IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 12 June 2024, accepted 3 July 2024, date of publication 10 July 2024, date of current version 19 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3425820

RESEARCH ARTICLE

Automatic Detection of Students' Engagement During Online Learning: A Bagging Ensemble Deep Learning Approach

MAYANDA MEGA SANTONI[1], **T. BASARUDDIN**[1], **KASIYAH JUNUS**[1], **AND OENARDI LAWANTO**[2]

[1]Faculty of Computer Science, Universitas Indonesia, Depok 16424, Indonesia
[2]Department of Engineering Education, Utah State University, Logan, UT 84322, USA

Corresponding authors: Mayanda Mega Santoni (mayanda.mega21@ui.ac.id) and T. Basaruddin (chan@cs.ui.ac.id)

This work was supported in part by the Directorate of Research and Development, Universitas Indonesih, Publikasi Terindeks Internasional (PUTI) 2023 under Grant NKB-571/UN2.RST/HKP.05.00/2023, and in part by the GPU computing server facilities provided by the Tokopedia-UI AI Center of Excellence, Universitas Indonesia.

ABSTRACT The COVID-19 pandemic has reshaped education and shifted learning from in-person to online. While this shift offers advantages such as liberating the learning process from time and space constraints and enabling education to occur anywhere and anytime, a challenge lies in detecting student engagement during online learning due to limited interaction. Student engagement, defined as the active involvement of students in the educational journey, is a critical factor influencing the overall learning experience. This research addresses this challenge by proposing a model using bagging (bootstrap aggregating) ensemble learning applied to 1-dimensional convolutional neural networks (1D CNN), 1-dimensional residual networks (1D ResNet), and hybrid ensemble deep learning models. Utilizing the DAiSEE dataset, our findings indicate that the bagging ensemble of the 1D CNN model achieves 93.25% accuracy, surpassing the individual model by 3.25%. The deep learning ensemble bagging attains 93.75%, outperforming the unique 1D ResNet model by 3.5%. Additionally, the hybrid ensemble bagging achieves the highest accuracy of 94.25%, a 1% improvement over the 1D CNN model and a 0.5% increase over the 1D ResNet model.

INDEX TERMS Bagging, convolutional neural networks, deep learning, engagement detection, ensemble learning, online learning, residual networks.

I. INTRODUCTION

Across the globe, the COVID-19 pandemic has created complex and multidimensional situations. The pandemic represents an unprecedented challenge within education, ushering in unforeseen consequences and unforeseen social transformations [1]. Educational institutions find themselves compelled to confront pressing challenges in maintaining the continuity of the learning process [2]. Fortunately, the current era of technological disruption offers a silver lining. The rapid evolution of communication technology provides an alternative avenue for learning amid the dangerous outbreak of the COVID-19 virus [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad J. Abdel-Rahman.

As manifestation of digital technology, the online learning modality shapes contemporary and future educational paradigms [4]. Online learning transcends the constraints of physical classrooms and time limitations. [5]. Apart from the inherent flexibility of online learning, the ability to document material digitally, including learning recordings, facilitates easy access and review of presented material at later dates.

Interaction, a key element of successful learning that involves dynamic exchanges between teachers and students [6], can take place in both offline and online settings. However, it is challenging to obtain such interaction during online learning. The lack of information for teachers regarding students' enthusiasm and participation levels is a significant hurdle. According to D'Errico et al. [7], engagement is an internal state constructed from various signals and may not be visually apparent. Engagement encompasses behavioral, emotional, and cognitive aspects [7], making research initiatives to detect students' engagement crucial to enhance the learning process.

More often than not, universities typically conduct summative evaluations at the end of each semester to evaluate students' achievements in learning; however, in the student-centered learning (SCL) environment, the emphasis shifts towards formative evaluation throughout the learning process. The assessment of student learning is a continuous effort, with feedback regularly communicated to offer direct assessment to students. In this context, formative evaluation seeks to measure student engagement within the classroom and serves as valuable material for students and teachers, fostering ongoing enhancements in the learning process. Recognizing the important role of teachers, providing timely feedback becomes imperative in cultivating student engagement, influencing perspectives, and fostering an optimal learning atmosphere [8]. Hence, there arises a need for an automated approach capable of discerning the levels of student engagement, thereby supporting the sustainable refinement of the learning process.

