International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

ResNet-50 in a Mobile Application with Facial Expression Recognition for Teacher Assessment

Aji Gautama Putrada [1], Mahmud Dwi Sulistiyo [2,*], Donni Richasdy [3], Aditya Firman Ihsan [4]

[1,2,3,4] School of Computing, Telkom University, Jl. Telekomunikasi No. 1, Bandung 40287, Indonesia

[1] ajigps@telkomuniversity.ac.id, [2] mahmuddwis@telkomuniversity.ac.id,
[3] donnir@telkomuniversity.ac.id, [4] adityaihsan@telkomuniversity.ac.id
[*]corresponding author

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: Facial expression is one of the measurement metrics in teacher assessment because facial expression is a non-verbal aspect of communication, and communication is an important aspect of teaching. However, teacher assessment has never used a mobile application with facial expression recognition. Our research aims to develop a mobile facial expression recognition application for teacher assessment measurements with optimum inference time. The first step of our research was to obtain the Jonathan Oheix face expression recognition dataset from Kaggle, which has seven labels: 'angry,' 'disgust,' 'fear,' 'happy,' 'neutral,' 'sad,' and 'surprise.' This dataset is used with the ResNet-50 model for facial expression recognition. We have two comparison models, which are shallow learning methods, namely k-Nearest Neighbor (KNN) and Support Vector Machine (SVM); then, two other comparison models are pre-trained deep learning methods: MobileNetv2 and SE-ResNet-50. The metrics we compare are accuracy, inference time, and frame rate. The test results show that fear has the best recall value and neutral has the worst. Then, disgust has the best precision value, while fear has the worst. Happy is the label with the best F1-score with a value of 0.56. Compared with the SVM, KNN, SE-ResNet-50, and MobileNetV2 methods, ResNet-50 is the model with the best accuracy, 0.5314. ResNet-50 has a worse inference time and frame rate than MobileNetV2. However, the ResNet-50 frame rate of 946 fps is still above the frame rate considered good, namely 15 fps. Our research is the first facial expression recognition in teacher assessment that uses the ResNet-50 model on the Jonathan Oheix dataset and has a mobile application.

Keywords: ResNet-50, Facial Expression Recognition, Teacher Assessment, Mobile Application, Transfer Learning

1. Introduction

Teacher assessment is a necessary professional requirement in education in many countries [1]. There are several ways to measure metrics in teacher assessment, including learning intentions, asking questions, feedback, peer and self-assessment, and in-class assessment [2]. Apart from that, facial expression is also one of the measurement metrics in teacher assessment [3]. This metric is because facial expressions are one of the non-verbal aspects of communication, and communication is an important aspect of teaching [4]. Previous research used six basic expressions plus neutral expressions in assessing teachers using facial expressions. These expressions are disgust, surprise, joy, anger, fear, and sadness.

In computer vision, facial expression recognition is a sub-interest that uses machine learning and deep learning techniques to recognize facial expressions from images or videos [5]. Several studies have carried out facial expression recognition with shallow learning. Yadav *et al.* [6] used the Viola-Johns method with Support Vector Machine (SVM) and k-Nearest Neighbors (KNN) on four different databases. The results show that KNN is a superior method with the best accuracy of 0.96. Other studies have tried to use deep learning methods from pre-trained models. Li *et al.* [7] brought ResNet-50 for facial expression recognition and implemented it on their self-made dataset. The best accuracy in that research is 0.95. Hu *et al.* [8] used MobileNetv2 for facial expression recognition from three datasets. The best accuracy of the study was 0.89. Ngo *et al.* [9] used SE-ResNet-50, an extension of ResNet-50 for facial expression recognition on the AffectNet dataset, where the best accuracy was 0.61.

Furthermore, Dahri *et al.* [10] said that using real-time mobile apps in teacher assessment can improve learning outcomes. On the other hand, Bouhali *et al.* [11] mentioned that inference time is important for a machine learning model applied in an application, especially if the application runs in real-time. For example, Niu *et al.* [11] stated that the ResNet-50 implementation they developed on a mobile app had an inference time of 26 ms. Searching and comparing the optimum inference time from several machine learning models for facial expression recognition in teacher assessment is a research opportunity.

Our research aims to develop a mobile facial expression recognition application for teacher assessment measurements with optimum inference time. The first step of our research was to obtain the Jonathan Oheix face expression recognition dataset from Kaggle. We then developed a ResNet-50 model for facial expression recognition. We have two comparison models, which are shallow learning methods, namely KNN and SVM, and then two other comparison models are pre-trained deep learning methods: MobileNetv2 and SE-ResNet-50. The metrics we compare are accuracy, inference time, and frame rate.

To the best of our knowledge, no research has developed mobile apps for facial expression recognition in teacher assessment. The following is a list of our research contributions:

1) A facial expression recognition using transfer learning on the Jonathan Oheix dataset.
2) An optimal facial expression recognition for teacher assessment using the transfer learning model, ResNet-50.
3) A mobile application for facial expression recognition in teacher assessment, where the mobile application has optimum inference time and frame rate.

The remainder of this research is organized as follows: Section 2 discusses the latest research directly related to our research. Section 3 discusses the research methodology and theories related to our development. Section 4 displays our test results. This section closes with a discussion that compares our test results with existing research and emphasizes the contribution of our research. Finally, Section 5 presents the conclusions of our study.

2. Related Works

In this section, we discuss several related papers and provide several constraints in these related papers. The first constraint is that the papers we discuss are the latest papers published in the last five years. Then, the scope of several papers we discussed is regarding teacher assessment and its relationship to facial expressions. The second scope of our discussion concerns facial expression recognition methods and the datasets involved in each study. The third scope concerns the involvement of real-time mobile applications in facial expression recognition. The final scope is the use of transfer learning in facial expression recognition.

Several studies have discussed facial expressions in teacher assessment measurements. In the survey paper by Utami *et al.* [4], they expressed the importance of recognizing a teacher's facial expressions during assessment because the expression is one of the non-verbal aspects of communication, which is an important aspect of teaching. This research only surveys facial expression recognition methods and has no implementation.

Furthermore, other studies have tried to use the transfer learning method. Savchenko *et al.* [13] conducted facial expression recognition to measure enthusiasm in an e-learning system. They compared two transfer learning methods in this research, MobileNet and EfficientNet, where the dataset used was AffectNet. Li *et al.* [7] brought ResNet-50 for facial expression recognition and implemented it on their dataset. This research obtained the best accuracy, namely 0.95.

Several studies have developed real-time facial expression recognition. Lee *et al.* [14] created real-time facial expression recognition with a novel method called EmotionNet Nano. This research uses the CK+ dataset, where the frame rate of this method is 25 FPS at 15 watts of power. Yang *et al.* [15] created an edge device from a Raspberry Pi that can recognize facial expressions. Their novel model uses a new deep learning method that utilizes a facial action unit (AU) to detect atomic muscle movements. Their best run time is 66 seconds.

