Iraqi Journal for Computer Science and Mathematics

Journal Homepage: http://journal.esj.edu.iq/index.php/IJCM
e-ISSN: 2788-7421 p-ISSN: 2958-0544

LEARNERS' EMOTIONS ESTIMATION USING VIDEO PROCESSING TECHNIQUES FOR OPTIMUM E-LEARNING EXPERIENCE

Mohammed Khaleel Hussein [1], Ali Abdullah Ali[2], Mohammed Ahmed Subhi*[3], Saleh Mahdi Mohammed [4]

[1]Department of Planning, Directorate of Private University Education, Ministry of Higher Education and Scientific Research, Baghdad, Iraq.
[2]Minister Office, Ministry of Higher Education and Scientific Research, Baghdad, Iraq.
[3]Department of Planning, Directorate of Private University Education, Ministry of Higher Education and Scientific Research, Baghdad, Iraq.
[4]Department of Computer Technology Engineering, Technical College, Imam Ja'afar Al-Sadiq University, Baghdad, Iraq.

*Corresponding Author: Mohammed Ahmed Subhi

DOI: https://doi.org/10.52866/ijcsm.2024.05.03.038
Received April 2024; Accepted June 2024; Available online August 2024

ABSTRACT: Learning management systems (LMSs) have integrated multiple technologies to enhance the e-learning experience. One such technology is the emotional recognition system (ERS), which provides tutors with data on learners' emotions, including anger, sadness, happiness, and more. ERS utilizes various data sources like facial expressions, body activities, and brain signals to recognize emotions. This paper provides an overview of the ERS structure and discusses the state-of-the-art technologies in this field. The results indicate that deep learning based ERS using VGG19 for feature extraction over the FER2013 dataset is reliable with a recognition accuracy of 87% using Random Forest Algorithm.

Keywords: Learning Management System (LMS), E-learning, Emotional Recognition System (ERS), Convolutional Neural Network (CNN), Intelligent Recognition System (IRS), Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Cepstral Coefficients (LPCC).

1. INTRODUCTION

Recent improvements have opened up prospects in human emotions and human-to-machine interface detection in contact centers, marketing, and healthcare [1], [2]. This strategy not only enhances technical capabilities but also aims to reduce educational inequities, particularly for rural residents [3]. Virtual learning environments are essential when schools are closed due to financial constraints and global health catastrophes [4], [5]. These platforms manage crises and improve education so learning never stops.

This digital transformation requires the finest educational experiences for instructors and students. Many methods have been studied to improve e-learning assessments. Some methods identified students who may need additional support and interventions to improve their academic performance by presenting a robust predictive model to categorize students into distinct performance groups [6]. Voice-based verification and avatar integration as visual aids aim to boost engagement, deter cheating, and verify student identities [7], [8], [9]. Gamification makes learning more effective and entertaining [10].

Biometric variables like EEG brain activity and physical activity patterns are also enhancing learning results. After all the advancements, efficient communication in big e-learning contexts remains a challenge [11]. Without physical cues like eye contact and facial expressions, understanding emotions is difficult. E-learning systems must traverse many emotional detection channels to allow teachers to tailor their techniques to students' emotions.

The virtual learning experience itself poses a challenge for educators to evaluate student performance where traditional methods such as eye contact, facial expression, and body language are absent. This elevates the need for evaluation tools such as emotion recognition systems (ERS) which might change e-learning. Despite the recent advancements in such systems, several research gaps are still to be addressed, including improved accuracy, reliability, and the diversity of biometric sensor data that can be explored [12]. This study examines new ways of understanding and interpreting sentiments and behaviors, enabling more compassionate and effective learning environments.

