SINERGI Vol. 28, No. 2, June 2024: 413-424
http://publikasi.mercubuana.ac.id/index.php/sinergi
http://doi.org/10.22441/sinergi.2024.2.020

Students' emotion classification system through an ensemble approach

Muhajir Muhajir[1], Kahlil Muchtar[1]*, Maulisa Oktiana[1], Akhyar Bintang[1,2]
[1]Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Indonesia
[2]Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Malaysia

Abstract
Emotion is a psychological and physiological response to an event or stimulus. Understanding students' emotions helps teachers and educators interact more effectively with students and create a better learning environment. The importance of understanding students' emotions in the learning process has led to exploring the use of facial emotion classification technology. In this research, an ensemble approach consisting of ResNet, MobileNet, and Inception is applied to identify emotional expressions on the faces of school students using a dataset that includes emotions such as happiness, sadness, anger, surprise, and boredom, acquired from students of Darul Imarah State Junior High School, Great Aceh District, Indonesia. Our dataset is available publicly, and so-called USK-FEMO. The performance evaluation results show that each model and approach has significant capabilities in classifying facial emotions. The ResNet model shows the best performance with the highest accuracy, precision, recall, and F1-score, which is 86%. MobileNet and Inception also demonstrate good performance, indicating potential in handling complex expression variations. The most interesting finding is that the ensemble approach achieves the highest accuracy, precision, recall, and F1-score of 90%. By combining predictions from the three models, the ensemble approach can consistently and accurately address emotion variations. Implementing emotion classification models, individually and in an ensemble format, can improve teacher-student interactions and optimize learning strategies that are responsive to students' emotional needs.

Keywords:
Emotion Classification;
Ensemble;
Inception;
MobileNet;
ResNet;

Article History:
Received: November 18, 2023
Revised: December 26, 2023
Accepted: January 10, 2024
Published: June 2, 2024

Corresponding Author:
Kahlil Muchtar
Department of Electrical and Computer Engineering, Universitas Syiah Kuala
Email: kahlil@usk.ac.id

INTRODUCTION

In the current digital era, technology is developing rapidly and has a big impact on various aspects of life, including in the field of education [1]. One important aspect of the world of education is the understanding of students' emotions. Emotions are a natural part of the human experience, and understanding and managing them well is an important aspect of personal well-being and healthy social interactions [2]. Emotions play a crucial role in the learning process and social interaction at school. Students' emotions can affect motivation, understanding, social interaction, and even perceptions of themselves as students [3]. Therefore, a deep understanding of students' emotions and the ability to detect and respond well to them is a key factor in improving the quality of learning.

Along with the development of science, more and more educators and researchers are realizing the importance of detecting students' emotions as an integral element in the learning experience. While direct interaction between teachers and students remains important, the challenges of managing a classroom that is diverse in terms of backgrounds, needs, and learning preferences have driven the adoption of technology as a support in monitoring students' emotions.

Detecting students' emotions is not just about grossly recognizing positive or negative emotions, but also about understanding the deeper complexities of how they feel as they engage in the learning process. When students feel frustrated, anxious, engaged, or, these changes can impact their engagement, ability to absorb information, and connection with the content being studied [4].

One way to detect student emotions is through emotion recognition technology, which uses data such as facial expressions, voice, or body behavior. Unlike a sentiment or complaint analysis approach that mainly utilizes text information [36][37], emotion recognition can provide an objective view of emotional changes in various learning situations, both in the physical classroom and in the increasingly common distance learning environments used. In this context, emotion recognition technology through facial expressions is becoming increasingly attractive. Therefore, developing a classification system for students' facial emotions can help educators understand and respond to students' emotional states more effectively. Figure 1 illustrates examples of various types of human emotions [5].

As computer vision and deep learning techniques advance, the study of human-computer interaction increasingly focuses on facial expression recognition. Deep learning is widely used in various computer vision tasks, including facial expression recognition, and has succeeded excellently [6].

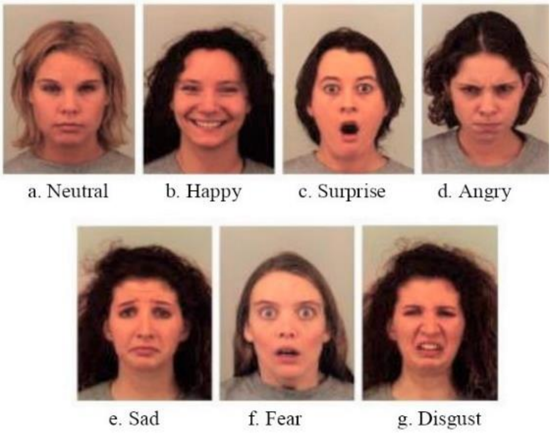

Figure 1. Examples of the types of emotions: (a) neutral, (b) happy, (c) surprise, (d) angry, (e) sad, (f) fear, and (g) disgusted

Ahmed has conducted related research, MZI et al. [7] classified emotions using the EEG method, then Ashraf, N et al [8] used the Random Forest (RF), Decision tree (DT), Sequential minimal optimization (SMO), AdaBoostM1, and Bagging), deep-learning algorithms (Convolutional Neural Networks (1D-CNN), Long short-term memory (LSTM), and LSTM with CNN features) and transformer-based baseline (BERT). Using the CNN method, Keshri, A [9] classifies emotions in real time. Li, Z et al. [10] use an end-to-end ERC model. Although many studies on emotion classification have only focused on a single classifier, this research proposes the Ensemble method, combining several classifiers to increase the resulting accuracy [11]. Each classifier has different advantages, and if several of these classifiers are combined, the results achieved will be better. The advantage of the ensemble method is that the classification results are more accurate because this method compares the results of the single classifiers used.

