RESEARCH ARTICLE

Identifying Employee Emotions based on Facial Expression Analysis in the IT Sector using Ensemble Learning Techniques

E. M. P. I. Roopasinghe[1*], A. Mohamed Aslam Sujah[1]

[1]*Department of Information and Communication Technology, Southeastern University, Oluvil, Sri Lanka.*

Submitted on: 2025.09.06 **Revised on: 2025.10.31** **Accepted on: 2026.01.07**

Abstract: In the contemporary workplace, especially in the IT sector, the emotional well-being of employees is pivotal to organizational success. High workloads and tight deadlines contribute to increased stress levels, making it essential to understand employee emotions to foster a positive work environment and enhance productivity. This research aims to develop a robust model for automatic emotion recognition through ensemble learning techniques, leveraging advancements in machine learning and facial expression analysis. The study is divided into three phases. The first phase involves the selection of the top three algorithms by evaluating five different transfer learning models based on their accuracy and processing time in emotion prediction. The second phase compares a custom-built Convolutional Neural Network (CNN) model against the selected transfer learning models from the first phase. The final phase constructs an ensemble model using the best three algorithms identified in the second phase. The dataset comprises 1,500 facial emotion samples from 84 individuals, categorized into positive and negative emotions for training and testing. The comparative analysis reveals that InceptionV3, EfficientNetV2, and MobileNetV2 exhibit superior performance, leading to the exclusion of the custom-built CNN from the ensemble model. The ensemble model achieves an impressive accuracy of 88% as evaluated using a confusion matrix.

Keywords: Emotion Detection, Employees, Facial Expression, Machine Learning, Transfer Learning, Ensemble Learning.

INTRODUCTION

In the dynamic environments of today's fast-paced workplaces, especially within the Information Technology (IT) sector [1], the emotional well-being of employees has become a foundation of organizational success [2]. The demands of heavy workloads, strict deadlines, and continuously changing technology often create a high-pressure environment and lead to emotional challenges that can negatively impact job satisfaction and productivity [3]. The risk of stress and burnout is increasing among IT professionals who must creatively solve problems and have technical expertise. Consequently, prioritizing employees' emotional health has evolved from being a personal concern to a strategic necessity [4] for maintaining a competitive edge and fostering workplace innovation.

Employees who feel emotionally supported and understood are often more motivated, engaged, and collaborative, leading to improved productivity and stronger workplace dynamics. Despite the promise, there are significant challenges associated with the implementation of facial recognition technologies in the workplace. It remains a complex task to translate facial expression analysis results into practical applications that can be easily integrated with the current systems.

Human emotions are complex, and cultural and personal variations in emotional manifestation become major challenges to emotion recognition systems [5].

**Corresponding author. E-mail address: poornimasbv@gmail.com (E. M. P. I. Roopasinghe)*

Moreover, the ethical aspects of privacy and data safety should be considered to avoid problems with the abuse of employee information or the violation of personal space.

Ensemble learning techniques have gained attention as a viable solution [6] to overcome these challenges. By leveraging the combined strengths of multiple machine learning models [7], ensemble methods enhance the accuracy, reliability, and adaptability of emotion detection systems These approaches work well when dealing with variation and complexities of facial expression so that the models used in emotion recognition can make consistent and reliable predictions. The proposed study seeks to develop and deploy an ensemble model of automatic recognition of emotions using facial expression analysis to suit the IT industry.

METHODOLOGY

In the context of emotion recognition, facial expression is the main principle. This research focuses on facial expression analysis and is considered vital. This research is constructed through an ensemble learning technique using a novel dataset. This chapter details the different analyses carried out to identify the best TL algorithms, implementation of the ensemble model [8], how the different models are evaluated to come up with the best algorithm, and how the ensemble model is constructed and tested.

A. Data Collection

The dataset collected from the real-world is designed to investigate compound emotion recognition. The two categories are positive and negative. All main emotions are compounded in those two categories and contain the dataset. The motivation for creating the dataset is to explore how well the model can perform and as well as a good impression on research topics. For now, data collection consists of 1500 front face images (male and female) with different emotions [9] captured from 84 people. The age range of people who are subject to the data collection is between 25 – 30. Subjects had different appearances such as hair styles, clothes, and accessories. Images were obtained in a controlled environment to overcome problems like background noise, head pose variations, and distance variations. And captured under the same environment. Image resolution 1080 x 1920. Room where images are captured with uniform light and uniform distance. The motivation to use such a controlled environment is to reduce preprocessing steps. High resolution images show detailed information. Each participant displayed six different emotions. The dataset is divided into two groups as a training set and test set. 80% of data is for training and remaining 20% of data for testing. Figure 1 and Figure 2 show two sets of data collection images.