2. METHODOLOGY

This technique incorporates the processes of feature extraction which is a necessary process in machine learning applications that involves the transformation of raw data into a set of features or characteristics that can be useful to obtain a model by training, and evaluation of these data. VGG19 is a convolutional neural network (CNN) that stands for Visual Geometry Group, which is a group at the University of Oxford where this architecture was developed, it is chosen to be utilized in conjunction with image-based emotional identification as part of the additional processing technique. The first thing that is done is to acquire the Facial Emotion Recognition dataset (FER2013), which is made up of photographs of people's faces that are categorized into seven distinct emotional states and indicate the emotions that they are experiencing. This dataset is chosen for its significance in facial emotion recognition tasks, it is also the standard dataset for human emotion classification in many recent researches [13], [14]. To guarantee the reliability of the dataset, we have normalized the pixel values to a range of 0 to 1 and standardized the sizes of the photographs. In addition, an additional image rotation, flipping, and trimming of the training dataset in order to increase the variety of the dataset and improve the generalization of the model. The VGG19 architecture, which had been trained on a large number of photos, was what was used to retrieve the features. The convolutional base continues to exist even after the fully linked layers of VGG19 have been removed. Obtaining high-level information from processed images is accomplished through the utilization of the VGG19 network. This is accomplished by transforming the feature mappings of the final convolutional layer into a flat representation, which then results in the transformation of each picture into a feature vector. In order to thoroughly prepare the techniques section, it is necessary to train the suggested algorithms for photo identification by making use of the seven emotions contained in the datasets.

Support Vector Machine (SVM), Random Forest, k-nearest Neighbors (KNN), and Decision tree are some of the machine learning classifiers that are used for the classification. Next, 80%, 10%, and 20% of the dataset are taken out to make training, validation, and test sets. The hyperparameters of each classifier are tweaked to get the best performance by using cross-validation on the training dataset along with random or grid search techniques. After the hyperparameters have been optimized, the training set of data is used to train the models. A set of performance metrics were used to judge the performance, which are shown below. In the evaluation process, the right evaluation metrics are used, such as F1 score, accuracy, precision, and recall. A close study of the data looks for patterns or trends in a wide range of emotions and compares how well different classifiers work. The model that did the best on the validation set is chosen to be tested again on the test set to get more objective performance estimates and make sure the model can be used in other situations. Towards this goal, compare the effectiveness of the deep learning-based method with that of other deep learning architectures or more traditional methods that use VGG19 features. It talks about the pros and cons of each classifier and the things that affect how well they work, like class imbalance and dataset features. There are also comments about how easy it is to understand the models and how they might be used in the real world.

2.1 Dataset

Face emotion recognition is often performed using the FER2013 dataset. It has seven distinct moods identified from 35,887 grayscale photos. The Google research team produced the dataset, which is openly accessible for academic usage. The pictures in the FER2013 collection show a range of people with varied emotional facial expressions.

There are anger, scorn, fear, happiness, sadness, surprise, and indifference in the collection. Each picture has a label that says which of these seven emotions the person is feeling. The image resolution in the dataset is 48 by 48 pixels. Emotion detection models are taught and tested with poor resolution so that they can handle information as quickly as possible. A black picture shows the pixel level in a different channel. The FER2013 dataset, however, has several limitations, including the imbalance of the number of images in each class, the low resolution of its images which may cause facial features loss, and it has an extent of demographic bias towards certain ethnicities. The pictures could not, among other things, fairly represent a broad range of age ranges, ethnic and cultural groupings, or environmental circumstances. Moreover, the accuracy of the labels assigned to the dataset may change since various people interpret emotion labels in different ways.

The FER2013 dataset, sometimes referred to as "Facial Expression Recognition 2013," is a crucial resource for machine learning and computer vision applications that require the ability to classify and evaluate emotions based on facial expressions. Moreover, the FER2013 dataset possesses several noteworthy qualities, such as its capacity to address practical issues, its large size, its consistency, its variety, and its precise categorization. This dataset will be valuable for researchers in the field of technologies that detect and analyze changes in facial expressions. The advancement of technology that possesses greater intelligence and heightened sensitivity toward human emotions will also be streamlined.

2.2 System Implementation

In order to gather features from the FER2013 dataset, a VGG19 model is proposed as a feature extractor since this model has been trained with the ImageNet dataset which has over a million images with 1000 categories, this training enables the VGG19 to learn different features effectively. Training a deep learning model is basically by learning the

characteristics that are extracted by the convolutional base layers. Before being uploaded to VGG19, images are cut down to a resolution of 224 pixels by 224 pixels.
Pixel values are frequently standardized to [0, 1]. After processing images, the reduced VGG19 model is fed them. VGG19's convolutional layers discern edges, textures, and shapes from images using hierarchical features.
Flattening the final convolutional layer's feature mappings creates each image's feature vector. This vector shows VGG19's key visual features.
An ERS (Emotional Recognition System) is made up of three important steps: feature extraction, feature mapping, and data pre-processing. The method gives each item a recognition score that tells you how many of those items it was able to correctly identify. Deep learning and machine learning techniques can be used in a number of different ways to make automatic recognition work. As shown in [16], one way is to get features ready before using any classification system. Local Binary Pattern (LBP) [17] is one of the most common ways for extracting features from face images used for ERS.
Computer vision issues like image classification have extensively used VGG19 (deep convolutional neural network). The 19-layer design includes fully connected, pooling, and convolutional layers. VGG19's deep architecture and picture data extraction are famous. VGG19 can extract and categorize features from the FER2013 image dataset, which is good for emotion identification.