Several studies have been conducted by Raza [12], Kim et al. [13], Dogan et al. [14], Kuncheva et al. [15], Brown [16], Chandra et al. [17], Hassen et al. [18], Liu and Zhang [19], demonstrating that ensemble methods can enhance classification accuracy compared to single-classifier systems when the classifiers are accurate and highly diverse [20].

In this study, analyzing the Facial Emotion Classification System of school students using the Ensemble method combines the results of several CNN architectures, namely ResNet, MobileNet, and Inception, using a dataset of facial expressions of students at SMP Negeri 1 Darul Imarah, Aceh Besar District. This research aims to increase the accuracy and reliability of recognizing emotions from students' facial expressions to positively contribute to increasing interaction and response in the educational environment.

The main contribution of our work is two-fold; (1) a voting-based student emotion classification model is proposed, and (2) the new face emotion (for junior high school level) dataset is introduced and provided publicly through this link https://muhajir2111.github.io/USK-FEMO-DATASET/. This work is an aid for educators in identifying students' emotions easily and taking appropriate actions to improve the quality of learning accordingly. In addition, this research can also be a basis for further research in combining emotion recognition technology with broader educational developments.

MATERIAL AND METHOD

Emotion Classification Scheme

Emotional classification schemes group or identify emotions based on various criteria [21]. One of the most popular schemes used is the basic emotion model developed by Paul Ekman [22]. Figure 2 shows an emotion classification scheme based on Ekman's model.

1. Happy: This emotion includes feelings of pleasure, satisfaction, and joy.
2. Fear: This emotion arises in response to a threat or danger.
3. Anger: This emotion appears in response to frustration, disapproval, or humiliation.
4. Sadness: This emotion is associated with feelings of loss, sadness, or regret.
5. Disgust: This emotion is associated with feelings of dislike for something that is considered disgusting or disgusting.
6. Surprise: This emotion appears when there is a sudden or unexpected event.

This emotion grouping scheme is useful as a general framework for understanding and categorizing different types of human emotions.

Figure 2 presents a schematic diagram of two machine learning classification systems types: Concatenation style and parallel style. This diagram shows how several classifiers process input images or data and then combine them using decision rules to produce the result.

Figure 3 shows two classification transformation methods in image data processing: concatenation and parallel styles. In concatenation style (a), image data is processed sequentially through various classifiers before reaching the final result. Each classification produces an output that becomes an input for the next classification, and this process continues until it reaches the "Final Result". Meanwhile, in parallel style (b), image data is processed by several classifiers simultaneously.

The results of all these classifications are then combined via a "decision rule" to produce a "Final Result". In the concatenation style, data flow occurs from left to right through each classification. In contrast, in parallel style, data flows to all classifications simultaneously, and the results are combined. It is important to note that we employ the original version of InceptionV3, MobileNetV2, and Resnet152 as our baseline models.

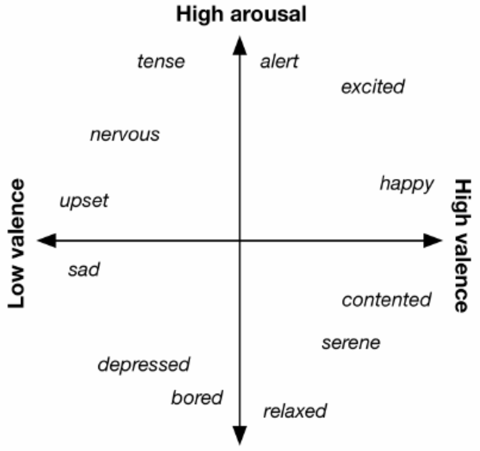

Figure 2. Scheme for Grouping Types of Emotions

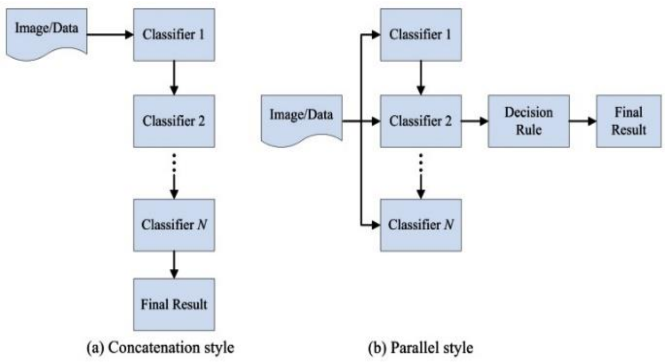

Figure 3. Emotion Classification System Scheme [23]

Inception

The Inception model is an artificial neural network architecture developed by a team of researchers at Google Research in 2014. Inception became known for its deep and efficient architectural design innovations, especially in image classification tasks. The main feature of the Inception model is the use of the Inception module, which consists of several convolution filters of different sizes applied in parallel at different levels of resolution in the image [24]. This allows the model to extract features at various levels of abstraction efficiently. The basic architecture of Inception is illustrated in Figure 4 [25].

The goal of Inception V3 was to maintain network efficiency while cutting down on the number of connections and parameters. The convolutional layer is broken down into three successive convolutions of size 3x3. Batch Normalization and ReLU activation blocks are used after each convolution, as explained in Figure 4.

MobileNet

MobileNetV2 is a Deep Neural Network that has been deployed for the classification problem. The foundation layers are defrosted to prevent any damage to previously acquired characteristics. To learn how to distinguish all emotion classes, new trainable layers are added, and these layers are trained using the gathered dataset. The weights are then saved after the model has been adjusted. By using pre-trained models, one can take the benefit of pre-biased weights without sacrificing previously learned features and prevent needless computational expenditures.