Various approaches, including sensor-based [9], [10] and computer-vision-based [11], [12], can be employed for automatic student engagement detection. Social Signal Processing (SSP), a field closely related to this issue, seeks to empower computers with the ability to sense and understand nonverbal human social signals [13], encompassing facial expressions [14], [15], eye contact [16], [17], body gestures [18], and more.

Deep learning, an evolution of neural network methodologies to address diverse problems, includes convolutional neural networks (CNN) for classifying image or video data. CNN have successfully addressed many classification problems by extracting high-level features without the need for explicitly defined features [19]. Ensemble learning, a method that combines multiple algorithms for more accurate predictions, has been proven effective by Yu et al. [20]. The study proposed an ensemble learning method for predicting a person's emotional expression. In this study, it is said that the ensemble approach can improve the solution's performance. Ensemble learning is a method that combines several algorithms to get more accurate prediction results. Ensemble learning aims to obtain a model with higher accuracy than if only one algorithm/model was used. Therefore, ensemble deep learning can improve the performance of deep learning methods.

This research aims to develop a model for detecting student engagement in online learning video recordings using an ensemble deep learning method called bagging ensemble learning. The video data is sourced from the DAiSEE (Dataset for Affective States in E-Environments) dataset [21]. Following up on our previous research [22], data imbalance was addressed using SMOTE undersampling and oversampling and performed dimension reduction through PCA (Principal Component Analysis) and SVD (Singular Value Decomposition). The previous study demonstrated superior performance compared to benchmark evaluations on the DAiSEE dataset, achieving the highest accuracy of 77.97% with the SVD dimension reduction technique.

In this research, contributions to the ensemble learning approach include:

- Proposing a bagging ensemble learning approach for the CNN model and another for the ResNet model to improve the benchmark evaluation score on the students' engagement detection model using the DAiSEE dataset. A hybrid bagging ensemble approach combining CNN and ResNet is also proposed to demonstrate the strength of the bagging ensemble approach in creating more accurate and stable predictions.
- Improving the number of best components in the SVD dimension reduction to identify a better set of important features. This feature extraction algorithm is crucial as it aims to overcome the dimensionality curse that can render machine learning techniques ineffective [23].

II. RELATED WORKS

A. STUDENT'S ENGAGEMENT DETECTION

Chen et al. [24] predicted student engagement in collaborative learning using computer vision. This research introduces a multi-modal deep neural network (MDNN) that integrates facial expression and gaze direction as two key components to forecast student engagement in collaborative learning. Simultaneously, Buono et al. [25] predicted student engagement in online learning by proposing a deep learning method using LSTM with features such as eye gaze, head pose, and facial action unit. Meanwhile, Ikram et al. [26] used a primary dataset of 32 in-person classroom videos to predict student engagement level with a deep learning approach using VGG16.

Nezami et al. [27] devised an engagement model employing deep learning, trained in two distinct stages. A deep learning model was initially employed for facial expression training, offering a comprehensive representation of facial features. Subsequently, the weights from this facial expression training model were utilized to initialize a model to recognize students' engagement in the learning process. The outcomes of this investigation involve a comparative analysis of the engagement model with deep learning models like CNN and VGGNet, as well as traditional learning models such as HOG + SVM. Notably, the results obtained from the engagement model demonstrate a superior accuracy evaluation compared to the other three algorithms. Moreover, Pabba and Kumar [28] presented an intelligent system to monitor student engagement in classroom learning through facial expression recognition. This intelligent system achieved training and testing accuracies of 78.70% and 76.90%, respectively.

B. WORKING ON DAISEE DATASET

In 2016, Gupta et al. [21] introduced the DAiSEE (Dataset for Affective States in E-Environments) dataset designed to categorize engagement levels into very low, low, high, and very high categories. This study employs various classification methods, including InceptionNet Frame Level, InceptionNet Video Level, C3D Training, C3D FineTuning, and LRCN. Despite the significant data imbalance in the dataset, the models are evaluated using accuracy metrics, resulting in a baseline accuracy of 51.07% for identifying engagement levels.

Selim et al.'s research [29] utilized VRESEE and DAiSEE datasets, recording videos as the primary data source. Employing Hybrid EfficientNetB7, TCN, LSTM, and Bi-LSTM techniques, the study aimed to categorize student engagement levels into four categories. The proposed model demonstrated increased effectiveness, showing significant progress compared to previous performance.