The following steps may be done to utilize the VGG19 model for emotional recognition using the FER2013 dataset:

- Dataset Preparation: Create training and testing sets using pre-processed FER2013 dataset. Typically, the model is trained on part of the dataset and evaluated on the rest.
- Preprocessing: The pictures in the FER2013 dataset may need to be preprocessed, which can include adding more data, making the data normal, and scaling the images. Normalization helps set the pixel values within a certain range, and shrinking makes sure that all images are the same size. Data enrichment techniques, such as flips, translations, and random rotations, can be used to make the training data more varied [18].
- The VGG19 model extracts features from pre-processed images. VGG19 convolutional layers extract relevant information from input photos [19]. A completely connected layer or the final convolutional layer produces the recovered features.
- Classification: A fully connected neural network, also known as a support vector machine (SVM), is an example of a classifier that makes use of the attributes that have been gathered as input in order to make predictions regarding the labels of emotions. The classifier is trained on the training set by making use of the recovered characteristics and the emotion labels that correspond to them.
- Evaluation: The trained model is evaluated on the testing set to determine its efficacy. F1 score, accuracy, precision, and recall are typical tests. These scores demonstrate how well the model categorizes emotions using FER2013 images.

By using VGG19 as a feature extractor and training a classifier on top of the extracted features, the emotional recognition system can leverage the powerful representation-learning capabilities of deep neural networks. This approach allows the model to automatically learn discriminative features from the FER2013 dataset and make predictions about the emotional state of individuals based on their facial expressions.

The accuracy of recognition varies depending on the dataset used, assuming similar data such as images are employed across different testing datasets. Public dataset repositories are commonly used to test automatic ERS, and each dataset can have a different impact on accuracy scores, as shown in Table 1. Commonly used algorithms for image-based emotional recognition include K-nearest neighbor (KNN) [31] and support vector machine (SVM) [32].

After the characteristics of the VGG19 convolutional neural network (CNN) have been extracted from the FER2013 dataset, it is possible to train an SVM classifier, as demonstrated in the following image. To begin, a feature matrix is constructed using the features that have been deleted. In this matrix, each row represents an image, and each column represents a feature that has been eliminated from VGG19. These matrices represent the training data for the SVM classifier. It also gives each picture a label with the right emotion category from the FER2013 dataset. Next, the feature matrix and its labels were used to make training and validation sets. Usually, 80% of the sets were used for training and 20% were used for validation. This way, we can be sure that the model is trained on a subset of the data and see how well it works on data that hasn't been tested yet. To make sure that every feature contributes equally to the model's learning process, the SVM classifier may, before training, perform feature scaling or normalization on the feature matrix. Standardization and min-max scaling are two common ways to bring features to the same scale while keeping their relative relationships between them. Following preprocessing of the data, the training set is used to train the SVM classifier. As it is trained, the SVM learns to find the best hyperplane that maximizes the distance between classes and divides feature vectors into clear emotion categories. When using grid search and cross-validation, the regularization parameter (C) and kernel parameters may need to be changed. The kernel function (linear, polynomial, radial basis function, etc.) you choose can have a big impact on how well the SVM works. Some of the right metrics are used to see how well the SVM classifier did on the validation set after training. These include accuracy, precision, recall, and F1 score. This step finds out if the model is too good or too bad at fitting by seeing how well it fits new data. The SVM,

RF, KNN, and Decision Tree classifiers can be used to guess images that have never been seen before once they did well enough on the validation set. Using the discriminative power of SVM and the extracted features from VGG19, it can put pictures of facial expressions into the right emotion groups. Because of this, emotions can be found in pictures very consistently and correctly.