Figure 1: Data collection Images – Set 1

Figure 2: Data collection Images – Set 2

B. Proposed System

This study consists of three phases. The first phase consists of selecting the best three algorithms based on their predictions on emotions by comparing five different transfer learning models according to their learning time and performance accuracy. The next phase consists of comparing one of the own CNN models with built-in transfer learning models that you choose in the first phase. Then the final phase is creating an ensemble model by using the best three algorithms chosen from the second phase. In the initial phase, a preprocessed human emotions dataset is stored in a folder with two different classes as positive emotions and negative emotions. Classification is performed to help the model uniquely identify features [10] with supervised learning. Next, we import python libraries which require them to carry out the processes. NumPy, matplotlib, and TensorFlow are used within the process. Dataset is used in InceptionV3, EfficientNetV2, ResNet50, MobileNetV2 and MobileNetV3 algorithms to evaluate the best one by comparing the learning time taken by algorithm and based on the performance accuracy shown by each model. Use the same dataset with all algorithms and fixed number of epochs. Finally figure out the best algorithms which have the lowest time taken and have high accuracy.

In the second phase, compares own built CNN model with built-in transfer learning models which were selected in the first phase. Own custom model is created for facial expression analysis and evaluated with transfer learning models according to model accuracy and time. To create this own model, we need to import python libraries. Then split the dataset into training and testing dataset. 80% of data was used for training purposes and the remaining portions used for testing purposes. The computational efficiency of each model will be calculated considering factors like training time, scores of accuracies, and losses. This analysis supports checking model feasibility. Models are evaluated through testing on unseen data samples. That process determines the performance of the model on data outside of the training set.

In the third phase, an ensemble model is constructed in the third stage using the best three algorithms selected in the phase two comparisons. To further boost general performance and resistance, a combination of numerous individual models will be integrated into the ensemble model. The algorithms with the best results are chosen in comparison from the second phase. They are chosen based on specific features of their effectiveness. These include accuracy, computational efficiency, and generalization properties. To decide on the best method that can be used for incorporating these chosen algorithms into one system, different methods for ensemble learning were tried. These include bagging, boosting, and stacking. When numerous models are used during prediction, prediction of an individual model often doesn't provide accurate predictions like what happens in ensemble methods. The integration of Algorithms into the ensemble model works such that the chosen algorithms are connected and their predictions aggregated using methodologies like averaging or weighted voting. Combining the strengths of models gives them better results during emotion recognition tasks. The ensemble model must be validated and optimized to ensure its effectiveness and reliability. This process involves testing the model using datasets, fine-tuning ensemble parameters, and optimizing performance metrics for best results. Figure 3 shows the flow chart of the proposed system.

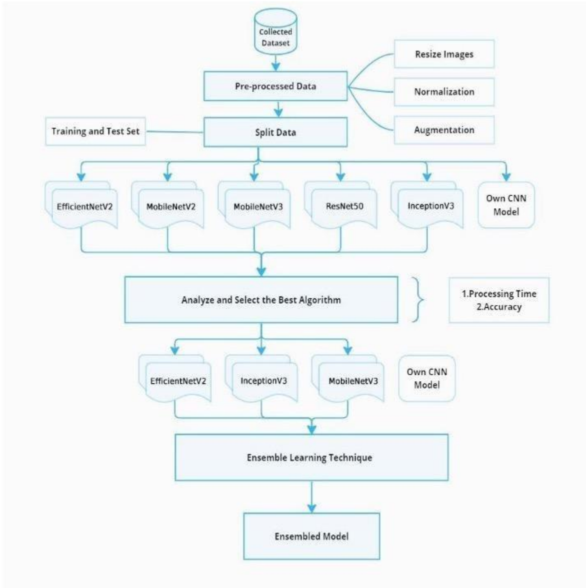

Figure 3: Flow chart of proposed system

C. Model Comparison

After evaluating the performance of each of the five models. Select the most appropriate three models to go. These top three models selected displayed significant precision and efficiency. This comparison shows the merits and shortcomings of each approach, providing useful insights for future optimization. The selected models are EfficientNetV2, InceptionV3, and MobileNetV3. EfficientNetV2 is a model that was discovered to be the most effective in classifying images.