Research by Paidja et al. [12] focused on the DAiSEE dataset, using video data for emotional engagement classification: very low, low, high, and very high, with a Convolutional Neural Network (CNN). The findings highlighted CNN's effectiveness in identifying emotional engagement, albeit it cannot be generalized as it was limited to a small subset of video data (77 out of 9068).

Abedi and Khan's study [30] utilized the DAiSEE dataset, employing hybrid networks, Residual Networks (ResNet), and Temporal Convolutional Networks (TCN) for classifying student engagement, namely very low, low, high, and very high. Achieving accuracy results of 63.9% for ResNet-TCN and 61.15% for ResNet-LSTM, the study demonstrated performance exceeding the baseline.

Mehta et al. [31] used the DAiSEE and EmotiW-EP datasets as data sources, with features that included students' facial expressions. This data was used to classify engagement levels into four classes, namely very low, low, high, and very high. In this study, the 3D DenseNet self-attention neural network architecture (DenseAttNet) was used by taking image cubes of students' facial expressions as input. The 3D DenseNet architecture consists of dense blocks and transition layers. This study evaluates the performance of the model using several evaluation metrics, such as accuracy, precision, recall, and F1-score. The results of this study showed that the DenseAttNet model successfully exceeded the baseline accuracy by 63.59% in classifying engagement levels. However, this study also identified some limitations. One of them is that it does not address the issue of data imbalance in the DAiSEE dataset.

Abedi et al. [32] used two distinct data sources: the IIITB Online SE dataset and the DAiSEE dataset, incorporating a range of features. These encompass affective aspects like Valence and arousal alongside behavioral components such as eye location, head pose, eye gaze direction, blink rate, and hand pose. The engagement levels were stratified into two classes for the IIITB Online SE dataset and four classes for the DAiSEE dataset. This research introduced a non-sequential methodology, termed the ''Bag of States'' (BoS), designed for measuring engagement levels in videos. A comparative analysis pitted the BoS approach against the Temporal Convolutional Network (TCN) in assessing student engagement. Results demonstrated the superior performance of the non-sequential BoS model over the TCN model [30], challenging the assumption that sequential modeling of behavioral and affective features is imperative for accurate engagement measurement. The BoS model achieved an accuracy of 66.58%. However, this study has some limitations, such as the inability to cope with videos of different lengths and the lack of explanation regarding the selection of features used in the study.

C. ENSEMBLE LEARNING

To improve the performance of deep learning classification, one of the ways that can be done is by performing ensemble learning [33], [34], [35], [36]. Ensemble models can improve evaluation results in engagement detection. Several other studies have proven that the use of bagging in deep learning, especially CNN, can improve classification performance. Zhang et al. [37] stated that the proposed deep learning bagging can predict the diagnosis of Covid 19 disease with an accuracy rate of 98.89%. When compared to methods without using ensemble learning, the ResNet algorithm has an accuracy of 97.22%. This shows that the use of deep learning bagging can improve classification accuracy results. Likewise, research by Deng et al. [38] concluded that the use of ensemble bagging strategies could provide the highest accuracy results when compared to the use of a single classification model. The proposed ensemble bagging model is robust and improves the overall accuracy of the model and, at the same time, reduces classification errors on a single model.

III. METHODOLOGY

The proposed bagging ensemble learning model can be seen in Figure 1. Detailed insights into addressing the imbalance within the DAiSEE dataset in this study can be found in our previous work [22]. Each video from the DAiSEE dataset undergoes extraction into 300 frames. The OpenFace library serves as the tool for feature extraction from each frame, providing a numerical vector consist 709 facial feature values encompassing facial landmark detection, head pose estimation, eye gaze estimation, and facial expressions in the form of facial action unit (AUs) features [39]. From the pool of 709 facial features, a meticulous feature selection process ensues to identify key attributes that authentically represent the dataset, enhancing the precision of the prediction model. This selection technique involves the application of Singular Value Decomposition (SVD). The optimal value of n-components is determined based on the tradeoff in variance generated through SVD. In tandem with this, a data augmentation stage is implemented to bolster the quantity and diversity of the dataset.