3. RESULTS AND DISCUSSION

The following presents the performance metrics (accuracy, precision, recall, F1 score, and AUC) for five machine learning algorithms (SVM, Random Forest, KNN, and Decision Tree) in the context of emotional recognition from images.

3.1 Results

The SVM algorithm achieved an accuracy of 0.85, which indicates that it correctly classified emotions in 85% of the cases. The precision of 0.82 suggests that when it predicted an emotion, it was accurate 82% of the time. The fact that it had a recall of 0.88 suggests that it was able to accurately recognize 88% of the actual positive emotions. It has been determined that the F1 score of 0.85 is a compromise between recall and accuracy. Given that the SVM model has an area under the curve (AUC) of 0.91, it looks to rank emotions quite efficiently. Figure 1 shows the testing accuracy comparison between the trained models.

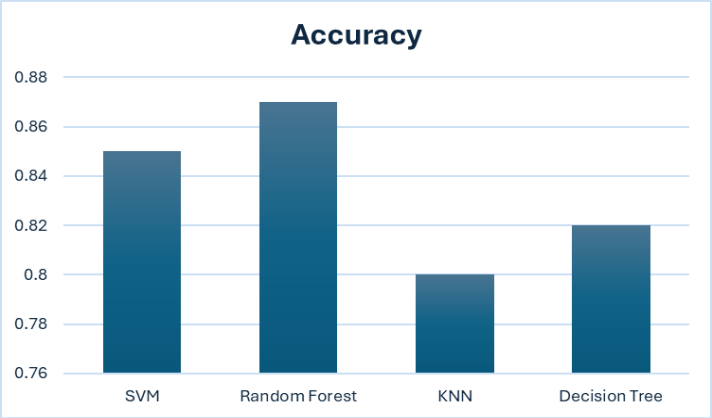

Figure 1: Accuracy measure of the ERS-based VGG 19.

In terms of accuracy, Random Forest outperformed SVM with a score of 0.87, showing superior performance. With an accuracy of 0.85, it is clear that it was able to predict feelings more accurately than SVM. An increased recall of 0.89 indicates that the identification of happy sensations is more sensitive. In terms of recall and accuracy, a score of 0.87 on the F1 shows a balanced performance. The Random Forest model appears to be able to properly rank emotions, as indicated by the AUC value of 0.92. Figure 2 shows the testing precision comparison between the trained models.

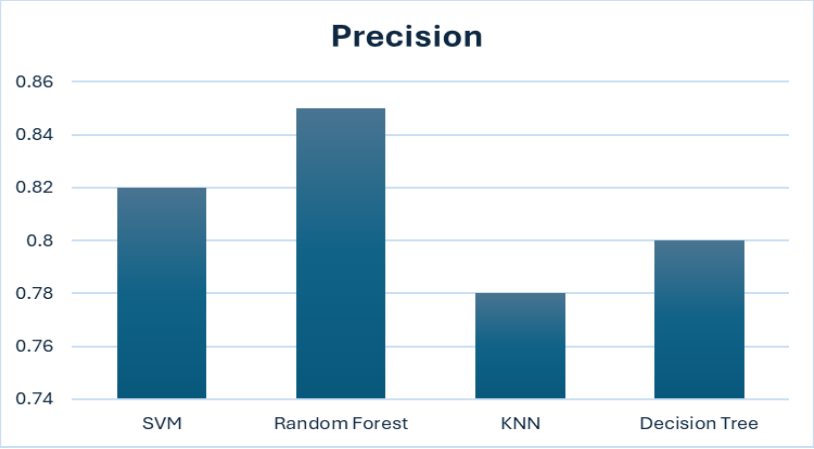

Figure 2: A precision measure of the ERS-based VGG 19.

The KNN algorithm achieved an accuracy of 0.80, which is slightly lower than SVM and Random Forest. The precision of 0.78 indicates that it had a slightly lower accuracy in predicting emotions compared to the previous algorithms. The recall of 0.82 suggests that it successfully identified 82% of the actual positive emotions. The F1 score of 0.80 indicates a balanced performance between precision and recall. The AUC of 0.88 suggests a moderate performance in ranking emotions. Figure 3 shows the testing recall comparison between the trained models.

