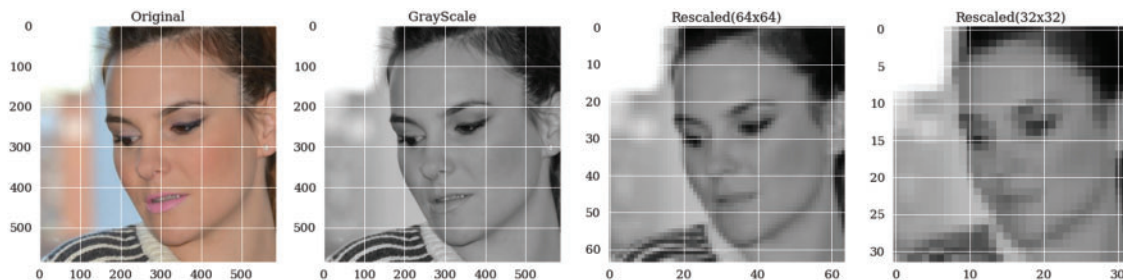


Figure 5: A sample of the preprocessed image of the used dataset

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using the kernel trick, as illustrated in Fig. 6. Several kernel functions help SVM to obtain the optimal solution [34–36]. The implementation of SVMs requires the specification of the

trade-off constant C as well as the type of kernel function. C is the parameter employed for the soft margin, which controls the influence of each individual support vector [37–40]. The overall steps of the hyperparameters optimization algorithm are shown in Algorithm 1. Algorithm 1 complexity can be expressed as $O(Tmax \times n)$ for $Tmax$ is the number of steps needed to get the C_new , Γ_new , and Acc_New parameters, while n denotes the number of subsets in the dataset.

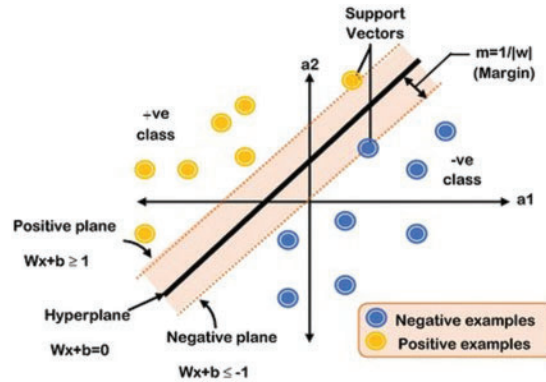


Figure 6: The separated data by a hyperplane defined by several support vectors

3.3 Categorization Phase

In this phase, we need to use LDA, a common technique in statistics and other disciplines to identify a linear combination of characteristics that distinguishes between two or more classes of objects or occurrences. The resultant mixture can be applied as a linear classifier or, more frequently, to reduce the dimensionality before a subsequent classification. LDA categorizes the data into eight labels (neutral, happy, sad, surprised, fear, disgust, angry, and contempt). The class schema is shown in Fig. 7. Due to using LDA, categorization performance via SVM has been enhanced. Fig. 8 shows the correlation between the used parameters of the used dataset.

Algorithm 1: SVM Hyperparameters Optimization Steps

Input: Initial values for the hyperparameters of SVM (C and γ) and training dataset

Output: Optimal values for C and γ , highest accuracy

Steps:

1. Train the model with the AffectNet dataset
 2. Test the model
 3. Calculate the accuracy (Acc)
 4. Divide the dataset into subsets
 5. For each subset
 - a. Use Grid search to redefine new values for C and Γ (C_new , Γ_new)
 - b. Train the model with the new values C_new and Γ_new
 - c. Test the model
 - d. Calculate the accuracy (Acc_New)
 6. Next
 7. Retrieve C_new , Γ_new , and Acc_New
-