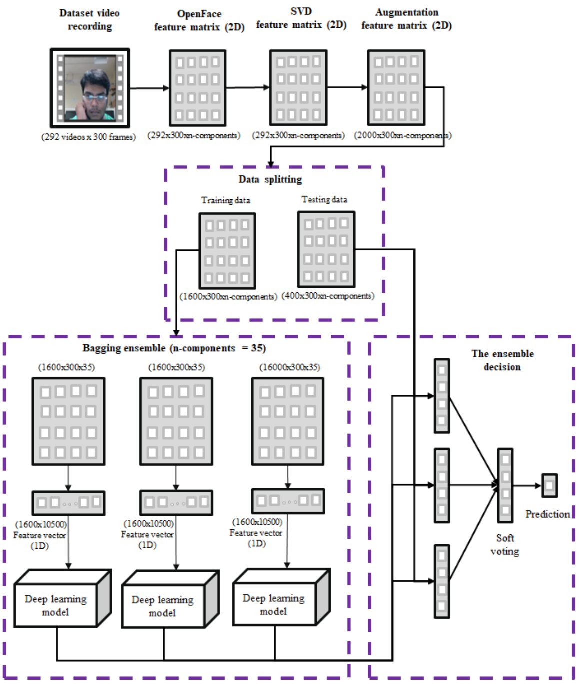

FIGURE 1. Proposed bagging ensemble learning model.

As this research proposes a bagging ensemble learning approach that demands additional resources, the data are transformed from multidimensional to 1D vectors after the bagging process to enhance computational efficiency. Any spatial information lost during this transformation will be addressed in the facial feature extraction process using the OpenFace library.

A. DATASET

DAiSEE is a multi-label video classification set consisting of 9068 10-second videos recorded from 112 students to identify students' affective state levels of boredom, confusion, engagement, and frustration. Boredom was defined as feeling tired or restless due to lack of interest. Confusion was defined as a noticeable lack of comprehension, while engagement was described as a state of interest arising from involvement in an activity. Frustration was described as displeasure or annoyance [40].

Table 1 is a snapshot of the data labeling for each video in the DAiSEE dataset [21]. Each video shows the level of the student's affective state denoted in numbers, 0 for very low, 1 for low, 2 for high and 3 for very high. For example, in Table 1, the video with clipID 1100011002.avi, has a boredom level of 0 (very low), an engagement level of 2 (high), a confusion level of 0 (very low), and a frustration level of 0 (very low). This research focuses on using only the affective group of students, namely engagement.

B. DATA SPLITTING

The data is split into two: training data and testing data. The test data is used to test the performance of the deep learning model generated from the training data. This test data is unseen data during the model training process. Meanwhile, during the training process of the deep learning classification model, the training data will be divided into two, namely training data and validation data. Unlike the testing data, which is used to evaluate the final model from the training data, the validation data is used to evaluate the performance of the model during the training process. The proportion of training, validation, and testing data is 64%, 16%, and 20%, with 1280 training data, 320 validation data, and 400 test data, respectively. To ensure that each data distribution is in the normal distribution, it is necessary to visualize the data distribution in each class. In Figure 2, it can be seen that the data is normally distributed.

TABLE 1. Example of data labeling on daisee dataset.

ClipID	Boredom	Engagement	Confusion	Frustation
1100011002.avi	0	2	0	0
1100011003.avi	0	2	0	0
1100011004.avi	0	3	0	0
1100011009.avi	0	2	1	0
1100011011.avi	0	3	0	0
1100011012.avi	0	2	2	0

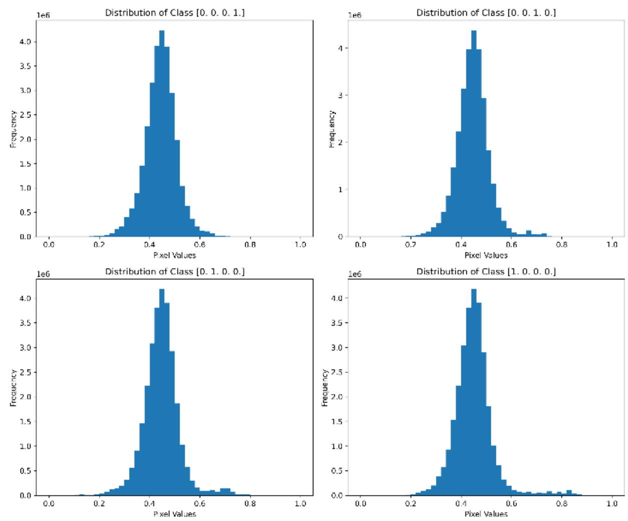

FIGURE 2. Data distribution for each class.