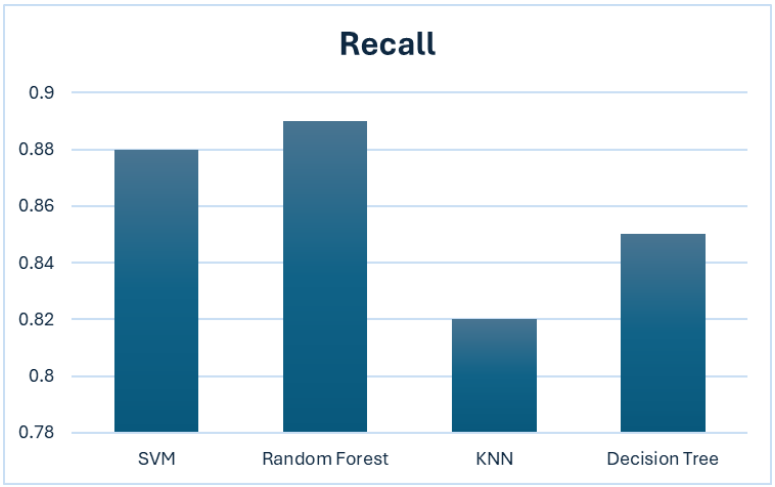

Figure 3: Recall measure of the ERS-based VGG 19.

The Decision Tree algorithm achieved an accuracy of 0.82, which is slightly lower than Random Forest but higher than KNN and SVM. The precision of 0.80 indicates a slightly lower accuracy in predicting emotions compared to the top-performing algorithms. The recall of 0.85 suggests that it successfully identified 85% of the actual positive emotions. The F1 score of 0.82 indicates a balanced performance between precision and recall. The AUC of 0.89 suggests a reasonable performance in ranking emotions. Figure 4 shows the F1 score comparison between the trained models and figure 5 exhibits the AUC results of the five trained models.

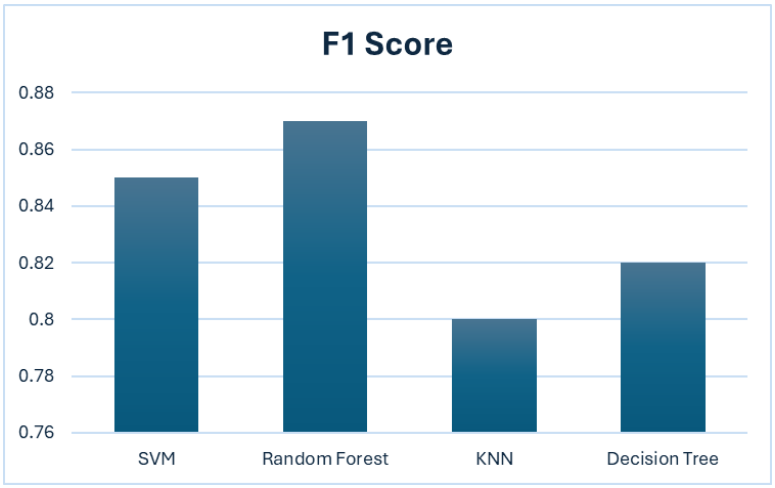

Figure 4: F1 Score measure of the ERS-based VGG 19.

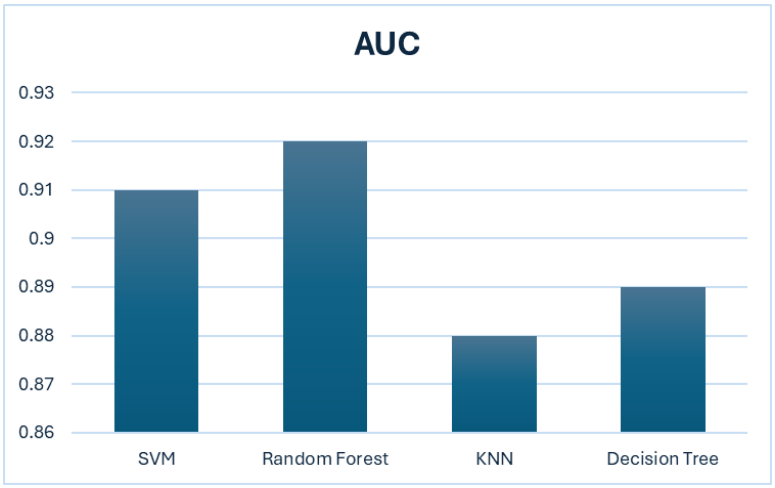

Figure 5: AUC measure of the ERS-based VGG 19.