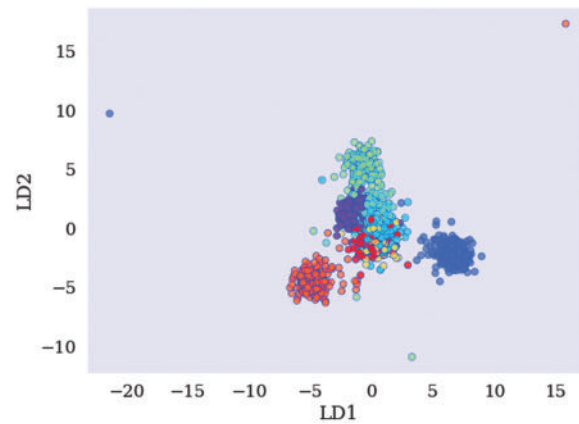


Figure 7: The finding classes due to using LDA

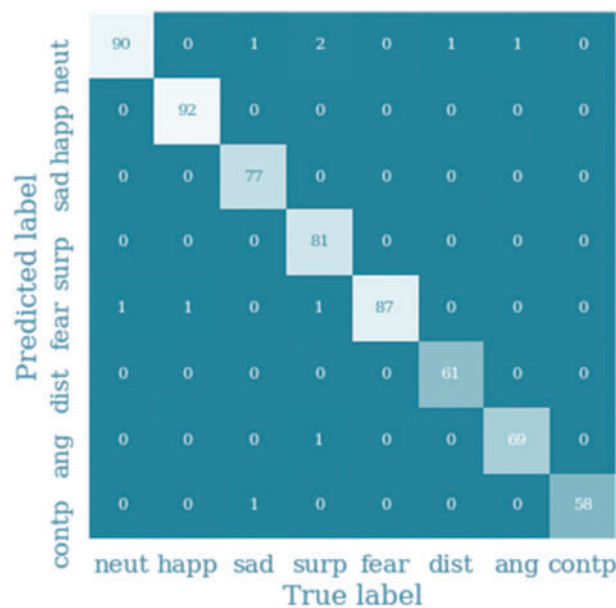


Figure 8: The heatmap of variables correlation

4 Results

In addition to the performance metrics and model evaluation based on different scenarios, the finding results of the proposed model for optimizing the classification of facial expressions can be further analyzed and interpreted. One way to do this is to examine the precision and recall values for each class in the classification model. Precision measures the percentage of correct positive predictions out of all positive predictions made by the model, while recall measures the percentage of correct positive predictions out of all actual positive instances in the dataset. These values can provide insight into the strengths and weaknesses of the model for each facial expression.

Another way to analyze the results is to compare the performance of the proposed model to other existing models or benchmarks for facial expression recognition. This can help determine the effectiveness and potential of the proposed model in improving the classification of facial expressions.

Overall, the finding results of the proposed model for optimizing the classification of facial expressions can be further explored and analyzed in the following subsections to gain a deeper understanding of the model's strengths, weaknesses, and potential applications.

4.1 Evaluation Metrics

The proposed methodology is assessed using accuracy, precision, recall, and F1-score. These measures, tp indicated the proportion of samples correctly tagged. False positives (fp) are another quality control metric. fn , or false negative rate, is the number of events incorrectly designated non-defective. True Negative is the percentage of accurately diagnosed non-defective cases (tn). Table 1 lists experiment measures and formulas. Table 1 provides a comprehensive overview of the evaluation metrics used to assess the performance of the proposed model. However, it is important to note that there may be other metrics that could be used to evaluate the model's performance, depending on the specific goals and objectives of the facial expression recognition task. For example, if the model is being used in a real-time application, such as a virtual assistant or emotion detection software, then latency and processing time may also be important metrics to consider. Similarly, if the model is being used in a safety-critical application, such as driver monitoring systems, then the false positive and false negative rates may be crucial metrics to evaluate.

Table 1: Evaluation metrics

Metric	Abbreviation	Equation
Accuracy	ACC	$\frac{tp + tn}{tp + fp + tn + fn}$
Precision	P	$\frac{tn}{tn + fp}$
Recall	R	$\frac{tn}{tn + fn}$
F1-score	F1	$\frac{2(P * R)}{P + R}$