C. BAGGING ENSEMBLE APPROACH

Bagging stands for bootstrap aggregating. A bagging ensemble is a type of ensemble learning that uses several models of the same algorithm and then is trained on the same dataset [39]. The prediction results from each training model will be combined using soft voting techniques. An illustration of bagging in general can be seen in Figure 3.

In this study, the number of baggings used for ensemble learning is three bagging training subsets. Applying sampling with replacement is used to sample these three bagging subsets. Replacement means that one instance can be sampled multiple times for the same classifier. To get the final decision of the three-bagging ensemble learning using the soft voting technique. This soft voting technique uses two approaches,

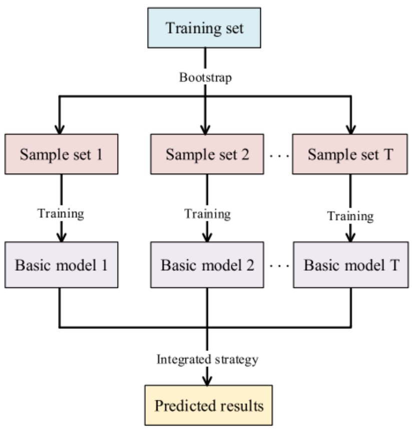

FIGURE 3. Bagging ensemble [41].

namely averaging soft voting and maximum soft voting, which are illustrated in Figure 4.

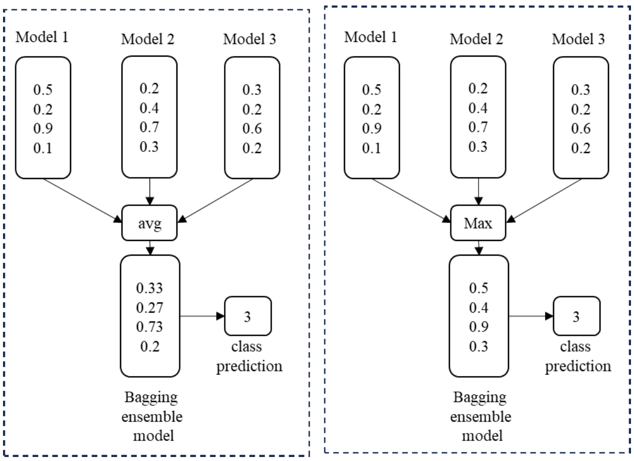

FIGURE 4. Soft Voting in Ensemble Decision, left for averaging soft voting, right for maximum soft voting.

Averaging soft voting calculates the average probability for each class across all bagging models. As illustrated in Figure 5, model 1 assigns probabilities of 0.5 to class 1, 0.2 to class 2, 0.9 to class 3, and 0.1 to class 4. Using only model 1, the prediction would be class 3 based on the highest probability. This process is similar for models 2 and 3. In ensemble bagging, the probabilities from each model are averaged for each class, resulting in a new probability distribution used for prediction. In this example, the ensemble averaging soft voting predicts the data as class 3. On the other hand, maximum soft voting makes the ensemble decision by selecting the highest probability from each model. In the illustrated example, the maximum soft voting also predicts the data as class 3.

IV. EXPERIMENTAL SETUP

A. FEATURES EXTRACTION USING SVD

The number of n-components in dimension reduction using SVD represents the extraction of important features in the data. In previous research, the value of n-components chosen was 300 with the consideration that the resulting feature matrix is square in size, namely 300 × 300. However, in this study, the feature matrix will later be converted into a 1-dimensional form (feature vector), so it is necessary to find the best value of n-components or features in order to produce a more accurate classification model. Table 2 shows that the greater the variance value, the greater the number of n-components needed. In this experiment, the component values of 300, 35, and 19 will be compared to determine which produce a more accurate model evaluation level.

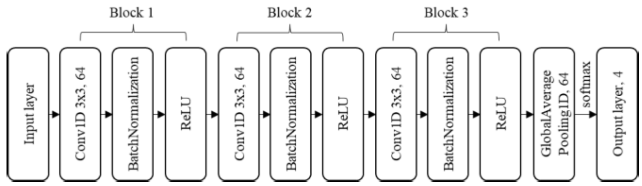

FIGURE 5. CNN architecture.

TABLE 2. Variance value for each of the N-components.