As seen in Figure 5, MobileNet changes standard convolution to convolution based on depth and 1×1 based on points. Point-based convolution is used to combine the outputs of depth-based convolution using filters for each input, thereby reducing the model size and computational process. MobileNet uses 3x3 depth convolution, reducing the computational process to 1/8 – 1/9 of other convolutions [26].

ResNet

ResNet, short for "Residual Network," is an artificial neural network architecture that is revolutionary in the development of deep models (deep learning) [27]. ResNet was developed by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun from Microsoft Research in 2015.

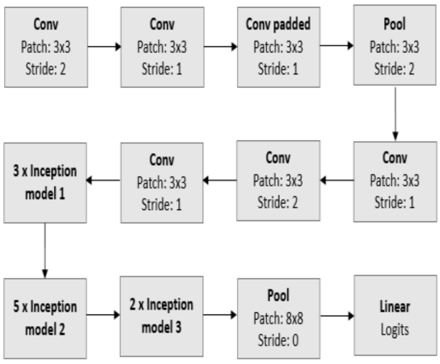

Figure 4. Basic Inception Architecture

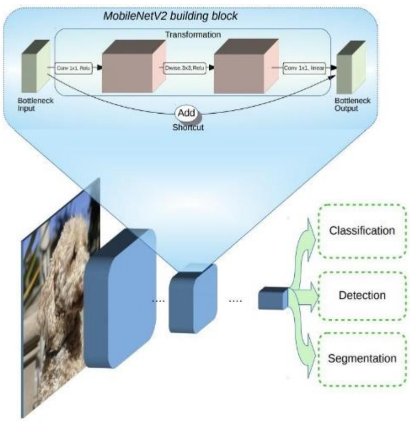

Figure 5. MobileNet Architecture

This architecture is famous for solving a problem known as the vanishing gradient problem which is often encountered in very deep artificial neural networks [27]. This architecture can be used in computer vision tasks such as image classification, object localization, and object detection.

ResNet consists of several residual blocks, which are composed of a convolutional layer, a batch normalization layer, and a shortcut that connects the original input to the output of the residual block. Figure 6 shows schematically how a Residual Block with an Identity Shortcut (RB-IS) and a Residual, Block with Projection Shortcut (RB-PS) operate.

Ensemble

The Ensemble model is one of the ensemble methods used in machine learning to improve the performance and accuracy of predictions [28]. This approach involves combining the predicted results from several individual models or bases, and then selecting the most frequent (majority) result as the final result. The key principle behind ensemble models is that combining predictions from multiple models can produce more reliable and consistent decisions than a single model. This is a powerful way to increase the prediction accuracy of a classification model. A schematic diagram of a typical ensemble learning system is shown in Figure 7 [12].

Methods

Figure 8 is an illustration of the flow of research conducted. That includes data collection, preprocessing, model architecture training, Ensemble, and model evaluation.