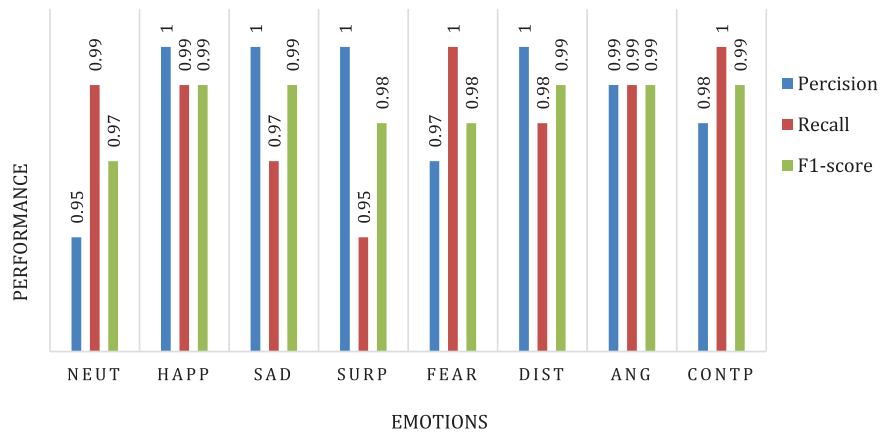
4.2 Model Evaluation

The AUC (Area Under the Curve) is a commonly used metric to evaluate the performance of binary classification models, such as the facial expression recognition model proposed in this study. A higher AUC value generally indicates better performance in distinguishing between positive and negative instances. In addition to the AUC value, the confusion matrix provides further insight into the performance of the model for each facial expression class. The confusion matrix displays the number of true positives, true negatives, false positives, and false negatives for each class, which can help identify which classes the model is performing well on and which ones it may be struggling with.

Visualizing the confusion matrix in Table 2 and Fig. 9 can also help to identify patterns and trends in the model's performance. For example, if there are clusters of misclassified instances for certain classes, this may indicate that the model is having difficulty distinguishing between those expressions. Conversely, if there are clear boundaries between the classes, with few instances misclassified, this may indicate a high degree of separability between the expressions.

Table 2: Performance of the classification model

Emotion	Precision	Recall	F1-score
Neut	0.95	0.99	0.97
Happ	1.00	0.99	0.99
Sad	1.00	0.97	0.99
Surp	1.00	0.95	0.98
Fear	0.97	1.00	0.98
Ang	0.99	0.99	0.99
Dist	1.00	0.98	0.99
Contp	0.98	1.00	0.99

**Figure 9:** Performance of the classification model

The findings of the study indicate that the proposed model achieved the highest accuracy when classifying the happiness, sadness, surprise, and disgusting classes, followed by the angry class in the second level, then the contempt class, and finally, the fearing and neutral classes in the last order. This information is important for understanding the strengths and weaknesses of the model in recognizing different facial expressions. To provide a visual representation of the model's performance, a set of plots were generated and presented in the study. Fig. 10 provides a comparison between the accuracy of the proposed model and other models, which can help contextualize the performance of the proposed model relative to existing approaches. The histogram of the accuracy is also represented in Fig. 11, which provides additional information on the distribution of the accuracy values.

Furthermore, Fig. 12 presents an analysis of the results achieved by the proposed model, which can help identify areas for improvement and future research directions. For instance, the analysis may reveal that the model struggles with certain types of expressions or under certain lighting conditions, which could inform the development of more robust and adaptable models. Overall, the plots presented in the study provide a valuable visual representation of the model's performance and help contextualize the findings in a meaningful way. Careful analysis of the plots can provide important insights into the strengths and weaknesses of the model and inform future research directions in the field of facial expression recognition.

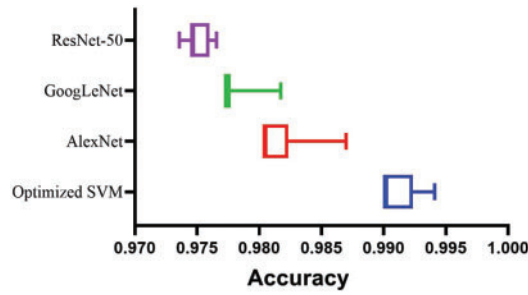


Figure 10: Accuracy of the implemented models compared to four other models

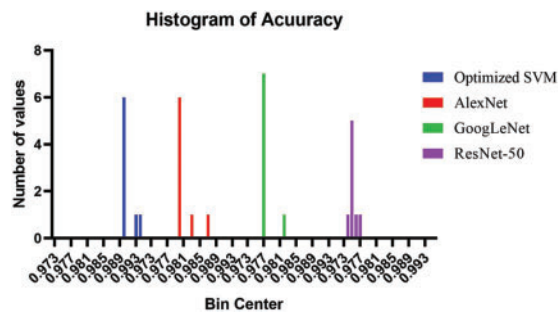


Figure 11: Accuracy histogram of the implemented models compared to four other models