Variance	n-components
99.99998%	300
99.99000%	35
99.95000%	19
99.90000%	14
99.00000%	6

B. 1D CNN MODEL ARCHITECTURE

The architecture of the ResNet training model can be seen in Figure 5. During the model learning process, the training parameters include batch size, epoch, learning rate, and optimizer. Parameter selection is done with a trial-error approach. Table 3 shows the CNN model parameters used in this study. An additional element implemented in the training process of the Convolutional Neural Network (CNN) model is Callbacks. It aims to supervise and organize the training process. This research uses several types of callbacks, namely ModelCheckpoint, EarlyStopping, ReduceLROnPlateau, and LearningRateScheduler.

TABLE 3. Learning parameters of 1D CNN & 1D resnet models.

Parameters	Parameters Value
Number of epochs	3000
Optimizer	Adam
Batch size	32, 64

C. 1D RESNET MODEL ARCHITECTURE

The architecture of the ResNet training model can be seen in Figure 6. Just like the CNN model, the training parameters used in the ResNet model are batch size, epoch, learning rate, and optimizer with the same parameter values as the CNN model (Table 3). Likewise, the type of callbacks used is the same as the CNN model.

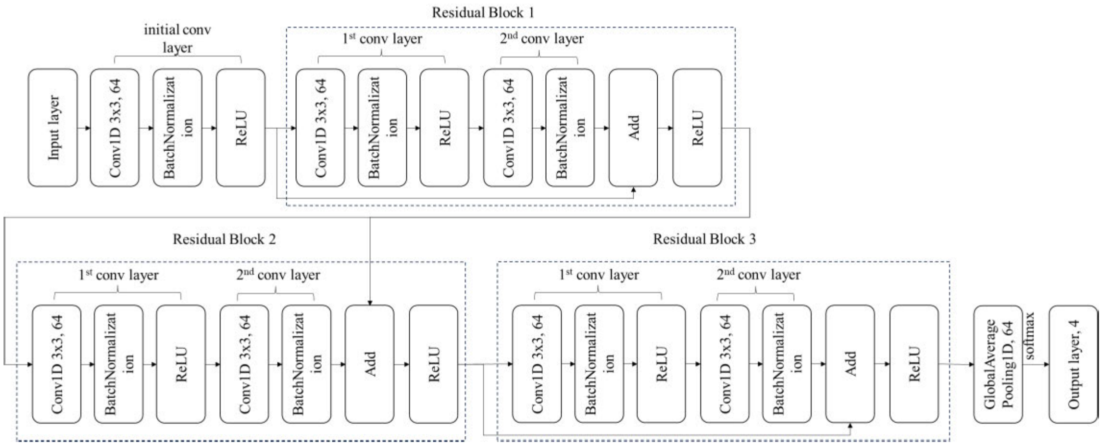

FIGURE 6. ResNet architecture.

D. EVALUATION METRICS

This research will use several evaluation metrics, namely accuracy, precision, recall, and F1-score. Accuracy measures how many predictions from all predicted data are correct according to the actual class [42]. Precision is also called positive predictive value. These metric measures, out of all instances classified as a certain class (true positives + false positives), how many instances actually belong to that class. True positives refer to instances that are correctly classified as positive by the model, while false positives refer to instances that are incorrectly classified as positive by the model when they are actually negative. If our model classifies something as positive, precision indicates how confident we are that it is actually positive.

Recall represents the ratio of instances truly belonging to a class (true positives + false negatives) that are accurately classified. False negatives refer to instances that are incorrectly classified as negative by the model when they are actually positive. Recall offers insights into the model's ability to categorize instances of a specific class correctly.

F1-Score is defined as the harmonic mean of precision and recall, and like recall and precision, its value ranges between 0 and 1. The closer to 1, the better our model is. F1-Score depends on both recall and precision [42].

V. EXPERIMENTAL RESULTS

The primary contribution of this research is the application of bagging ensemble learning to CNN, ResNet, and hybrid deep learning models. Before evaluating the performance of these models, we will discuss dimensionality reduction, which aims to select important features using SVD. Dimensionality reduction aims to remove irrelevant and redundant features. A proper selection of dimensionality reduction can help increase processing speed and reduce the time and effort required to extract important information [43], [44]. The number of components in dimensionality reduction using SVD represents the extraction of important features in the data.

TABLE 4. Comparison of testing accuracy values on the number of SVD components.