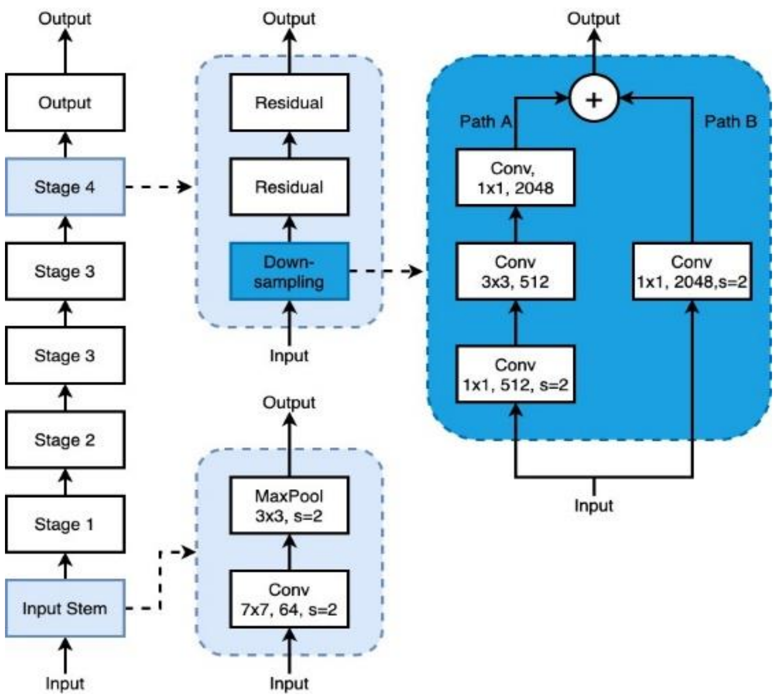

Figure 6. ResNet Architecture

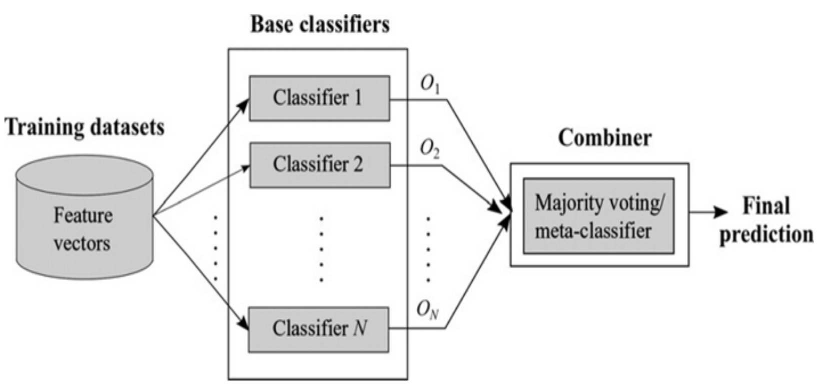

Figure 7. Ensemble Learning Architecture

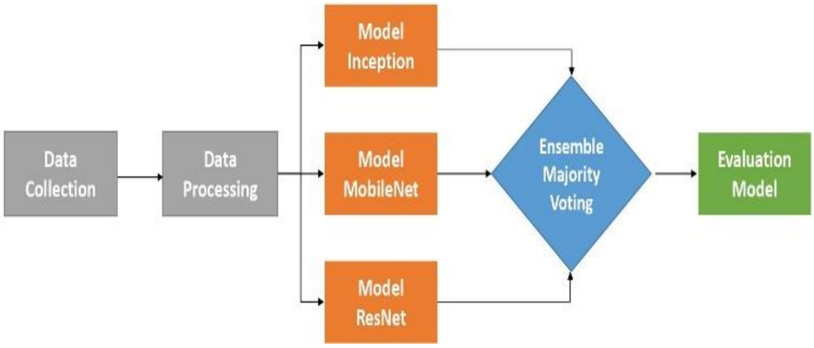

Figure 8. Research Flow

Data Collection

The facial emotion dataset was obtained from the Great Aceh Regency, Aceh, who were randomly selected. Each participant gave a variety of facial expressions including happy, sad, angry, surprised, and bored. These emotions were chosen because they are generally felt by students during the learning process. After taking the images, the images are grouped into a training folder and a test folder. The total amount of data is 1250 images. The training folder has 80% image data, while the validation folder contains 20% image data of the total data. Table 1 lists the details of the data set.

The dataset collection was carried out using a Nikon D5300 digital camera. Pictures are taken in certain environments that allow good visibility. In particular, shooting was carried out in the morning between 08.00 and 12.00 local time. The picture was taken in a well-lit room using an aperture setting of f/ 5.6, 1/80 sec shutter, and autofocus. The camera is placed on a tripod with a distance of about 1 meter from the subject and a white wall as the background. In total, 1250 images were captured, with each category consisting of 250 images. Figure 9 shows the scenario of collecting the dataset [29].

Figure 10 shows the results of the Emotion image dataset taken using a Digital Camera which consists of 5 types of emotions including 10. a happy emotion, 10.b angry emotions, 10.c sad emotions, 10.d surprised emotions, 10.e bored emotions, respectively.

Table 1. The details on the distribution of the dataset

Classes	Training Data	Testing Data
Happy	192	58
Sad	192	58
Anger	192	58
Surprised	192	58
Bored	192	58
Data Total	1250	

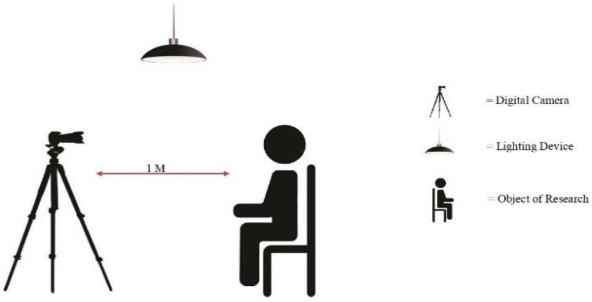

Figure 9. Dataset Collection Scenario

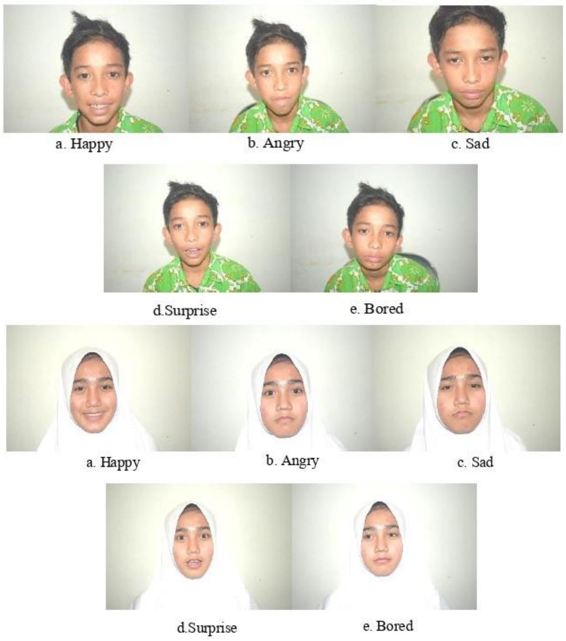

Figure 10. Example of emotional image data Captured by Digital Camera

Data Processing

1. Cropping
   Cropping is done so that the system can more easily understand the facial expression that the image is conveying. Apart from that, cropping is useful for improving the accuracy of results obtained from training and validation. Cropping is done by cutting unimportant parts of the image, such as the

background, which is too dominant over the face, making it difficult for the system to classify the type of expression. Apart from the background, disturbing objects will be cropped.

2. Resize
   Resizing is the process of changing the image size to a smaller or larger size [30]. At the stage of taking photos, facial expressions from the image have different sizes. Therefore, all images of facial expressions in this study were converted to 224×224 pixels. The image is converted to a smaller size in order to speed up the image processing process.