Apart from AffectNet and CK+, the Jonathan Oheix dataset from Kaggle is also a dataset for facial expression recognition, which has been used in several studies. Priyanka *et al.* [16] created facial expression recognition to organize music playlists automatically. This research uses the Jonathan Oheix dataset and the CNN method. Weladi *et al.* [17] also used CNN on the same dataset. This research can detect happy, neutral, and sad faces. There is a three-fold research opportunity:

1) There has never been any research that has done facial expression recognition for teacher assessment using ResNet-50.
2) There has never been research on real-time facial expression recognition using ResNet-50 with mobile application implementation.
3) There has never been research on facial expression recognition using the Jonathan Oheix dataset and real-time mobile application implementation.

TABLE I. compares all the research discussed in this chapter and highlights our contribution.

TABLE I. RELATED WORKS ON FACIAL EXPRESSION RECOGNITION FOR TEACHER ASSESSMENT WITH MOBILE APPLICATION.

Reference	Teacher Assessment	Facial Expression Recognition	ResNet-50	Real-Time	Jonathan Oheix Dataset
Utami *et al.* [4]	✓	✗	✗	✗	✗
Savchenko *et al.* [13]	✓	✓	✗	✗	✗
Li *et al.* [7]	✗	✓	✓	✗	✗
Lee *et al.* [14]	✗	✓	✗	✓	✗
Yang *et al.* [15]	✗	✓	✗	✓	✗
Priyanka *et al.* [16]	✗	✓	✗	✗	✓
Weladi *et al.* [17]	✗	✓	✗	✗	✓
Proposed Method	✓	✓	✓	✓	✓

3. METHOD

We provide a way to accomplish our study goals. Getting the Jonathan Oheix face expression recognition dataset from Kaggle was the initial step in our investigation. Next, we created a face expression recognition model called ResNet-50. Two of our comparison models, KNN and SVM, are shallow learning techniques. The other two, MobileNetv2 and SE-ResNet-50, are pre-trained deep learning techniques. We compare three metrics: frame rate, inference time, and correctness. A block diagram of our study method's system flow is shown in Figure 1. .

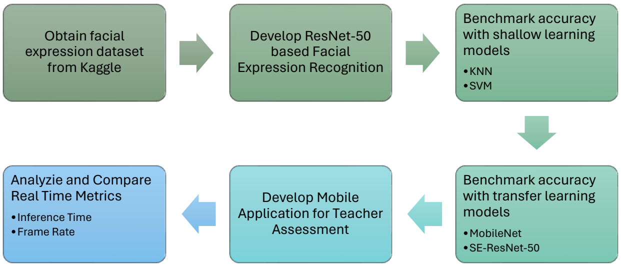

Obtain facial expression dataset from Kaggle
Develop ResNet-50 based Facial Expression Recognition
Benchmark accuracy with shallow learning models
•KNN
•SVM
Benchmark accuracy with transfer learning models
•MobileNet
•SE-ResNet-50
Develop Mobile Application for Teacher Assessment
Analyzie and Compare Real Time Metrics
•Inference Time
•Frame Rate

Figure 1. The workflow of our proposed research.

A. Facial Expression Dataset and Pre-Processing

We obtained Jonathan Oheix's facial expression dataset from Kaggle [18]. The data was uploaded in 2019, and the facial expression image was 48×48 pixels in size and grayscale in color [19]. There are seven subfolders in two folders in the dataset, where each subfolder represents six basic expressions plus one neutral expression: Happy, sad, angry, surprised, fearful, and disgust, while two folders represent the train and validation data [20]. There are 35,887 images in the dataset, of which 28,821 are training data, and 7,066 are validation data. The angry label has 3,993 images; disgust has 436 images; fear has 4,103 images; happy has 7,164 images; neutral has 4,982 images; sad has 4,938 images; and surprise has 3,205 images. Figure 2. is an example of an image in our dataset, an expression with a neutral label.

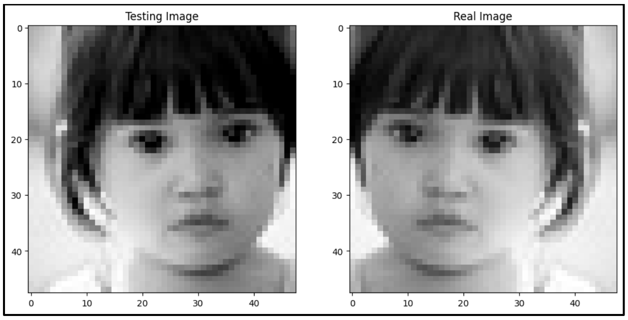


Figure 2. Example image in the facial expression dataset courtesy of Jonathan Oheix from Kaggle. The image shows a neutral expression.

We use several pre-processing stages before entering the facial expression recognition stage. Data augmentation in computer vision is a step to increase the quantity and quality of images in a dataset so that machine learning methods have better performance and are more generalizable [21]. We use advanced blur, which makes the image blurrier, thereby increasing the generalizability of the prediction model. The rotation process allows the model to predict objects from different points of view. *Random brightness contrast* is a data augmentation method that changes the brightness and contrast of the image, increasing the variation in the image. Finally, horizontal flip-type data augmentation allows the model to make predictions without considering the orientation of the image [22].

We use other pre-processing methods, such as using an image loader, dividing the dataset for training, and randomizing the data. The image loader can speed up the training process [23]. We divide the dataset for validation and testing, where the test dataset ensures the robustness of our model. Randomizing the data before training can help the training process achieve convergence, both in convex and non-convex cases [24].

B. Facial Expression Recognition with ResNet-50

Facial expression recognition is a machine learning method that recognizes people's expressions from images or videos and classifies them into expressions [25]. Several studies use the six basic expressions coined by Paul Ekman and Wallace Friesan to classify expressions: happy, sad, angry, disgust, fear, and surprise [26]. Apart from teacher assessment, other implementations of facial expression recognition include video games, suspicious people detection, hospital patient pain measurement, and online meetings [27].

We propose facial expression recognition with ResNet-50 as a transfer learning method. ResNet-50 has obtained weights from training on a general image dataset with many images [28]. Transfer learning is a method of re-training a model trained on a large dataset to a new one with a distinct case study [29]. The specificity of ResNet-50 is that this model has a skip connection, which allows it to skip a layer to the next layer. Furthermore, ResNet-50 has four stages of convolutional and pooling layers, a bottleneck architecture for deeper networks, and a fully linked layer consisting of 1,000 neurons [30].