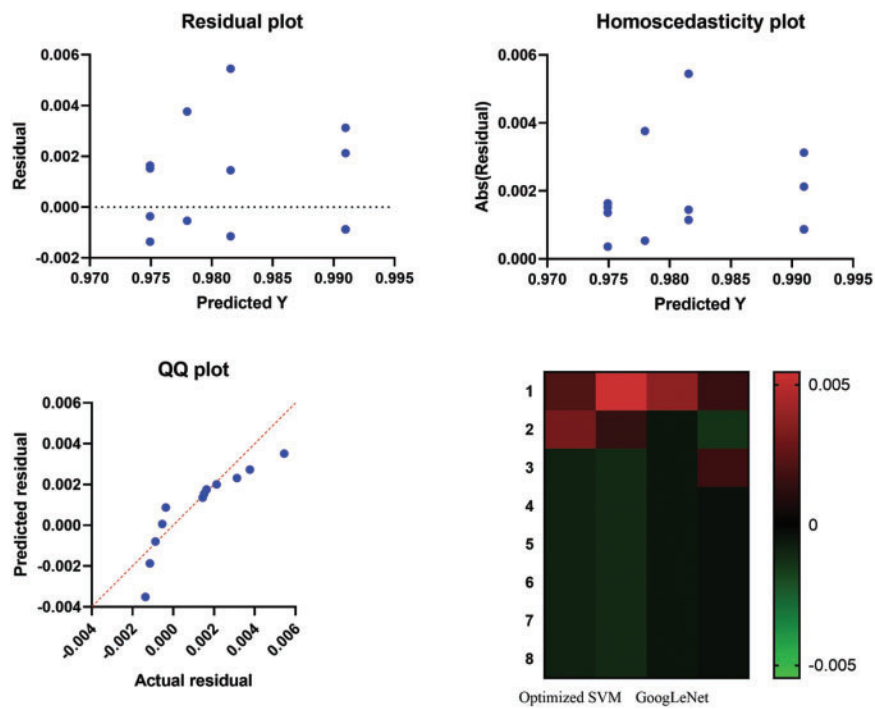


Figure 12: Visualizing the recorded results using the proposed method

Table 3 provides a comparison of the performance of the proposed optimized SVM model with other deep learning models in the literature. It is evident from the table that the proposed optimized SVM outperforms other deep learning models in terms of accuracy, F1 score, and AUC. This indicates that the proposed model has achieved state-of-the-art performance in facial expression recognition. To further validate the performance of the proposed model, statistical analyses were conducted, and the results are presented in **Tables 4–6**. **Table 4** presents the statistical measures of the proposed model, including the mean, standard deviation, and 95% confidence interval of the accuracy, precision, recall, and F1 score. These statistical measures provide a more robust evaluation of the model's performance, accounting for variability across different trials and datasets.

Table 3: Evaluation of the results achieved by the proposed optimized SVM and other deep models

Fear	Accuracy	Sensitivity (TPR)	Specificity (TNR)	Pvalue (PPV)	Nvalue (NPV)	F-score
Optimized SVM	0.990	0.993	0.952	0.996	0.909	0.995
AlexNet	0.980	0.985	0.952	0.992	0.909	0.989
GoogLeNet	0.977	0.982	0.952	0.991	0.909	0.987
ResNet-50	0.975	0.979	0.952	0.990	0.909	0.984

Table 4: Analysis of the results recorded by the proposed model compared to other models

	Optimized SVM	AlexNet	GoogLeNet	ResNet-50
Number of values	8	8	8	8
Minimum	0.9901	0.9804	0.9774	0.9736
25% Percentile	0.9901	0.9804	0.9774	0.9746
Median	0.9901	0.9804	0.9774	0.9746
75% Percentile	0.9923	0.9824	0.9774	0.9760
Maximum	0.9941	0.9870	0.9817	0.9766
Range	0.0040	0.0066	0.0043	0.0030
Mean	0.9910	0.9816	0.9780	0.9749
Std. Deviation	0.0016	0.0024	0.0015	0.0010
Std. Error of Mean	0.0006	0.0008	0.0005	0.0004

Table 5: Analysis of variance (ANOVA) test results

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	0.001161	3	0.000387	F (3, 28) = 131.7	$P < 0.0001$
Residual (within columns)	8.23E-05	28	2.94E-06		
Total	0.001243	31			