Model Test

1. Inception
   Inception is an architecture known as the use of modules Inception which has multiple convolution filters of different sizes at each step. This allows the model to understand features at different levels of resolution. Inception is used in the classification of emotions by changing the final layer and adjusting it to the number of emotion categories. Inception is also trained with a dataset of facial images that have emotion labels. The structure of the Inception module can be seen in Figure 2.
2. MobileNet
   MobileNet is an artificial neural network architecture designed for computational efficiency on low-power devices. It uses depthwise separable convolution to reduce the number of parameters and operations required. MobileNet is used in emotion classification by replacing its final layer with the appropriate layer for the task. The last layer consists of a shipping layer and an output layer with the number of neurons in each emotion category identified. MobileNet trained using a facial image dataset with emotion labels. During training, the model will learn appropriate feature representations to classify emotions.
3. ResNet
   ResNet is known for its deep layers and unique structure with skip connections or shortcut connections. ResNet is used in emotion classification by adding a final layer appropriate for the task. Like MobileNet, the layer consists of a shipping layer and an output layer. ResNet also trained with a dataset of facial images for each emotion label. Deep architecture helps models to understand more abstract and complex features.

Ensemble

Model Ensemble is an ensemble method used in machine learning to improve the performance of emotion classification by combining the prediction results from several individual models and choosing the emotion class that gets the most votes (majority) as the final prediction. The step of working on this method is to calculate the results of each classifier (Inception, MobileNet, and Resnet), then it will be weighted again using Ensemble weighting, namely majority voting, average of probability and weighted voting [31]. As discussed in [33-35], such approaches are able to achieve very competitive research in some fields.

In Figure 11, there is a design and simulation flow for 3 models, where model A (InceptionV3), model B (MobileNetV2) and model C (ResNet152) will take the final weight from each model without changing the contents of each model. The final weights of each model will be combined using Average voting.

Model Evaluation

In this study, a comprehensive evaluation was carried out on the performance of three different facial emotion classification models: Ensemble Model, ResNet, MobileNet, and Inception. This evaluation aims to provide in-depth insight into the ability of each model to identify and classify emotional expressions on the faces of school students. The parameters used are:

1. Accuracy
   Accuracy measures the percentage of correct predictions out of the total predictions made by the model. This parameter provides a general idea of how well the model recognizes emotional expressions in facial images.
2. Precision
   Precision measures the proportion of correct positive predictions out of all positive predictions made. In this context, precision indicates how precise the model is in predicting positive emotions.

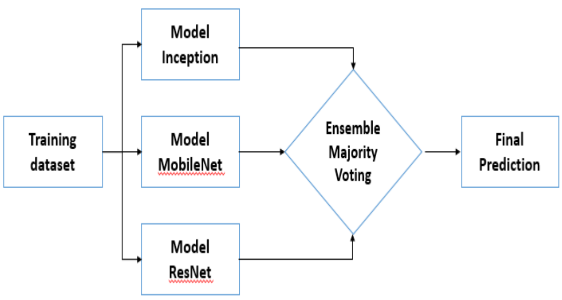

Figure 11. Ensemble Block Diagram

3. Recall
   Recall, also known as sensitivity, measures the proportion of correct positive predictions out of all positive cases in the actual data. This parameter shows the extent to which the model is able to detect all existing emotional cases.
4. F-1 Score
   F1-score is the harmonic meaning between precision and recall. This metric provides insight into the balance between positive accuracy and sensitivity in the model.

RESULTS AND DISCUSSION

After a series of experiments carried out involving three different classification models, namely ResNet, MobileNet, and Inception, as well as an ensemble approach. The performance results of the model can be seen in Table 2.

Based on Table 2, it can be identified that all models and ensemble approaches have good performance with accuracy above 0.8. This shows that all models are able to recognize emotional expressions on students' faces at a high level. The ResNet model has the best performance with the highest accuracy, precision, recall, and F1-score, which is 0.86, and the ensemble model achieves the highest performance with the highest accuracy, precision, recall, and F1-score, 0.9. This shows that combining the prediction results from all models provides better results than individual models.

Graphical visualization of the accuracy-loss and confusion matrix results for each model can be seen in Figure 12-19, and the results show that there is a difference in the accuracy of each model.

Table 2. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1 Score
MobileNet	0.85	0.85	0.85	0.85
Inception	0.84	0.85	0.84	0.84
ResNet	0.86	0.86	0.86	0.86
Ensemble	0.9	0.9	0.9	0.9

Figure 12. Accuracy-Loss Graph of the MobileNet model.

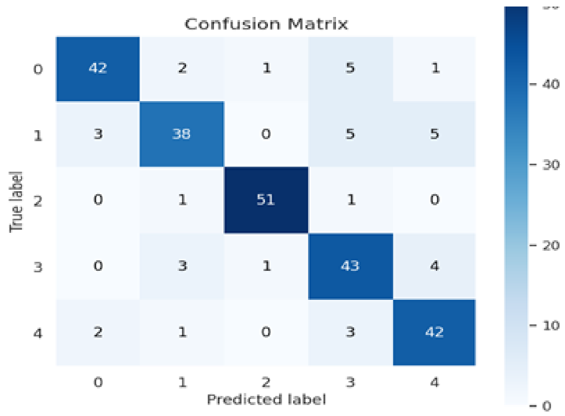

Figure 13. Confusion Matrix of the MobileNet Model

Figure 12 shows the accuracy and loss of the facial expression classification model using MobileNet during training. The graph on the left shows the model's accuracy on training data (blue line) and validation data (orange line) increasing over time. The graph on the right shows the model loss on training data (blue line) and validation data (orange line) decreasing over time.