ResNet-50 was named so because it has 50 residual layer networks [31]. One of the characteristics of ResNet-50 is the use of residual blocks, where this model can learn from residuals due to the identity mapping [32]. The following is the formula for the identity block, which is part of the identity mapping:

$$Input \xrightarrow{Conv} Output = Input\ +\ Conv(Input) \qquad (1)$$

Alternatively, if F is considered a residual function:

$$Output = Input\ +\ F(Input) \qquad (2)$$

Furthermore, the convolutional block is the part where the input and output dimensions are different due to changes in filter size or spatial dimensions. The formula for the convolutional block in ResNet-50 is as follows:

$$Input \xrightarrow{Conv1\times1} Conv(Input) \xrightarrow{Conv3\times3} Output \qquad (3)$$

Alternatively, the form of the mathematical formula is as follows:

$$Output = Conv3\times3(Conv1\times1(Input))\ +\ Shortcut(Input) \qquad (4)$$

Finally, here is a simplified form of the overall ResNet-50 architecture:

$$Input \xrightarrow{Conv7\times7/64} MaxPool3\times3/2 \xrightarrow{Residual} AveragePool \xrightarrow{Fully\ Connected} \qquad (5)$$

In our architecture, we combine ResNet-50 with a fully connected layer consisting of three dense layers, the first consisting of 4,096 neurons, the second consisting of 512 neurons, and the last according to our number of classes, namely seven. Before the dense layer, there is a flattened layer, which makes it easier to change dimensions between layers. Then, there is a dropout layer between the two layers with a dropout value of 0.5 each. The output layer uses Softmax for the activation function. The other layers use a linear activation function. We use Adam as an optimizer in training. Figure 3. shows our ResNet-50 architecture, combining pre-trained and fully connected parts.

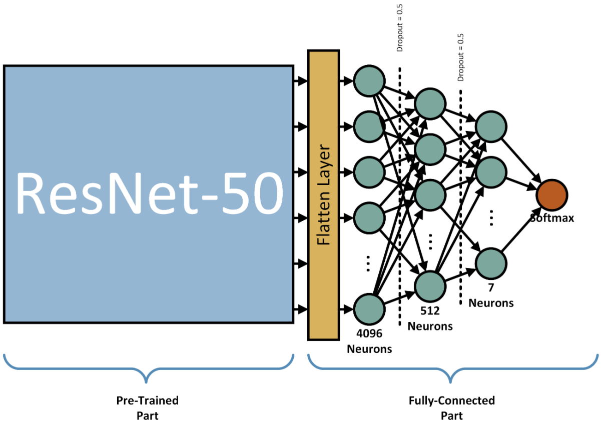


Figure 3. The proposed ResNet-50 architecture for facial expression recognition.

We compare our ResNet-50 architecture with several other models, including two shallow learning models, SVM and KNN. SVM performs classification by looking for a hyperplane in the form of a linear plane that maximizes the separation of two data in feature space [33]. The data closest to the hyperplane has a vital role and is called a support vector [34]. If the data is not linearly separable, the SVM will perform a kernel trick, transforming the data to a higher dimension with one of the kernels: radial basis function (RBF), sigmoid, or polynomial [35]. Even though it is shallow learning, SVM can perform image recognition by learning discriminative features [36]. KNN is a shallow learning method that uses the concept of distance between data in feature space to make decisions [37]. The k value in KNN functions to determine how many nearest neighbors of new data are considered when making decisions [38]. KNN is called a lazy learner because it does not undergo a training process, but all the training data becomes part of the prediction model [39].

Apart from several shallow learning methods, we benchmarked the ResNet-5 method with two other transfer learning methods, MobileNetV2 and SE-ResNet-50. Like ResNet-5, MobileNetV2 is a modification of CNN for the image classification [40]. MobileNetV2 is a lightweight model developed by Google that is made to run on mobile devices [41]. MobileNetV2 is pre-trained on large-scale datasets such as ImageNet, which contains millions of images and thousands of categories [42]. MobileNetV2 can be lightweight because it uses fewer parameters to predict images [43]. On the other hand, SE-ResNet-50 is a modification of ResNet-50 that implements a “squeeze-and-excitation” [44]. Squeeze means the dimension reduction process by applying global average pooling. Squeeze is used to carry out recalibration after the excitation process, multiplying the feature map to emphasize important features and suppress unimportant ones [45]. With these features, SE-ResNet-50 often outperforms ResNet-50 in several case studies [46].

C. Mobile Application for Teacher Assessment

Teacher assessment is an activity that evaluates a teacher's teaching suitability, including aspects such as ability, knowledge, and effectiveness of the teacher in the classroom [47]. Several types of teacher assessment are self-assessment, formative assessment, and summative assessment, where self-assessment is a type that allows teachers to assess themselves [48]. On the other hand, formative assessment is an assessment of a teacher while the teacher is in the teaching process [49]. Finally, summative assessment is an assessment of a teacher when the teacher has gone through the teaching process and is carried out at the end of a process [50]. Based on these understandings, the facial expression recognition we developed for teacher assessment will be an example of the application of formative assessment.

Facial expression recognition can automatically assess the teacher's emotions while teaching, improving the learning atmosphere and saving time [51]. Happy and surprise can be categorized as positive expressions of the six basic expressions [52], [53]. Meanwhile, sadness, anger, fear, and disgust can be categorized as negative expressions [54]. Then, positive expressions that represent positive emotions can be interpreted as good engagement and a form of good learning environment [55]. On the other hand, negative expressions can be interpreted as poor classroom management, difficulty in mastering the material, or issues regarding the well-being of the teacher who teaches [56], [57], [58]. In the end, the mobile application that we designed for teacher assessment can record teachers' faces during teaching practice and then classify their facial expressions, categorize the polarity of facial expressions, interpret the teacher's abilities and the teaching environment, and convey this in a formative assessment [59]. Figure 4. shows the explanation that we convey in the form of a flow chart.

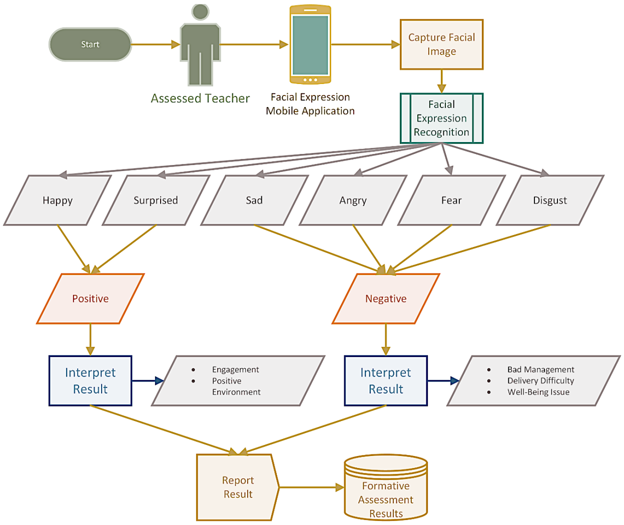


Figure 4. The workflow of the facial expression recognition mobile application on teacher assessment.

Our mobile application is real-time because it directly involves computer vision of the teacher's face [60]. We use TensorFlow Lite for our mobile application development, where TensorFlow Lite has an extension so that the deep learning developed can be deployed on iOS and AndroidOS [61]. Then, in this research, we use inference time and frame rate as metrics for measuring the designed real-time mobile application. Frame rate is the number of frames processed in a second [62]. Several studies reveal that the threshold limit considered good for frame rate values is 15 fps, while a frame rate of 10 fps is considered acceptable [63]. On the other hand, inference time is required for deep learning to predict an image that has never been processed [64].