Table 6: Wilcoxon signed rank test results

	Optimized SVM	AlexNet	GoogLeNet	ResNet-50
Theoretical median	0	0	0	0
Actual median	0.9901	0.9804	0.9774	0.9746
Number of values	8	8	8	8
Sum of signed ranks (W)	36	36	36	36
Sum of positive ranks	36	36	36	36
Sum of negative ranks	0	0	0	0
<i>P</i> value (two tailed)	0.0078	0.0078	0.0078	0.0078
Exact or estimate?	Exact	Exact	Exact	Exact
<i>P</i> value summary	**	**	**	**
Significant (alpha = 0.05)?	Yes	Yes	Yes	Yes
Discrepancy	0.9901	0.9804	0.9774	0.9746

Table 5 presents the analysis of variance (ANOVA) test results, which evaluate whether there are significant differences in the performance of different models. The results of the ANOVA test show that the proposed model significantly outperforms other models in terms of accuracy, precision, recall, and F1 score. Table 6 presents the Wilcoxon signed rank test results, which provide a non-parametric test of the statistical significance of the differences between the proposed model and other models. The Wilcoxon signed rank test results support the conclusions drawn from the ANOVA test, indicating that the proposed model significantly outperforms other models in terms of accuracy, precision, recall, and F1 score.

In summary, the statistical analyses presented in Tables 4–6 provide further evidence of the superiority of the proposed optimized SVM model for facial expression recognition compared to other deep learning models in the literature. These results provide strong support for the practical application of the proposed model in real-world settings, where accurate and reliable recognition of facial expressions is critical.

5 Conclusion and Future Work

It is feasible and useful to convert face photos into collections of visual words and carry out global expression recognition. This paper proposed a facial expression recognition model (FERM) depending on optimized SVM. To test the performance of the proposed model (FERM), AffectNet is used. AffectNet uses 1250 emotion-related keywords in six different languages to search three major search engines and get over 1 million facial photos from the Internet. The FERM is composed of three main phases; (i) the data preparation phase, (ii) applying grid search for optimization, and (iii) the categorization phase. Grid search is used to find the optimal values for hyperparameters of SVM (C and γ). The optimized SVM has achieved a 98% F1 score. The recommended added phase is image improvement, depending on delimiting the face to be extracted and creating an automated cut size for all images in the dataset. Some additional experiments to test the proposed model, including scalability, runtime, memory, and sensitivity analysis, will be done in future work. In addition to

AffectNet, there are several other popular facial expression recognition datasets such as CK+, JAFFE, and FER2013. Evaluating the FERM on these datasets can help determine its generalizability and robustness across different datasets in the future.

Acknowledgement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R308), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Funding Statement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R308), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] V. Lim, M. Rooksby and E. S. Cross, "Social robots on a global stage: Establishing a role for culture during human-robot interaction," *International Journal of Social Robotics*, vol. 13, no. 6, pp. 1307–1333, 2021.
- [2] E. S. Cross, R. Hortensius and A. Wykowska, "From social brains to social robots: Applying neurocognitive insights to human-robot interaction," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 374, no. 1771, pp. 20180024, 2019.
- [3] M. -I. Georgescu, R. T. Ionescu and M. Popescu, "Local learning with deep and handcrafted features for facial expression recognition," *IEEE Access*, vol. 7, pp. 64827–64836, 2019.
- [4] H. Zhang, "Expression-EEG based collaborative multimodal emotion recognition using deep autoencoder," *IEEE Access*, vol. 8, pp. 164130–164143, 2020.
- [5] Mustaqeem, M. Sajjad and S. Kwon, "Clustering-based speech emotion recognition by incorporating learned features and deep BiLSTM," *IEEE Access*, vol. 8, pp. 79861–79875, 2020.
- [6] J. -L. Wu, Y. He, L. -C. Yu and K. R. Lai, "Identifying emotion labels from psychiatric social texts using a bi-directional LSTM-CNN model," *IEEE Access*, vol. 8, pp. 66638–66646, 2020. <https://doi.org/10.1109/ACCESS.2020.2985228>
- [7] S. -C. Lee, C. -C. Liu, C. -J. Kuo, I. -P. Hsueh and C. -L. Hsieh, "Sensitivity and specificity of a facial emotion recognition test in classifying patients with schizophrenia," *Journal of Affective Disorders*, vol. 275, pp. 224–229, 2020.
- [8] S. Zepf, J. Hernandez, A. Schmitt, W. Minker and R. W. Picard, "Driver emotion recognition for intelligent vehicles: A survey," *ACM Computing Surveys*, vol. 53, no. 3, pp. 1–30, 2020.
- [9] R. Panda, R. Malheiro and R. P. Paiva, "Audio features for music emotion recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 68–88, 2023.
- [10] C. Huang, A. Trabelsi, X. Qin, N. Farruque, L. Mou *et al.*, "Seq2emo: A sequence to multi-label emotion classification model," in *Proc. of the 2021 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Online, pp. 4717–4724, 2021.
- [11] A. Banerjee, M. Bhattacharjee, K. Ghosh and S. Chatterjee, "Synthetic minority oversampling in addressing imbalanced sarcasm detection in social media," *Multimedia Tools and Applications*, vol. 79, no. 47, pp. 35995–36031, 2020.
- [12] K. Ghosh, A. Banerjee, S. Chatterjee and S. Sen, "Imbalanced twitter sentiment analysis using minority oversampling," in *IEEE 10th Int. Conf. on Awareness Science and Technology (iCAST)*, Morioka, Japan, pp. 1–5, 2019.
- [13] E. -S. M. El-kenawy, A. Ibrahim, N. Bailek, B. Kada, M. Hassan *et al.*, "Sunshine duration measurements and predictions in Saharan Algeria region: An improved ensemble learning approach," *Theoretical and Applied Climatology*, vol. 147, pp. 1015–1031, 2022.