A confusion matrix is a table that shows the performance of a classification model on test data. This image has rows representing actual labels and columns representing predicted labels. Each cell in the table shows the number of samples that had a particular true label and were predicted to be a particular label. In the case of facial expression classification using the MobileNet model, the confusion matrix shows how well the model predicts each facial expression. From this image, the model can correctly predict 42 images of Angry, 38 Bored, 51 Surprised, 43 Happy and 42 Sad face.

Figure 14 shows the results of the Inception model training. The graph on the left shows the accuracy of the model, while the graph on the right shows the loss values. Both graphs have an x-axis that ranges from 0 to 50, which may indicate the number of training epochs.

Figure 14. Inception model Accuracy-Loss graph

The blue lines on both graphs represent training data, while the orange lines represent validation data.

From the accuracy graph, it can be seen that the training accuracy increases steadily from 0.3 to 0.9, while the validation accuracy varies between 0.7 and 0.9. This shows that the model is able to learn patterns from training data well. And from the loss graph, it can be seen that the training loss value decreases steadily from 2.5 to 0.0, while the validation loss value varies between 1.0 and 1. This shows that the model is able to minimize prediction errors in the training data. Overall, this graph shows that the Inception model has been trained properly and is able to learn patterns from the training data.

From Figure 15, it can be concluded that the model can predict 39 sad images, 52 Surprised, 43 Angry, 47 Happy and 37 Bored images. The model is very accurate in predicting surprising images. This is evidenced by the second column which has a value of 52 and the other rows have a value of zero, but the difficulty in predicting bored images is evidenced by the fifth column, the fifth row which is smaller than the others, namely 37.

This graph shows the comparison between accuracy and loss on training and validation data during the training of a facial emotion classification model using ResNet.

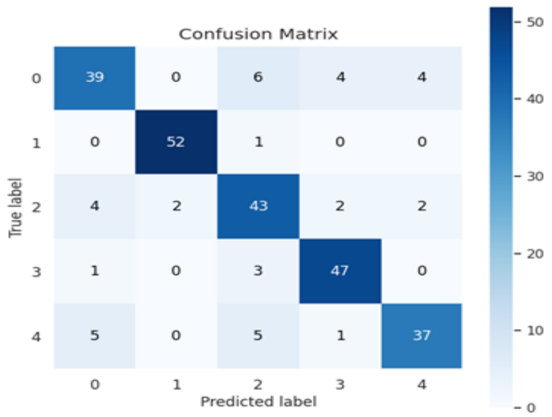

Figure 15. Confusion Matrix Model Inception

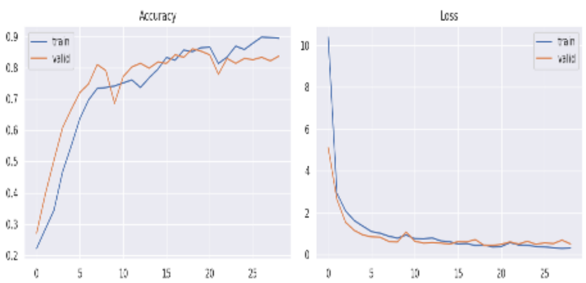

Figure 16. ResNet Model Accuracy-Loss Graph

The graph on the left shows that accuracy on training data increases steadily from 0.3 to 0.9, while accuracy on validation data increases from 0.3 to 0.7 and then decreases slightly. The graph on the right shows that the loss on the training data decreases steadily from 10 to 4, while the loss on the validation data decreases from 10 to 6 and then increases slightly. With the EarlyStopping callback, the training process stops at the 28th epoch because there has been no increase in validation accuracy for 10 consecutive epochs. This indicates that the model has reached its optimum point and will not improve further.

A confusion matrix is a table used to evaluate the performance of a machine learning model. This table has rows representing actual labels and columns representing predicted labels [32]. The numbers in the table indicate the number of events that fall into each category. Diagonal values represent the number of events correctly predicted by the model, while off-diagonal values represent the number of events predicted incorrectly by the model. Cell colors range from light blue to dark blue, with darker colors indicating higher values.

From Figure 17 it can be concluded that the model can correctly predict images of Sad 41, shocked, 52, angry 43, happy 46 and bored 40. The model is very accurate in predicting startled images. This is proven by the second column is worth 52 and the other row is zero.

Based on Figure 18, the ensemble model of three models, namely ResNet, Inception, and MobileNet, has been trained well. The accuracy graph shows that training and validation accuracy increases over time, and both are close to 1.0 at the end of training. The loss graph shows that the training and validation losses decrease with time, and both approach 0 at the end of the training. This shows that the model is able to learn patterns from training data and is also able to generalize well to validation data.