These have been a popular choice for many kinds of applications. InceptionV3 has a complex structure and numerous convolutional layers. And excels in handling complex visual data. MobileNetV3 is designed for mobile and edge devices that provide a lightweight but capable solution for real-time image processing. All these models contribute to different skills and meet various needs in the fields of deep learning and emotional recognition. The three models with the highest accuracy are EfficientNetV2 (0.8333), MobileNetV3 (0.8367), and InceptionV3 (0.7567). These three models have shown that there are new ways of classifying images, and each one of the models has a balance between accuracy and efficiency.

The performance of MobileNetV3 and EfficientNetV2 is advantageous when it comes to resource-constrained systems, whereas InceptionV3 is a suitable option when more computational units are demanded. MobileNetV3 and EfficientNetV2 could deliver impressive results on devices with limited resources. InceptionV3, on the other hand, is reputed for tasks demanding higher computational power. The versatility of these models allows developers to choose the most suitable option based on the specific requirements of their projects.

The custom CNN [11] was made and applied with the use of the Keras Sequential API to compare directly to the transfer learning models. The model was configured to accept the input of the greyscale image of 48 x 48 x 1. The architecture has been created in three major convolutional blocks; A Conv2D layer (with 32 filters) and a 3 x 3 kernel is successively followed by a MaxPooling2D layer (with 2 x 2 size). Conv2D, which has 64 filters, and 3 x 3 kernel is followed by MaxPooling2D of size 2 x 2. It is followed by a Conv2D layer of 128 filters and 3 x 3 kernel and MaxPooling2D that has a size of 2 x 2. Every convolutional layer in the Rectified Linear Unit (ReLU) activation works.

A Flatten layer was then applied to the convolutional output, creating a fully connected network. In this network, there is a hidden Dense layer of 128 units (ReLU activation), a Dropout layer with a regularization rate of 50%, and a Dense output layer which had a Sigmoid activation function, which was used to classify the binary data. The model was trained during 30 epochs with the help of Adam optimizer and categorical cross-entropy loss function, and the image data were augmented. Figure 4 demonstrates the own built CNN model accuracy and loss changs with epochs.

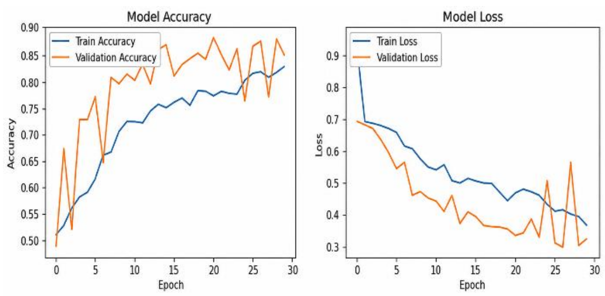

Figure 4: CNN model accuracy and loss changs

4. Accuracy and Loss of own-built CNN model

This own CNN model was found to exceed the three mentioned models in terms of accuracy, boasting an amazing accuracy score of 0.8500. Furthermore, with training times in the 983s, the trade-off between training time and accuracy is critical to consider when comparing these. The Own CNN model achieves the highest accuracy among other three models presented because of the input image sizes we used 48 x 48 pixels. The Custom model exceeds EfficientNetV2 (0.8333), MobileNetV3 (0.8367), and InceptionV3 (0.7567) by a large pixel size of 224 x 224. Therefore, own CNN model takes around 983 seconds to train. This is considerably quicker than training times for more complicated models such as EfficientNetV2, MobileNetV3, and InceptionV3, which are typically longer due to larger architectures and deeper layers. The own model is quicker training period can be useful, particularly in situations where training resources are limited, or rapid prototyping is required. Finally, the custom-created CNN model demonstrates promising performance with high accuracy and a quick training period. Despite reaching an impressive 85% accuracy, the custom-built CNN model was ultimately excluded from the final ensemble.