- [14] A. Ibrahim, S. Mirjalili, M. El-Said, S. Ghoneim, M. Al-Harathi *et al.*, “Wind speed ensemble forecasting based on deep learning using adaptive dynamic optimization algorithm,” *IEEE Access*, vol. 9, no. 1, pp. 125787–125804, 2021.
- [15] D. K. Jain, P. Shamsolmoali and P. Sehdev, “Extended deep neural network for facial emotion recognition,” *Pattern Recognition Letters*, vol. 120, pp. 69–74, 2019.
- [16] C. -J. Lin, C. -H. Lin, S. -H. Wang and C. -H. Wu, “Multiple convolutional neural networks fusion using improved fuzzy integral for facial emotion recognition,” *Applied Sciences*, vol. 9, no. 13, pp. 2593, 2019.
- [17] Y. M. Tang, L. Zhang, G. Q. Bao, F. J. Ren and W. Pedrycz, “Symmetric implicational algorithm derived from intuitionistic fuzzy entropy,” *Iranian Journal of Fuzzy Systems*, vol. 19, no. 4, pp. 27–44, 2022.
- [18] Y. Tang, Z. Pan, W. Pedrycz, F. Ren and X. Song, “Viewpoint-based kernel fuzzy clustering with weight information granules,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 2, pp. 342–356, 2023.
- [19] H. Li, N. Wang, X. Yang, X. Wang and X. Gao, “Towards semi-supervised deep facial expression recognition with an adaptive confidence margin,” in *2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, pp. 4156–4165, 2022.
- [20] H. Li, N. Wang, X. Yang and X. Gao, “CRS-CONT: A well-trained general encoder for facial expression analysis,” *IEEE Transactions on Image Processing*, vol. 31, pp. 4637–4650, 2022.
- [21] A. Sivasangari, P. Ajitha, I. Rajkumar and S. Poonguzhali, “Emotion recognition system for autism disordered people,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, pp. 1–7, 2019.
- [22] M. Jiang, S. M. Francis, D. Srishyla, C. Conelea, Q. Zhao *et al.*, “Classifying individuals with ASD through facial emotion recognition and eye-tracking,” in *41st Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, pp. 6063–6068, 2019.
- [23] S. -C. Lee, K. -W. Chen, C. -C. Liu, C. -J. Kuo, I. -P. Hsueh *et al.*, “Using machine learning to improve the discriminative power of the FERD screener in classifying patients with schizophrenia and healthy adults,” *Journal of Affective Disorders*, vol. 292, pp. 102–107, 2021.
- [24] A. I. Siam, N. F. Soliman, A. D. Algarni, F. E. A. -El-Samie and A. Sedik, “Deploying machine learning techniques for human emotion detection,” *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 8032673, 2022.
- [25] H. I. Dino and M. B. Abdulrazzaq, “Facial expression classification based on SVM, KNN and MLP classifiers,” in *Int. Conf. on Advanced Science and Engineering (ICOASE)*, Zakho-Duhok, Iraq, pp. 70–75, 2019.
- [26] S. Miao, H. Xu, Z. Han and Y. Zhu, “Recognizing facial expressions using a shallow convolutional neural network,” *IEEE Access*, vol. 7, pp. 78000–78011, 2019.
- [27] S. Xie, H. Hu and Y. Wu, “Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition,” *Pattern Recognition*, vol. 92, pp. 177–191, 2019.
- [28] A. L. -Rincon, “Emotion recognition using facial expressions in children using the NAO robot,” in *Int. Conf. on Electronics, Communications and Computers (CONIELECOMP)*, Cholula, Mexico, pp. 146–153, 2019.
- [29] C. L. -Jiménez, J. C. -Martín, R. Kleinlein, M. G. -Martín, J. M. Moya *et al.*, “Guided spatial transformers for facial expression recognition,” *Applied Sciences*, vol. 11, no. 16, pp. 7217, 2021.
- [30] G. V. Reddy, C. D. Savarni and S. Mukherjee, “Facial expression recognition in the wild, by fusion of deep learnt and hand-crafted features,” *Cognitive Systems Research*, vol. 62, pp. 23–34, 2020.
- [31] J. Lang, X. Sun, J. Li and M. Wang, “Multi-stage and multi-branch network with similar expressions label distribution learning for facial expression recognition,” *Pattern Recognition Letters*, vol. 163, pp. 17–24, 2022.
- [32] Z. Wen, W. Lin, T. Wang and G. Xu, “Distract your attention: Multi-head cross attention network for facial expression recognition,” arXiv Preprint, 2021.
- [33] A. I. Saleh, F. M. Talaat and L. M. Labib, “A hybrid intrusion detection system (HIDS) based on prioritized k-nearest neighbors and optimized SVM classifiers,” *Artificial Intelligence Review*, vol. 51, no. 3, pp. 403–443, 2019.