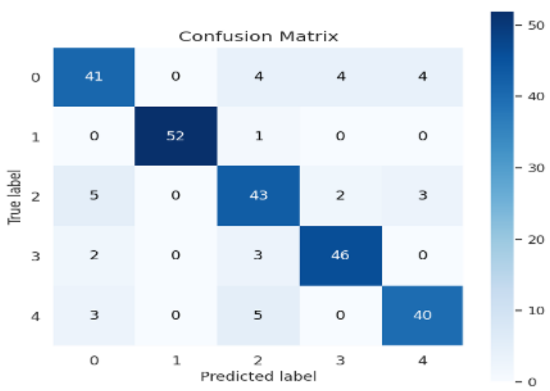

Figure 17. ResNet Confusion Model Matrix

Figure 18. Ensemble Model Accuracy-Loss Graph

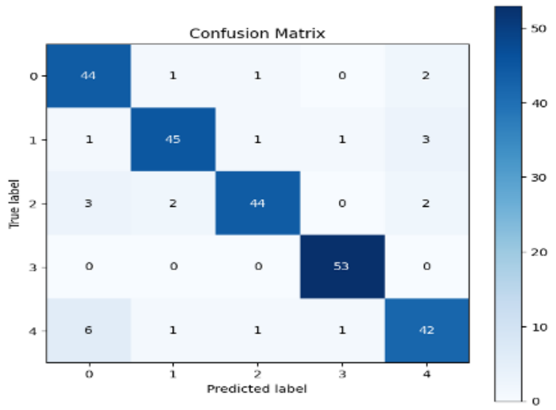

Figure 19. Confusion Matrix Ensemble Model

From Figure 19, it can be concluded that the model can predict 44 Bored images, 45 Happy, 44 Sad, 53 Surprised and 42 Angry. The model is very accurate in predicting surprising images. This is proven to be worth 53 but the difficulty in predicting the Angry image is worth 42 images. By providing an extensive comparative evaluation, future research will be able to evaluate the SOTA (state-of-the-art) approaches in this specific field.

CONCLUSION

Based on the results discussed earlier, it can be concluded that the ensemble approach has the best performance in recognizing and classifying students' facial emotions, namely by obtaining an accuracy of 90%. The ResNet model is also the individual model with the highest performance with an accuracy of 86%, while MobileNet and Inception still provide satisfactory results with an accuracy of 85% and 84%. The use of an ensemble approach or the ResNet model can be considered for implementation in an educational environment, taking into account the specific needs and available resources.

ACKNOWLEDGMENT

This research was supported by the Ministry of Education, Culture, Research, and Technology, Indonesia under Grant 685/UN11.2.1/PT.01.03/DPRM/2023.

REFERENCES

[1] A. Haleem, M. Javaid, M. A. Qadri, and R. Suman, “Understanding the role of digital technologies in education: A review,” *Sustainable Operations and Computers*, vol. 3, pp. 275–285, 2022, doi: 10.1016/j.susoc.2022.05.004.

[2] L. Dwyer, “Destination competitiveness and resident well-being,” *Tourism Management Perspectives*, vol. 43, pp. 100996, 2022, doi: 10.1016/j.tmp.2022.100996.

[3] J. C. Richards, “Exploring emotions in language teaching,” *Relc Journal*, vol. 53, no. 1, pp. 225-239, 2022, doi: 10.1177/0033688220927531.

[4] R. Miroshnyk, I. Matviy, I. Bahlai and L. Halaz, "Peculiarities of E-learning in Modern Conditions and Prospects for its Development," *2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT)*, Lviv, Ukraine, 2023, pp. 1-4, doi: 10.1109/CSIT61576.2023.10324039.

[5] P. Tarnowskiet al., “Emotion recognition using facial expressions,” *Procedia Computer Science*, vol. 108, pp. 1175-1184, 2017, doi: 10.1016/j.procs.2017.05.025.

[6] S. Li and W. Deng, “Deep facial expression recognition: A survey,” *IEEE Transactions on affective computing*, vol. 13, no. 3, pp. 1195-1215, 2020, doi: 10.1109/TAFFC.2020. 2981446.

[7] M. Z. I. Ahmed et al., “Automated feature extraction on AsMap for emotion classification using EEG,” *Sensors*, vol. 22, no. 6, pp. 2346, 2022, doi: 10.3390/s22062346.

[8] N. Ashraf et al., “Multi-label emotion classification of Urdu tweets,” *PeerJ Computer Science*, vol. 8, pp. e896, 2022, doi: 10.7717/peerj-cs.896.

[9] A. Keshri et al., “Automatic detection and classification of human emotion in real-time scenario,” *Journal of IoT in Social, Mobile, Analytics, and Cloud*, vol. 4, no. 1, pp. 41-53, 2022, doi: 10.36548/jismac.2022.1.005.

[10] Z. Li, F. Tang, M. Zhao, and Y. Zhu, “Emocaps: Emotion capsule based model for conversational emotion recognition,” *ArXiv preprint arXiv*: 2203.13504, 2022, doi: 10.48550/arXiv.2203.13504.

[11] V. R. R. Chirra, S. R. Uyyala, and V. K. K. Kolli, “Virtual facial expression recognition using deep CNN with ensemble learning,” *Journal of Ambient Intelligence and Humanized Computing*, pp.1-19, 2021, doi: 10.1007/s12652-020-02866-3.

[12] K. Raza, “Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule,” *In U-Healthcare Monitoring Systems,* pp. 179-196, Academic Press, 2019, doi: 10.1016/B978-0-12-815370-3.00008-6.

[13] H. Kim, H. Kim, H. Moon, and H. Ahn, “A weight-adjusted voting algorithm for ensembles of classifiers,” *Journal of the Korean Statistical Society*, vol. 40, no. 4, pp.437-449, 2011, doi: 10.1016/j.jkss. 2011.03.002.

[14] A. Dogan and D. Birant, "A Weighted Majority Voting Ensemble Approach for Classification," *2019 4th International Conference on Computer Science and Engineering (UBMK)*, Samsun, Turkey, 2019, pp. 1-6, doi: 10.1109/UBMK.2019.8907028.

[15] L. I. Kuncheva and J. J. Rodríguez, “A weighted voting framework for classifiers ensembles,” *Knowledge and information systems,* vol. 38, pp. 259-275, 2014, doi: 10.1007/s10115-012-0586-6.

[16] G. Brown and L. I. Kuncheva, “”Good” and “Bad” diversity in majority vote ensembles,” In: *Multiple Classifier Systems: 9th International Workshop,* MCS 2010, Cairo, Egypt, April 7-9, 2010. Proceedings 9. Springer Berlin Heidelberg, 2010. p. 124-133. doi: 10.1007/978-3-642-12127-2_13.