4. RESULTS AND DISCUSSION

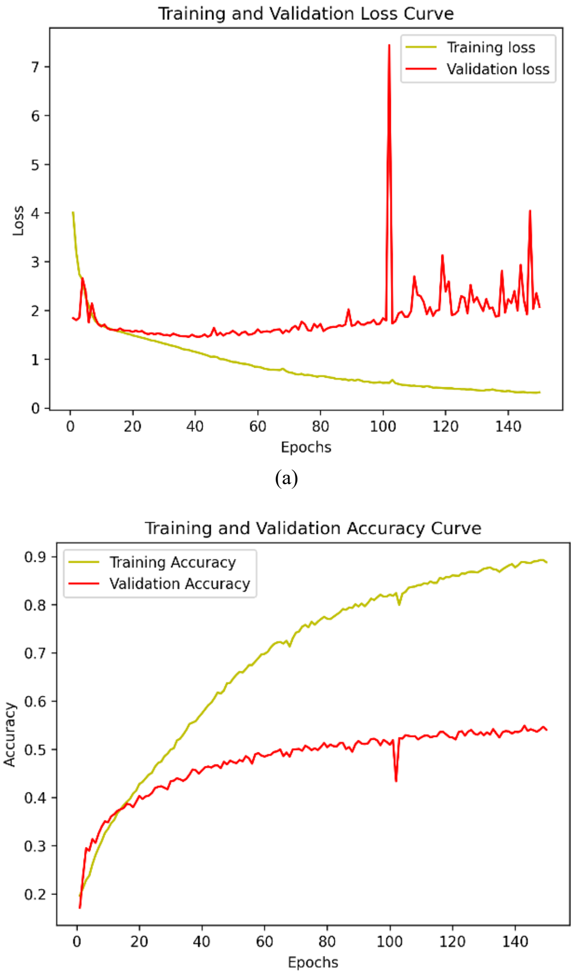
A. Results

We download facial expression data from Kaggle and then do pre-processing. There are 28,821 images for training, then 7,066 images for validation. From some validation data, we took some for testing so that there were 6,716 images for validation and then 350 images for testing. We ensure that each folder has seven labels, namely ‘angry,’ ‘disgust,’ ‘fear,’ ‘happy,’ ‘neutral,’ ‘sad,’ and ‘surprise.’ We use NVIDIA-SMI 535.104.05 with the Tesla T4 GPU provided by Google Colab for training.

In carrying out the ResNet-50 training process, we observed several hyperparameters that significantly influence the training curve performance. The first hyperparameter is the batch size per epoch; the more the batch size increases, the better the learning curve. However, this hyperparameter is inversely proportional to training time, so finding the optimal batch size per epoch is necessary. The optimal batch size per epoch for us is 120.

The second hyperparameter is the epoch, the same as the batch size per epoch. A higher epoch value also improves the quality of the learning curve. The optimum epoch value is 150. The third hyperparameter is the learning rate of the Adam optimizer—the smaller the optimizer’s learning rate, the higher the learning curve quality. The optimum value of the learning rate is 10-4. The fourth factor is validation samples; the more validation samples, the better the learning curve. The optimum validation sample value is 3,500, with training samples of 28,000.

With these four optimal hyperparameter values, the average time required per epoch is 36s, and the time required to train the ResNet-50 model with initial hardware specifications is one hour, 30 minutes, and 40.2 seconds. The final training and validation loss in the training process is 0.32 and 2.07, respectively. Meanwhile, the final training and validation accuracy in the training process were 0.89 and 0.54, respectively.



(a)

(b)

Figure 5. The ResNet-50 training and validation loss and accuracy curve.

We tested our ResNet-50 model on 350 images for testing. As a result, we got an accuracy of 0.5314. Figure 6. shows the confusion matrix of the test results, where the diagonal elements represent the classes that were predicted correctly in the confusion matrix. The 'happy' label has the best prediction performance, with 662 data predicted correctly, while 'angry' has the worst prediction performance, with 227 data predicted correctly. The joint highest out-misclassification happens between 'disgust' and 'angry,' where 12 'disgust' labels are predicted as 'angry,' also between 'neutral' and 'sad,' where 124 'neutral' labels are predicted as 'sad.' At the same time, the highest in-misclassification is also the occurrence between 'neutral' and 'sad.'

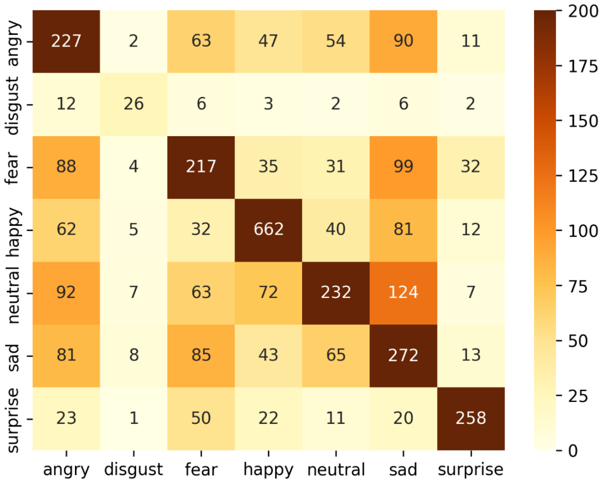


Figure 6. Confusion matrix of our proposed ResNet-50 on the facial expression dataset.

The performance of each label in the confusion matrix can be analyzed further using several metrics, namely precision, recall, and F1-score. Figure 7. shows a bar chart that compares each class's performance in the facial expression recognition by ResNet-50 that we are implementing. In the bar chart, 'surprise' has the best precision, meaning that the label receives the fewest wrong labels from other labels compared to other labels. On the other hand, 'angry' has the worst precision. At the same time, 'happy' has the best recall, meaning that the label was predicted correctly the most. Then, 'angry' has the worst recall. Finally, the f1-score shows the label with the most balanced precision and recall values, whereas the best f1-score belongs to the 'happy' label with a value of 0.75. The 'happy' label has a precision of 0.75 and a recall of 0.74.