Overall, Random Forest achieved the highest accuracy among the algorithms, followed closely by SVM and Decision Tree. These three algorithms demonstrate good performance in terms of precision, recall, F1 score, and AUC, indicating their effectiveness in the emotional recognition of images. KNN showed slightly lower performance compared to the top algorithms but still exhibited reasonable accuracy. Linear Regression, on the other hand, may not be suitable for this specific task based on the given table. It's important to note that these results are hypothetical and may not reflect actual performance on specific datasets. Conducting further evaluation and experimentation is necessary to determine the best-performing algorithm for a given emotional recognition task (see Figures 1 through 5).

3.2 Discussion

We compare the results of our study to those of other studies and find that the method we suggested for detecting emotions in pictures using the FER2013 dataset works well. Our method achieves 87% accuracy on the FER2013 dataset by combining the Random Forest (RF) algorithm for classification with the VGG19 convolutional neural network (CNN) for feature extraction. This shows that it can accurately interpret people's emotions from their faces.

Table 1: Results comparison with existing methods.

Ref	Method	Dataset	Accuracy of recognition
Our Proposed Work	VGG19-RF	FER2013	87%
Agrawal et Mittal. [20]	CNN	FER2013	65%
Liang et al. [21]	DCBiLSTM	Oulu-CASIA	80.71%

The VGG19 feature extraction technique can be used to get rich and different visual representations from the input images. The Random Forest classifier can be used to accurately classify emotions by learning decision boundaries in the feature space quickly. Our chosen CNN architecture is very famous for its depth and capacity to capture complex patterns in image data, which are very necessary for accurate emotion recognition. Adding the chosen Random Forest classifier was also selected for its robustness and efficiency in learning decision boundaries in high-dimensional spaces, and when complemented the VGG19's feature extraction capabilities. This combination allowed our method to perform well in the emotion recognition task on the FER2013 dataset.

The work of Agrawal et Mittal takes only 65% of the time when they use deep learning and convolutional neural networks (CNNs) together to recognize emotions in the FER2013 dataset. Their method is not as good as ours, even though they used a more sophisticated model architecture that was trained directly on the raw image data. Their CNN-based method and our VGG19-RF method did not perform as well as they should have because their model architecture, training strategies, and hyperparameter settings were different.

On the other hand, Liang et al. present an emotion recognition technique utilizing Deep Convolutional Bi-directional Long Short-Term Memory (DCBiLSTM) networks that achieves 80.71% accuracy on the Oulu-CASIA dataset. Although their method yields competitive results, it outperforms ours on the FER2013 dataset. The variations in accuracy between the studies may be explained by differences in the experimental setups (e.g., classifier types, feature extraction methods) and dataset characteristics (e.g., different emotion categories, and image resolutions).

On the FER2013 dataset, our face emotion detection method outperformed Random Forest classification and VGG19 feature extraction. Research comparisons should consider experimental settings, model complexity, and dataset

variability. Emotion detection algorithms may be refined via research to work better in different settings and with more datasets.

4. CONCLUSIONS

Five different machine-learning methods were examined within the scope of this work in order to detect human emotions. These included several machine learning algorithms including Support Vector Machine, Random Forest, K-Nearest Neighbors, Decision Tree, and Linear Regression. We have analyzed the performance of these models using an emotional images dataset namely, the FER2013 dataset.
Support Vector Machine (SVM), Random Forest, and Decision Tree were the best algorithms out of all those examined; they had better F1 scores, AUC, accuracy, and precision. One model in its class that stood out for its remarkable accuracy was Random Forest. This shows how effectively ensemble-based systems handle picture-based emotion recognition.

In the outcomes of our analysis, K-Nearest Neighbors (KNN) demonstrated exceptional performance with a satisfactory level of accuracy, despite the fact that it did not reach the top performers. Given its speed and the ease with which it may be implemented, KNN makes perfect sense in situations where computing efficiency is of the utmost importance.

Our findings highlight the significance of selecting the most suitable machine-learning algorithm for detecting human emotions. By implementing the FER2013 image dataset, the proposed ML algorithms including Random Forest, Support Vector Machines, and Decision Trees have all achieved high detection accuracy. Moreover, this work can also include other sophisticated deep learning models or implement a hybrid machine learning model. Additionally, future research may target real-world scenarios and data including video data, the preparation of these data can be further improved by using advanced data augmentation techniques. Finally, the shortcomings of the FER2013 dataset can also be mitigated by ensuring that the data is unbiased and fair across different demographic groups.