This choice was made based on technical justifications aimed at increasing the entire system's robustness and efficiency. First and foremost, a lack of feature diversity was a problem. Since the custom CNN was trained just on the relatively short 1500 image dataset, the patterns it caught were regarded inadequately complementary to the huge feature spaces learnt by the pre-trained models (Inception V3, EfficientNetV2, and MobileNetV2). Second, resource constraints encourage exclusion. While the specific CNN trained quicker, EfficientNetV2 and MobileNetV2 are specifically designed for deployment on resource-constrained platforms.

RESULTS

The overall accuracy of each model was measured on the test dataset. Accuracy indicated the proportion of correctly classified images out of the total number of images. The confusion matrix provided insights into the specific strengths and weaknesses of each model in classifying different emotions.

Figure 5, Figure 6, and Figure 7 show the confusion matrices for EfficientNetV2, MobileNetV3, and InceptionV3 respectively.

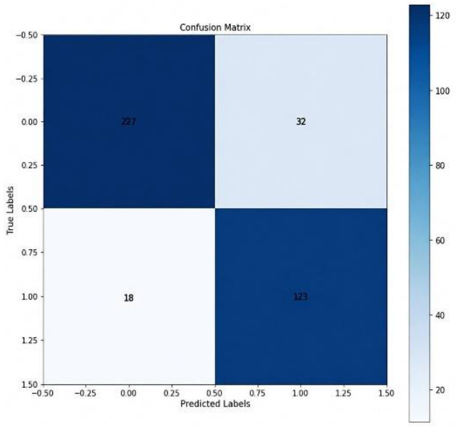

Figure 5: Confusion Matrix of EfficientNetV2

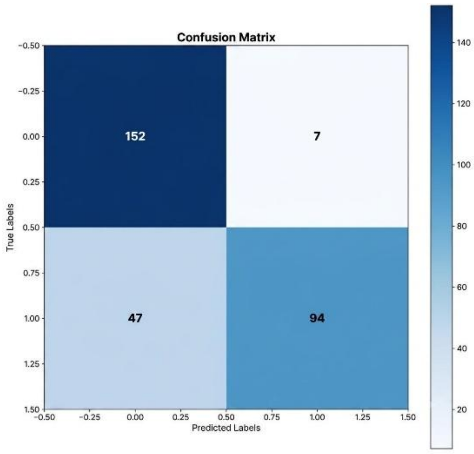

Figure 6: Confusion matrix of MobileNetV3

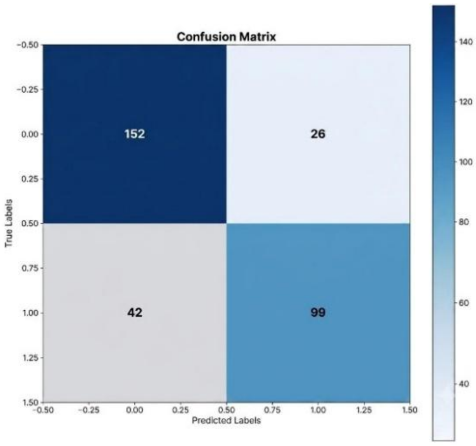

Figure 7: Confusion matrix of InceptionV3

Once the ensemble model has been defined, build it and train it on dataset, noting the training and validation of accuracy and loss. After training, assess the model on a test dataset, create a confusion matrix to analyses its performance, and plot the training and validation metrics to see how the model learns over time. This strategy makes use of the capabilities of various models, potentially improving total performance. With 30 epochs, the ensemble model [12] shows 88% accuracy. This can be compared to a significant improvement than the individual models. The ensemble method utilizes the strengths of each model as well as minimizing its flaws. The predictions of the ensemble model can be used together to come up with more accurate and strong conclusions. Figure 8 presents the ensemble model accuracy and loss variation with epochs.

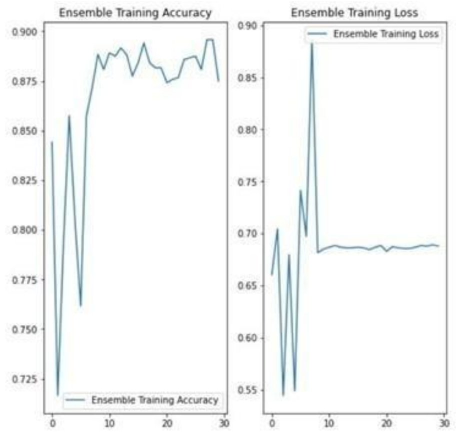

Figure 8: Ensemble model accuracy and loss variation.