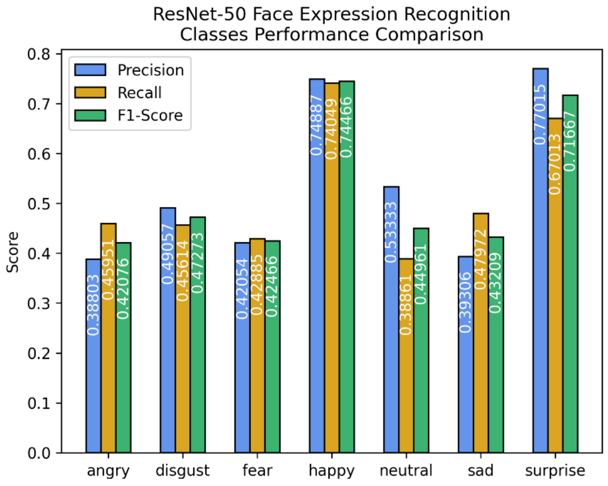
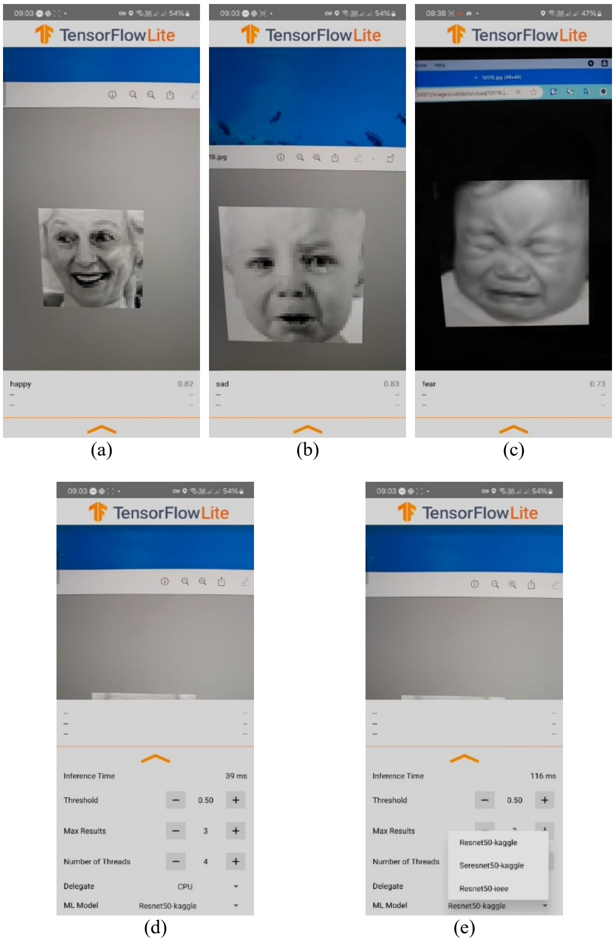


Figure 7. A bar chart that displays a comparison of the performance of each class in facial expression recognition by ResNet-50 that we are carrying.

Figure 8. shows screenshots of a mobile application for facial expression recognition in teacher assessment that we successfully developed with TensorFlow Lite. Figure 8. (a)–(c) shows when we direct the application to the dataset images, among which are happy, sad, and afraid faces. The number value to the right is the output value of the Softmax activation function, which ranges from 0.00 to 1.00. Figure 8. (d) shows the settings menu we created, including setting the Softmax threshold, maximum desired results, the maximum number of threads to run deep learning-based recognition, selecting the type of accelerator, and selecting the decision-making model. Lastly, Figure 8. (e) shows the choice of deep learning models for facial expression recognition, including ResNet-50 and SE-ResNet-50.



happy 0.82

sad 0.83

fear 0.73

(a) (b) (c)

Inference Time 39 ms
Threshold − 0.50 +
Max Results − 3 +
Number of Threads − 4 +
Delegate CPU
ML Model Resnet50-kaggle

Inference Time 116 ms
Threshold − 0.50 +
Max Results
Number of Threads
Delegate
ML Model Resnet50-kaggle
Resnet50-kaggle
Seresnet50-kaggle
Resnet50-ieee

(d) (e)

Figure 8. Screenshots of the developed mobile application of facial expression recognition for teacher assessment.

B. Discussion

Paper [6] has carried out facial expression recognition using SVM and KNN. On the other hand, ResNet-50 is a deep learning method that has feature learning capabilities. ResNet-50 should have better capabilities than SVM and KNN. We tried SVM and KNN on the Jonathan Oheix dataset we used in this research. TABLE II. compares ResNet-50, which we proposed, with SVM and KNN as state-of-the-art shallow learning methods in facial expression recognition. In terms of accuracy, ResNet-50 has a better value, namely 0.5314. At the same time, KNN has the worst performance, with an accuracy of 0.2543.

TABLE II. RESNET-50 PERFORMANCE COMPARISON WITH STATE-OF-THE-ART SHALLOW LEARNING METHODS ON FACIAL EXPRESSION RECOGNITION.

Model	Accuracy	Inference Time (s)	Frame Rate (fps)
kNN [6]	0.2543	1.1847	295
SVM [6]	0.3086	25.2765	14
Proposed Method	**0.5314**	**0.3698**	**946**

Furthermore, regarding inference time, ResNet-50 has a better value than SVM and KNN, with a value of 0.3698 seconds. SVM has the worst inference time, 25.2765 seconds, 68 times slower than ResNet-50. Kramer *et al.* [65] said that SVM takes a long time because it has a time complexity of O(sv), where sv is the number of support vectors. Lastly, ResNet-50 has the best frame rate with a value of 946 fps. KNN is the method with the second-best frame rate with 295 fps. The SVM frame rate is below the good value but still acceptable at 14 fps.

Several transfer learning methods have been used for facial expression recognition, namely MobileNetV2 [8] and SE-ResNet-50 [9]. Here, we compare these methods' use with the proposed method, namely ResNet-50. We implemented MobileNetV2 and SE-ResNet-50 on the Jonathan Oheix dataset we used in this study. TABLE III. shows a comparison between the ResNet-50 that we are promoting with MobileNetV2 and SE-ResNet-50 as state-of-the-art transfer learning methods in facial expression recognition. Regarding accuracy, our proposed method outperforms MobileNetV2 and SE-ResNet-50 from training, validation, and testing. However, regarding the number of parameters involved, model size, inference time, and frame rate, MobileNetV2 is the superior method. Lin *et al.* [43] said that MobileNetV2 is lightweight because it uses small amounts of parameters. That knowledge could be the reason why ResNet-50 has better performance than MobileNetv2, but MobileNetv2 has a faster inference time and fewer parameters.

On the other hand, Chen *et al.* [63] stated that a frame rate of 15 fps is considered a good frame rate. Because the ResNet-50 frame rate is still above that value, even though it is not the best, ResNet-50 is still the optimal method in terms of accuracy, inference time, and frame rate. These state-of-the-art methods have never been tried on the Jonathan Oheix dataset, so facial expression recognition using transfer learning on the Jonathan Oheix dataset is a research contribution.

TABLE III. COMPARISON OF STATE-OF-THE-ART TRANSFER LEARNING METHODS ON FACIAL EXPRESSION RECOGNITION, WHERE THE PROPOSED METHOD IS RESNET-50.

Model	Accuracy			Parameters			Inference Time (s)	Frame Rate (fps)
	Training	*Validasi*	*Testing*	*Trainable*	*Non-trainable*	*Total*		
MobileNetV2 [8]	0.2364	0.2514	0.1429	**25,300,743**	**34,112**	**25,334,855**	**0.1705**	**2052**
SE-ResNet-50 [9]	0.4962	0.3119	0.2743	61,698,807	53,120	61,751,927	0.3699	946
Proposed Method	**0.7799**	**0.5954**	**0.5314**	59,160,266	45,574	59,205,840	0.3698	946

Other research, such as paper [13], has carried out facial expression recognition for teacher assessment using MobileNetV5. On the other hand, paper [7] carried out facial expression recognition with ResNet-50, but not for teacher assessment, and neither implemented a real-time mobile application. So, the contribution of our research is two-fold. First, we propose an optimal facial expression recognition for teacher assessment using a transfer learning model, ResNet-50. Second, we propose a mobile application for facial expression recognition in teacher assessment, where the mobile application has optimum inference time and frame rate.