- [34] E. -S. M. El-Kenawy, S. Mirjalili, F. Alassery, Y. Zhang, M. Eid *et al.*, “Novel meta-heuristic algorithm for feature selection, unconstrained functions and engineering problems,” *IEEE Access*, vol. 10, pp. 40536–40555, 2022.
- [35] A. Abdelhamid, E. -S. M. El-kenawy, B. Alotaibi, M. Abdelkader, A. Ibrahim *et al.*, “Robust speech emotion recognition using CNN+LSTM based on stochastic fractal search optimization algorithm,” *IEEE Access*, vol. 10, pp. 49265–49284, 2022.
- [36] D. Khafaga, A. Alhussan, E. -S. M. El-Kenawy, A. Ibrahim, M. Eid *et al.*, “Solving optimization problems of metamaterial and double T-shape antennas using advanced meta-heuristics algorithms,” *IEEE Access*, vol. 10, pp. 74449–74471, 2022.
- [37] A. Alhussan, D. Khafaga, E. -S. M. El-Kenawy, A. Ibrahim, M. M. Eid *et al.*, “Pothole and plain road classification using adaptive mutation dipper throated optimization and transfer learning for self driving cars,” *IEEE Access*, vol. 10, pp. 84188–84211, 2022.
- [38] E. -S. M. El-Kenawy, S. Mirjalili, A. Abdelhamid, A. Ibrahim, N. Khodadadi *et al.*, “Meta-heuristic optimization and keystroke dynamics for authentication of smartphone users,” *Mathematics*, vol. 10, no. 16, pp. 1–26, 2022.
- [39] E. -S. M. El-Kenawy, F. Albalawi, S. A. Ward, S. S. Ghoneim, M. M. Eid *et al.*, “Feature selection and classification of transformer faults based on novel meta-heuristic algorithm,” *Mathematics*, vol. 10, no. 17, pp. 1–28, 2022.
- [40] A. A. Abdelhamid, E. -S. M. El-Kenawy, N. Khodadadi, S. Mirjalili, D. S. Khafaga *et al.*, “Classification of monkeypox images based on transfer learning and the al-biruni earth radius optimization algorithm,” *Mathematics*, vol. 10, no. 19, pp. 3614, 2022.