The accuracy obtained with ensuing model exhibits a tremendous move towards emotion identification. This technical success is not a measure of the efficiency of algorithms only, but also organizational empathy. The strength of this model was spread in its variety of feature extraction where it managed to merge the intelligence of InceptionV3, EfficientNetV2, and MobileNetV3.

All the models add their own features, which are trained with their own weaknesses, and strike a balance between performance and efficiency. The combination was a strong, generalized classifier. This observation is very influential in automatic identification of emotions of the employees and is useful in enhancing employee health, decreasing stress levels, and consequently increasing productivity at the workplace.

CONCLUSION

This research was able to create a successful emotion recognition system for the IT sector based on an ensemble learning approach on a specifically compiled dataset of 1,500 facial emotion samples. The combination of three progressive models of transfer learning, including Inception V3, EfficientNetV2, and MobileNetV2, and the ultimate ensemble model through the advantages of the complementary learning approaches resulted in an accuracy of 88% in classifications. This result demonstrates better generalization and classification of employee emotion into positive and negative categories using ensemble methods than using a single model architecture. The system will work well in enabling IT organizations to monitor and control the emotional status of employees automatically, leading to improved work environment and productivity.

REFERENCES

[1] A. Hassouneh, A. M. Mutawa, and M. Murugappan, Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods', *Inform Med Unlocked*, vol. 20, p. 100372, 2020, doi: 10.1016/j.imu.2020.100372.

[2] J. Guo *et al.*, 'Dominant and Complementary Emotion Recognition from Still Images of Faces', *IEEE Access*, vol. 6, pp. 26391–26403, 2018, doi: 10.1109/ACCESS.2018.2831927.

[3] S. S. Panchal, A. Hiremath, and N. R. Toravi, 'Automatic face emotion recognition', *Proceedings of the International Conference on Smart Systems and Inventive Technology, ICSSIT 2018*, no. Icssit, pp. 402–405, 2018, doi: 10.1109/ICSSIT.2018.8748847.

[4] N. Mishra and A. Bhatt, 'Feature extraction techniques in facial expression recognition', *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, no. Iciccs, pp. 1247–1251, 2021, doi: 10.1109/ICICCS51141.2021.9432192.

[5] W. Li, M. Luo, P. Zhang, and W. Huang, 'A Novel Multi-Feature Joint Learning Ensemble Framework for Multi-Label Facial Expression Recognition', *IEEE Access*, vol. 9, pp. 119766–119777, 2021, doi: 10.1109/ACCESS.2021.3108838.

[6] M. Hussain, E. U. H. Qazi, H. A. Aboalsamh, and I. Ullah, 'Emotion Recognition System Based on Two-Level Ensemble of Deep-Convolutional Neural Network Models', *IEEE Access*, vol. 11, no. January, pp. 16875–16895, 2023, doi: 10.1109/ACCESS.2023.3245830.

[7] Ronak Kosti; Jose M. Alvarez; Adria Recasens;, and Agata Lapedriza, Context Based Emotion Recognition Using EMOTIC Dataset', vol. 42, no. 11, 2019.

[8] H. P. Martinez, Y. Bengio, and G. Yannakakis, 'Learning deep physiological models of affect', *IEEE Comput Intell Mag*, vol. 8, no. 2, pp. 20–33, 2013, doi: 10.1109/MCI.2013.2247823.

[9] [M. M. Thiruthuvanathan, B. Krishnan, and M. Rangaswamy, 'Engagement Detection through Facial Emotional Recognition Using a Shallow Residual Convolutional Neural Networks', *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 2, pp. 236–247, 2021, doi: 10.22266/ijies2021.0430.21.

[10] Sherstinsky, Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network', Physica D, vol. 404, p. 132306, 2020, doi: 10.1016/j.physd.2019.132306.

[11] S. Ren, K. He, R. Girshick, and J. Sun, 'Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks', *IEEE Trans Pattern Anal Mach Intell*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[12] B. Myroniv, C.-W. Wu, Y. Ren, A. Christian, E. Bajo, and Y.-C. Tseng, 'Analyzing User Emotions via Physiology Signals', *Ubiquitous International*, vol. 1, no. 2, pp. 11–25, 2017.