ACKNOWLEDGMENT

This paper is an output and part of research entitled 'Facial Expression Recognition for Teacher Assessment,' which received a research grant from the Directorate of Research and Community Service at Telkom University, Indonesia, under the Internal Partnership Grant scheme, Grant Number PPM/Tel-U/XX/2023. We want to thank our research partner, Festivo.

REFERENCES

[1] C. DeLuca, D. LaPointe-McEwan, and U. Luhanga, "Teacher assessment literacy: a review of international standards and measures," *Educ. Assess. Eval. Account.*, vol. 28, no. 3, pp. 251–272, Aug. 2016, doi: 10.1007/s11092-015-9233-6.

[2] Z. Lysaght, M. O'Leary, and L. Ludlow, "Measuring Teachers' Assessment for Learning (AfL) Classroom Practices in Elementary Schools," *Int. J. Educ. Methodol.*, vol. 3, no. 2, pp. 103–115, Dec. 2017, doi: 10.12973/ijem.3.2.103.

[3] X.-Y. Tang, W.-Y. Peng, S.-R. Liu, and J.-W. Xiong, "Classroom Teaching Evaluation Based on Facial Expression Recognition," in *Proceedings of the 2020 9th International Conference on Educational and Information Technology*, Oxford United Kingdom: ACM, Feb. 2020, pp. 62–67. doi: 10.1145/3383923.3383949.

[4] P. Utami, R. Hartanto, and I. Soesanti, "A Study on Facial Expression Recognition in Assessing Teaching Skills: Datasets and Methods," *Procedia Comput. Sci.*, vol. 161, pp. 544–552, Jan. 2019, doi: 10.1016/j.procs.2019.11.154.

[5] Abutalib K, Amandeep Gautam, Amandeep Gautam, Amandeep Gautam, and Aditya Dayal Tyagi, "Facial Expression Based Music Recommendation System," *Int. J. Adv. Res. Sci. Commun. Technol.*, pp. 316–325, Apr. 2023, doi: 10.48175/IJARSCT-9046.

[6] K. S. Yadav and J. Singha, "Facial expression recognition using modified Viola-John's algorithm and KNN classifier," *Multimed. Tools Appl.*, vol. 79, no. 19, pp. 13089–13107, May 2020, doi: 10.1007/s11042-019-08443-x.

[7] B. Li and D. Lima, "Facial expression recognition via ResNet-50," *Int. J. Cogn. Comput. Eng.*, vol. 2, pp. 57–64, 2021.

[8] L. Hu and Q. Ge, "Automatic facial expression recognition based on MobileNetV2 in Real-time," in *Journal of Physics: Conference Series*, IOP Publishing, 2020, p. 022136. Accessed: Mar. 08, 2024. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1549/2/022136/meta

[9] Q. T. Ngo and S. Yoon, "Facial expression recognition based on weighted-cluster loss and deep transfer learning using a highly imbalanced dataset," *Sensors*, vol. 20, no. 9, p. 2639, 2020.

[10] N. A. Dahri, M. S. Vighio, Omar A. Alismaiel, and Waleed Mugahed Al-Rahmi, "Assessing the Impact of Mobile-Based Training on Teachers' Achievement and Usage Attitude," *Int. J. Interact. Mob. Technol. IJIM*, vol. 16, no. 09, pp. 107–129, May 2022, doi: 10.3991/ijim.v16i09.30519.

[11] N. Bouhali, H. Ouarnoughi, S. Niar, and A. A. El Cadi, "Execution Time Modeling for CNN Inference on Embedded GPUs," in *Proceedings of the 2021 Drone Systems Engineering and Rapid Simulation and Performance Evaluation: Methods and Tools Proceedings*, Budapest Hungary: ACM, Jan. 2021, pp. 59–65. doi: 10.1145/3444950.3447284.

[12] W. Niu, X. Ma, Y. Wang, and B. Ren, "26ms Inference Time for ResNet-50: Towards Real-Time Execution of all DNNs on Smartphone." arXiv, May 02, 2019. doi: 10.48550/arXiv.1905.00571.

[13] A. V. Savchenko, L. V. Savchenko, and I. Makarov, "Classifying emotions and engagement in online learning based on a single facial expression recognition neural network," *IEEE Trans. Affect. Comput.*, vol. 13, no. 4, pp. 2132–2143, 2022.

[14] J. R. Lee, L. Wang, and A. Wong, "Emotionnet nano: An efficient deep convolutional neural network design for real-time facial expression recognition," *Front. Artif. Intell.*, vol. 3, p. 609673, 2021.

[15] J. Yang, T. Qian, F. Zhang, and S. U. Khan, "Real-time facial expression recognition based on edge computing," *IEEE Access*, vol. 9, pp. 76178–76190, 2021.

[16] S. S. Priyanka, P. Jayahladini, M. S. Shankar, and S. T. Sri, "FACE DETECTION TO RECOGNIZE MOOD AND SUGGEST SONGS ACCORDINGLY", Accessed: Mar. 08, 2024. [Online]. Available: https://www.ijetcse.com/admin/uploads/FACE%20DETECTION%20TO%20RECOGNIZE%20MOOD%20AND%20SUGGEST%20SONGS%20ACCORDINGLY_1626164258.pdf

[17] P. Weladi, S. Wattamwar, P. Wardhe, N. Wategaonkar, A. Yadav, and A. Wankhede, "FACIAL EXPRESSION DETECTION SYSTEM," *Enhancing Product. Hybrid Mode Begin. New Era*, p. 51.

[18] P. K. Sidhu, A. Kapoor, Y. Solanki, P. Singh, and D. Sehgal, "Deep Learning Based Emotion Detection in an Online Class," in *2022 IEEE Delhi Section Conference (DELCON)*, Feb. 2022, pp. 1–6. doi: 10.1109/DELCON54057.2022.9752940.

[19] M. J. Awan, A. Raza, A. Yasin, H. M. F. Shehzad, and I. Butt, "The Customized Convolutional Neural Network of Face Emotion Expression Classification," *Ann. RSCB*, vol. 25, no. 6, pp. 5296–5304, 2021.

[20] S. Harish, V. P. Rathish Kumar, and P. Tharun Raj, "REAL-TIME FACE EMOTION RECOGNITION USING DEEP LEARNING", Accessed: Mar. 08, 2024. [Online]. Available: https://www.irjmets.com/uploadedfiles/paper/issue_6_june_2022/26991/final/fin_irjmets1656337411.pdf

[21] K. Nanthini, D. Sivabalaselvamani, K. Chitra, P. Gokul, S. KavinKumar, and S. Kishore, "A Survey on Data Augmentation Techniques," in *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India: IEEE, Feb. 2023, pp. 913–920. doi: 10.1109/ICCMC56507.2023.10084010.