[17] T. B. Chandra et al., “Coronavirus disease (COVID-19) detection in chest X-ray images using majority voting based classifier ensemble,” *Expert systems with applications*, vol. 165, pp.113909, 2021, doi: 10.1016/j.eswa.2020.113909.

[18] O. A. Hassen et al., “Realistic Smile Expression Recognition Approach Using Ensemble Classifier with Enhanced Bagging.” *Computers, Materials & Continua*, vol. 70, no. 2, 2022, doi: 10.32604/cmc.2022.019125

[19] C. Liu, et al., “Multi-modal Expression Recognition with Ensemble Method,” *ArXiv preprint,* 2023, doi: 10.48550/arXiv.2303. 10033.

[20] A. Ozcift and A. Gulten, “Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms,” *Computer methods and programs in biomedicine,* vol. 104, no. 3, pp. 443-451, 2011, doi: 10.1016/j.cmpb.2011. 03.018.

[21] S. M. Rosfi et al., “Classification of Emotions of Characters in Novels Above the Pen You are Applying for Me by E. Shabila El Raihany and Its Contribution as Teaching Materials in High School (David Krech's Study of Literary Psychology),” *Journal of Asian Multicultural Research for Social Sciences Study*, vol. 3, no. 2, pp. 15-20, 2022, doi: 10.47616/jamrsss.v3i2.269.

[22] M. Z. Asghar et al., ”Sentence-Level Emotion Detection Framework Using Rule-Based Classification,” *Cognitive Computation,* vol. 9, pp. 868-894, 2017, doi: 10.1007/s12559-017-9503-3.

[23] P. Du et al., “Multiple Classifier System for Remote Sensing Image Classification: A Review,” *Sensors,* vol.12, no. 4, pp. 4764-4792, 2012, doi: 10.3390/s120404764.

[24] A. Kumar et al., "An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification," in IEEE *Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 31-40, Jan. 2017, doi: 10.1109/JBHI.2016.2635663.

[25] L. D. Nguyen, D. Lin, Z. Lin and J. Cao, "Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation," *2018 IEEE International Symposium on Circuits and Systems (ISCAS),* Florence, Italy, 2018, pp. 1-5, doi: 10.1109/ISCAS.2018.8351550.

[26] A. G. Howard et al., “Mobilenets: Efficient Convolutional Neural Networks For Mobile Vision Applications,” *ArXiv preprint, 2017,* doi: 10.48550/arXiv.1704.04861.

[27] H. Alaeddine and M. Jihene, “Deep residual network in network,” *Computational Intelligence and Neuroscience*, pp. 1-9, 2021, doi: 10.1155/2021/6659083.

[28] U. Zahoora et al., ”Zero-day Ransomware Attack Detection using Deep Contractive Autoencoder and Voting based Ensemble Classifier,” *Appl Intell*, vol. 52, pp. 13941–13960,2022. Doi: https://doi.org/10.1007/s10489-022-03244-6.

[29] M. Muhajir et al., "USK-FEMO: A Face Emotion Dataset using Deep Learning for Effective Learning," *2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE),* Banda Aceh, Indonesia, 2023, pp. 199-203, doi: 10.1109/COSITE60233.2023.10249834.

[30] S. Avidan and A. Shamir., “Seam carving for content-aware image resizing,” In *Seminal Graphics Papers: Pushing the Boundaries*,

2023, vol. 2, pp. 609-617, doi: 10.1145/3596711.3596776.

[31] Y. et al., “Ensemble Learning of Multiple Deep CNNs Using Accuracy-Based Weighted Voting for ASL Recognition,” *Applied Sciences,* vol. 12, no. 22, pp. 11766, 2022, doi: 10.3390/app122211766.

[32] M. Heydarian, T. E. Doyle and R. Samavi, "MLCM: Multi-Label Confusion Matrix," in *IEEE Access*, vol. 10, pp. 19083-19095, 2022, doi: 10.1109/ACCESS.2022.3151048.

[33] A. Rojarath & W. Songpan, “Probability-weighted voting ensemble learning for classification model,” *Journal of Advances in Information Technology*, vol. 11, no. 4, 2022, doi: 10.12720/jait.11.4.217-227.

[34] E. Tasci, C. Uluturk, and A. Ugur, ”A voting-based ensemble deep learning method focusing on image augmentation and preprocessing variations for tuberculosis detection.” *Neural Computing and Applications*, vol. 33, no. 22, pp. 15541-15555, 2021, doi: 10.1007/s00521-021-06177-2.

[35] F. Ülker and A. Ahmet Küçüker, “Probabilistic weighted voting model using multiple machine learning methods for fault detection and classification,” *COMPEL-The international journal for computation and mathematics in electrical and electronic engineering*, vol. 41, no. 5, pp. 1542-1565, 2022, doi: 10.1108/COMPEL-06-2021-0200.

[36] A. Irwanto and L. Goeirmanto. "Sentiment Analysis from Twitter about Covid-19 Vaccination in Indonesia Using Naïve Bayes and XGboost Classifier Algorithm,” *SINERGI,* vol. 27, no. 2, pp. 145-152, 2023, doi: 10.22441/sinergi.2023.2.001.

[37] M. A. Khadija and W. Nurharjadmo. "Enhancing Indonesian customer complaint analysis: LDA topic modelling with BERT embeddings," *SINERGI,* vol. 28, no. 1, pp. 153-162, 2024, doi: 10.22441/sinergi.2024. 1.015.