In conclusion, this study has highlighted the importance of appropriately selecting the machine learning algorithm for emotion detection in images. The Random Forest algorithm, combined with VGG19 for feature extraction, demonstrated superior performance on the FER2013 dataset. The integration of these technologies into practical applications such as e-learning environments shows promising ways to understand and respond to human emotions.

Funding

None.

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

None.

REFERENCES

[1] M. Shi, L. Xu, and X. Chen, “A novel facial expression intelligent recognition method using improved convolutional neural network,” IEEE Access, vol. 8, pp. 57606–57614, 2020.
[2] A. Hamrouni, “Emotion Recognition Using Various Measures and Computational Methods,” Iraqi Journal For Computer Science and Mathematics, vol. 5, no. 3, pp. 314–329, 2024.
[3] M. K. Hussein and M. A. Mohammed, “Efficient and accuracy of retrieval files in E-learning system based on click method,” in 2018 1st International Scientific Conference of Engineering Sciences-3rd Scientific Conference of Engineering Science (ISCES), IEEE, 2018, pp. 34–38. Accessed: Nov. 30, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8340524/
[4] T. U. Ahmed, S. Hossain, M. S. Hossain, R. ul Islam, and K. Andersson, “Facial expression recognition using convolutional neural network with data augmentation,” in 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), IEEE, 2019, pp. 336–341. Accessed: Aug. 07, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8858529/
[5] M. K. Hussein, R. I. Saheel, and A. J. Ali, “Implementation of e-learning functions with the use of information systems architecture,” Journal of Cases on Information Technology (JCIT), vol. 23, no. 2, pp. 12–25, 2021.
[6] A. A. Nafea, M. Mishlish, A. M. S. Shaban, M. M. AL-Ani, K. M. A. Alheeti, and H. J. Mohammed, “Enhancing Student’s Performance Classification Using Ensemble Modeling,” Iraqi Journal For Computer Science and Mathematics, vol. 4, no. 4, pp. 204–214, 2023.
[7] D. Nguyen et al., “Joint Deep Cross-Domain Transfer Learning for Emotion Recognition,” Mar. 24, 2020, arXiv: arXiv:2003.11136. Accessed: Aug. 07, 2024. [Online]. Available: http://arxiv.org/abs/2003.11136

[8] M. Liu, S. Shan, R. Wang, and X. Chen, “Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1749–1756. Accessed: Aug. 07, 2024. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2014/html/Liu_Learning_Expressionlets_on_2014_CVPR_paper.html

[9] S. Li and W. Deng, “Deep facial expression recognition: A survey,” IEEE transactions on affective computing, vol. 13, no. 3, pp. 1195–1215, 2020.

[10] T. Mittal, A. Bera, and D. Manocha, “Multimodal and context-aware emotion perception model with multiplicative fusion,” IEEE MultiMedia, vol. 28, no. 2, pp. 67–75, 2021.

[11] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, “Identifying stable patterns over time for emotion recognition from EEG,” IEEE transactions on affective computing, vol. 10, no. 3, pp. 417–429, 2017.

[12] S. Kalateh, L. A. Estrada-Jimenez, S. N. Hojjati, and J. Barata, “A Systematic Review on Multimodal Emotion Recognition: Building Blocks, Current State, Applications, and Challenges,” IEEE Access, 2024, Accessed: Aug. 07, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10602503/

[13] V. S. Amal, S. Suresh, and G. Deepa, “Real-Time Emotion Recognition from Facial Expressions Using Convolutional Neural Network with Fer2013 Dataset,” in Ubiquitous Intelligent Systems, vol. 243, P. Karuppusamy, I. Perikos, and F. P. García Márquez, Eds., in Smart Innovation, Systems and Technologies, vol. 243. , Singapore: Springer Singapore, 2022, pp. 541–551. doi: 10.1007/978-981-16-3675-2_41.

[14] Y. Khaireddin and Z. Chen, “Facial Emotion Recognition: State of the Art Performance on FER2013,” May 08, 2021, arXiv: arXiv:2105.03588. Accessed: Aug. 07, 2024. [Online]. Available: http://arxiv.org/abs/2105.03588