[22] S. AbuSalim, N. Zakaria, N. Mokhtar, S. A. Mostafa, and S. J. Abdulkadir, "Data Augmentation on Intra-Oral Images Using Image Manipulation Techniques," in *2022 International Conference on Digital Transformation and Intelligence (ICDI)*, Kuching, Sarawak, Malaysia: IEEE, Dec. 2022, pp. 117–120. doi: 10.1109/ICDI57181.2022.10007158.

[23] I. Ofeidis, D. Kiedanski, and L. Tassiulas, "An Overview of the Data-Loader Landscape: Comparative Performance Analysis," 2022, doi: 10.48550/ARXIV.2209.13705.

[24] Q. Meng, W. Chen, Y. Wang, Z.-M. Ma, and T.-Y. Liu, "Convergence Analysis of Distributed Stochastic Gradient Descent with Shuffling." arXiv, Sep. 29, 2017. doi: 10.48550/arXiv.1709.10432.

[25] M. Ahmad *et al.*, "Facial expression recognition using lightweight deep learning modeling," *Math. Biosci. Eng.*, vol. 20, no. 5, pp. 8208–8225, 2023, doi: 10.3934/mbe.2023357.

[26] M. Murtaza, M. Sharif, M. AbdullahYasmin, and T. Ahmad, "Facial expression detection using Six Facial Expressions Hexagon (SFEH) model," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, USA: IEEE, Jan. 2019, pp. 0190–0195. doi: 10.1109/CCWC.2019.8666602.

[27] A. Patwal, M. Diwakar, A. Joshi, and P. Singh, "Facial expression recognition using DenseNet," in *2022 OITS International Conference on Information Technology (OCIT)*, Bhubaneswar, India: IEEE, Dec. 2022, pp. 548–552. doi: 10.1109/OCIT56763.2022.00107.

[28] K. Balavani, D. Sriram, M. B. Shankar, and D. S. Charan, "An Optimized Plant Disease Classification System Based on Resnet-50 Architecture and Transfer Learning," in *2023 4th International Conference for Emerging Technology (INCET)*, Belgaum, India: IEEE, May 2023, pp. 1–5. doi: 10.1109/INCET57972.2023.10170368.

[29] L. Zhang, Y. Bian, P. Jiang, and F. Zhang, "A Transfer Residual Neural Network Based on ResNet-50 for Detection of Steel Surface Defects," *Appl. Sci.*, vol. 13, no. 9, p. 5260, Apr. 2023, doi: 10.3390/app13095260.

[30] S. S., T. R., A. S., and Y. P. S., "ResNet50 Architecture Based Dog Breed Identification Using Deep Learning," *Appl. Comput. Eng.*, vol. 2, no. 1, pp. 300–308, Mar. 2023, doi: 10.54254/2755-2721/2/20220651.

[31] P. Nagpal, S. A. Bhinge, and A. Shitole, "A Comparative Analysis of ResNet Architectures," in *2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, Bangalore, India: IEEE, Dec. 2022, pp. 1–8. doi: 10.1109/SMARTGENCON56628.2022.10083966.

[32] S. A. Agrawal, V. D. Rewaskar, R. A. Agrawal, S. S. Chaudhari, Y. Patil, and N. S. Agrawal, "Advancements in NSFW Content Detection: A Comprehensive Review of ResNet-50 Based Approaches," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 4, pp. 41–45, 2023.

[33] B. A. Fadillah, A. G. Putrada, and M. Abdurohman, "A Wearable Device for Enhancing Basketball Shooting Correctness with MPU6050 Sensors and Support Vector Machine Classification," *Kinet. Game Technol. Inf. Syst. Comput. Netw. Comput. Electron. Control*, 2022.

[34] B. H. Farizan, A. G. Putrada, and R. R. Pahlevi, "Analysis of Support Vector Regression Performance in Prediction of Lettuce Growth for Aeroponic IoT Systems," in *2021 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, IEEE, 2021, pp. 1–6. Accessed: Mar. 10, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9702093/

[35] A. G. Putrada, N. Alamsyah, M. N. Fauzan, and S. F. Pane, "NS-SVM: Bolstering Chicken Egg Harvesting Prediction with Normalization and Standardization," *JUITA J. Inform.*, vol. 11, no. 1, pp. 11–18, 2023.

[36] A. G. Putrada, N. Alamsyah, M. N. Fauzan, and D. Perdana, "PCA-SVM for a Lightweight ASL Hand Gesture Image Recognition," in *2023 International Conference on Electrical Engineering and Informatics (ICEEI)*, IEEE, 2023, pp. 1–6. Accessed: Mar. 09, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10346744/

[37] A. G. Putrada, M. Abdurohman, D. Perdana, and H. H. Nuha, "Q8KNN: A Novel 8-Bit KNN Quantization Method for Edge Computing in Smart Lighting Systems with NodeMCU," in *Intelligent Systems and Applications*, vol. 824, K. Arai, Ed., in Lecture Notes in Networks and Systems, vol. 824. , Cham: Springer Nature Switzerland, 2024, pp. 598–615. doi: 10.1007/978-3-031-47715-7_41.

[38] A. G. Putrada, M. Abdurohman, D. Perdana, and H. H. Nuha, "EdgeSL: Edge-Computing Architecture on Smart Lighting Control With Distilled KNN for Optimum Processing Time," *IEEE Access*, vol. 11, pp. 64697–64712, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3288425.

[39] F. Ghassani, M. Abdurohman, and A. G. Putrada, "Prediction of smarthphone charging using k-nearest neighbor machine learning," in *2018 Third International Conference on Informatics and Computing (ICIC)*, IEEE, 2018, pp. 1–4.

[40] A. Hadi, R. R. Pahlevi, and A. G. Putrada, "Office Room Smart Lighting Control with Camera and SSD MobileNet Object Localization," in *2022 International Conference on Advanced Creative Networks and Intelligent Systems (ICACNIS)*, IEEE, 2022, pp. 1–5. Accessed: Dec. 06, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10055274/?casa_token=B6RAcV3b8uMAAAAA:VUGswT2nUxdLzrKhg7hhLW0lBnhDJSP4DT5vbOJym0qGPeoYnPWQEuQs1wwtt8c4XWd5hRg9Pg8

[41] T. Xu and Z. Zhang, "License plate classification based on MobileNetV2," *Appl. Comput. Eng.*, vol. 4, no. 1, pp. 171–178, Jun. 2023, doi: 10.54254/2755-2721/4/20230442.

[42] Y. Qin, Q. Tang, J. Xin, C. Yang, Z. Zhang, and X. Yang, "A Rapid Identification Technique of Moving Loads Based on MobileNetV2 and Transfer Learning," *Buildings*, vol. 13, no. 2, p. 572, Feb. 2023, doi: 10.3390/buildings13020572.

[43] K. Lin, R. Hao, S. Zhang, J. Tang, and Z. Qin, "Algae Image Classification Algorithm Based on the Improved MobileNetV2," in *Proceedings of the 5th International Conference on Computer Science and Software Engineering*, Guilin China: ACM, Oct. 2022, pp. 75–79. doi: 10.1145/3569966.3569988.

[44] M. Patacchiola, J. Bronskill, A. Shysheya, K. Hofmann, S. Nowozin, and R. E. Turner, "Contextual Squeeze-and-Excitation for Efficient Few-Shot Image Classification," 2022, doi: 10.48550/ARXIV.2206.09843.

[45] Y. Hu, H. Wang, and B. Li, "SQET: Squeeze and Excitation Transformer for High-accuracy Brain Age Estimation," in *2022 IEEE International Conference on Bioinformatics and*

Biomedicine (BIBM), Las Vegas, NV, USA: IEEE, Dec. 2022, pp. 1554–1557. doi: 10.1109/BIBM55620.2022.9995270.

[46] Z. Liu, C. Zhang, X. Chen, J. Xu, L. Zhao, and R. Yan, “SE-ResNet-based noise reduction for steady-state micro-thrust measurement,” in *2022 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD)*, Harbin, China: IEEE, Nov. 2022, pp. 1–5. doi: 10.1109/ICSMD57530.2022.10058468.

[47] P. Edelenbos and A. Kubanek-German, “Teacher assessment: the concept of ‘diagnostic competence,’” *Lang. Test.*, vol. 21, no. 3, pp. 259–283, Jul. 2004, doi: 10.1191/0265532204lt284oa.

[48] R. Khoii and S. Sargolzehi, “An Application and Re-Evaluation of Borg’s Self-Assessment Tool for English Language Teachers (2018) in the Iranian EFL Context,” in *Conference Proceedings. Innovation in Language Learning 2022*, 2022. Accessed: Mar. 09, 2024. [Online]. Available: https://conference.pixel-online.net/library_scheda.php?id_abs=5687

[49] A. J. Lekwa, L. A. Reddy, and E. S. Shernoff, “The magnitude and precision of estimates of change in formative teacher assessment.,” *Sch. Psychol.*, vol. 35, no. 2, p. 137, 2020.

[50] N. Birnaz and C. Osoianu, “Summative evaluation in technical professional education: general aspects,” *Stud. Univ. Mold. Ser. Ştiinţe Ale Educ.*, no. 5(165), pp. 74–78, Jul. 2023, doi: 10.59295/sum5(165)2023_13.

[51] K. Zheng, D. Yang, J. Liu, and J. Cui, “Recognition of Teachers’ Facial Expression Intensity Based on Convolutional Neural Network and Attention Mechanism,” *IEEE Access*, vol. 8, pp. 226437–226444, 2020, doi: 10.1109/ACCESS.2020.3046225.

[52] M. Mortillaro, M. Mehu, and K. R. Scherer, “Subtly Different Positive Emotions Can Be Distinguished by Their Facial Expressions,” *Soc. Psychol. Personal. Sci.*, vol. 2, no. 3, pp. 262–271, May 2011, doi: 10.1177/1948550610389080.

[53] M. N. Shiota, B. Campos, and D. Keltner, “The Faces of Positive Emotion: Prototype Displays of Awe, Amusement, and Pride,” *Ann. N. Y. Acad. Sci.*, vol. 1000, no. 1, pp. 296–299, Dec. 2003, doi: 10.1196/annals.1280.029.

[54] S. H. Yoo and S. E. Noyes, “Recognition of Facial Expressions of Negative Emotions in Romantic Relationships,” *J. Nonverbal Behav.*, vol. 40, no. 1, pp. 1–12, Mar. 2016, doi: 10.1007/s10919-015-0219-3.

[55] E. Guz and M. Tetiurka, “Positive Emotions and Learner Engagement: Insights from an Early FL Classroom,” in *Positive Psychology Perspectives on Foreign Language Learning and Teaching*, D. Gabryś-Barker and D. Gałajda, Eds., in Second Language Learning and Teaching. , Cham: Springer International Publishing, 2016, pp. 133–153. doi: 10.1007/978-3-319-32954-3_8.

[56] M. Mohamed and M. Mogahed, “Tennis Serving Technique to Cope with Student’s Negative Comments,” *Asian J. Humanit. Soc. Stud.*, vol. Volume 03, pp. 2321–2799, May 2015.

[57] J. Yoo and J. Kim, “Capturing difficulty expressions in student online Q&A discussions,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2014. Accessed: Mar. 09, 2024. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/8718

[58] M. Descoeudres, V. Cece, and V. Lentillon-Kaestner, “The emotional significant negative events and wellbeing of student teachers during initial teacher training: The case of physical education,” *Front. Educ.*, vol. 7, Sep. 2022, doi: 10.3389/feduc.2022.970971.

[59] F.-C. Kuo, J.-M. Chen, H.-C. Chu, K.-H. Yang, and Y.-H. Chen, “A Peer-Assessment Mobile Kung Fu Education Approach to Improving Students’ Affective Performances,” *Int. J. Distance Educ. Technol.*, vol. 15, pp. 1–14, Jan. 2017, doi: 10.4018/IJDET.2017010101.

[60] A. Gupta, D. Yadav, A. Raj, and A. Pathak, “Real-Time Object Detection Using SSD MobileNet Model of Machine Learning,” *Int. J. Eng. Comput. Sci.*, vol. 12, no. 05, pp. 25729–25734, May 2023, doi: 10.18535/ijecs/v12i05.4735.

[61] B. Satya, Hendry, and D. H. F. Manongga, “Object Detection Application for a Forward Collision Early Warning System Using TensorFlow Lite on Android,” in *Third Congress on Intelligent Systems*, vol. 613, S. Kumar, H. Sharma, K. Balachandran, J. H. Kim, and J. C. Bansal, Eds., in Lecture Notes in Networks and Systems, vol. 613. , Singapore: Springer Nature Singapore, 2023, pp. 821–834. doi: 10.1007/978-981-19-9379-4_59.

[62] X. Li and Q. He, “Frame Rate Control in Distributed Game Engine,” in *Entertainment Computing - ICEC 2005*, F. Kishino, Y. Kitamura, H. Kato, and N. Nagata, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005, pp. 76–87. doi: 10.1007/11558651_8.

[63] J. Y. C. Chen and J. E. Thropp, “Review of Low Frame Rate Effects on Human Performance,” *IEEE Trans. Syst. Man Cybern. - Part Syst. Hum.*, vol. 37, no. 6, pp. 1063–1076, Nov. 2007, doi: 10.1109/TSMCA.2007.904779.

[64] S. Kim, J. Kim, N. Kim, M. Kang, and J. Seo, “Improving Inference Time of Deep Learning Model with Partial Skip of ReLU-fused Matrix Multiplication Operations,” in *2022 International Conference on Electronics, Information, and Communication (ICEIC)*, Jeju, Korea, Republic of: IEEE, Feb. 2022, pp. 1–4. doi: 10.1109/ICEIC54506.2022.9748210.

[65] K. A. Kramer, L. O. Hall, D. B. Goldgof, A. Remsen, and Tong Luo, “Fast Support Vector Machines for Continuous Data,” *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 39, no. 4, pp. 989–1001, Aug. 2009, doi: 10.1109/TSMCB.2008.2011645.

Author 1 Name and a short biography …

Author 2 Name and a short biography …

Author N Name and a short biography …

Author N Name and a short biography …