Batch size	Model	Number of Components		
		300	35	19
32	CNN – Bagging Subset 1	69.50	89.50	82.75
	CNN – Bagging Subset 2	67.50	88.25	86.25
	CNN – Bagging Subset 3	62.75	88.25	87.00
	ResNet – Bagging Subset 1	65.75	90.25	86.00
	ResNet – Bagging Subset 2	76.75	89.50	90.00
	ResNet – Bagging Subset 3	67.50	89.50	90.00
64	CNN – Bagging Subset 1	72.25	90.00	89.00
	CNN – Bagging Subset 2	73.50	87.75	88.75
	CNN – Bagging Subset 3	65.50	89.25	88.75
	ResNet – Bagging Subset 1	72.25	88.50	85.50
	ResNet – Bagging Subset 2	73.50	92.00	90.25
	ResNet – Bagging Subset 3	65.50	90.50	87.75

TABLE 5. Relationship of batch size and test accuracy based on T-test value.

Model	T-test batch size 32 vs 64 (alpha = 0.05)		
	300	35	19
CNN	0.065	0.572	0.030
ResNet	0.988	0.795	0.456

Referring to Table 4, The terms bagging subset 1, 2, and 3 refer to the individual models in each bagging. In addition, the terms bagging ensemble averaging soft voting (BEA) and bagging ensemble maximum soft voting (BEM) refer to the soft voting technique in making the final decision on the bagging ensemble model. Table 4 shows that the optimal batch size parameter for 1D CNN is identified as 64, while for 1D ResNet, it stands at 32. However, upon conducting a T-Test (Table 5), batch sizes 32 and 64 exhibit no significant difference in test accuracy, as evidenced by p-values exceeding alpha. Consequently, for subsequent experiments, batch size 64 will be utilized for the 1D CNN model and batch size 32 for 1D ResNet.

TABLE 6. N-component relationship and test accuracy based on annova test values.

Model	N-component 300 vs 35 vs 19 (alpha = 0.05)
CNN	1.05E-06
ResNet	1.07E-12

TABLE 7. Accuracy values for bagging ensemble deep learning models 1D CNN and 1D resnet.

Model	Testing Accuracy	
	1D CNN	1D ResNet
Bagging Subset 1	**90.00**	**90.25**
Bagging Subset 2	87.75	89.50
Bagging Subset 3	89.25	89.50
Bagging Ensemble (Averaging Soft Voting)	**93.25**	**93.75**
Bagging Ensemble (Maximum Soft Voting)	92.25	93.00

Moreover, the choice of n-components proves influential on test accuracy outcomes. As revealed in Table 6 through the ANOVA test, the p-values for the CNN and ResNet models are below alpha, signifying a substantial difference in test accuracy among the n-component groups. Notably, for feature selection via SVD, n-component 35 outperforms values of 300 and 19 in terms of accuracy. Importantly, n-component 35 retains data information by 99.99%, as indicated in Table 3. Hence, moving forward, we will employ n-component 35 across all models.

Table 7 shows evaluation values for CNN and ResNet classification models. Individual models of bagging subset 1, 2, and 3 yield accuracies of 90.00, 87.75, and 89.25 for the CNN model and 90.25, 89.50, and 89.50 for the ResNet model. Upon combining models through ensemble bagging, accuracy values surpass those of individual models. The 1D CNN individual model achieves a maximum accuracy of 90.00 while bagging ensemble learning elevates this to 93.25, an increase of 3.25. Similarly, the 1D ResNet individual model achieves a peak accuracy of 90.25, with bagging ensemble learning enhancing this to 93.75, an increase of 3.5. This enhancement is deemed significant based on T-Test results, where the p-value for both CNN and ResNet models is 0.001, where p-value < alpha, indicating a significant difference in test accuracy between bagging ensemble learning and individual deep learning models.

Aligned with accuracy results, precision, recall, and F1 score values (Table 8) demonstrate the superiority of bagging ensemble learning over individual models in Bagging subset 1, 2, and 3. Bagging subset 1 yields superior evaluation values compared to Bagging subset 2 and subset 3, with precision, recall, and F1-score values of 90.14, 90.00, and 90.00 for the 1D CNN model and 90.31, 90.25, and 90.25 for the 1D ResNet model. Additionally, for the two ensemble decision techniques in bagging ensemble, averaging soft voting outperforms maximum soft voting, with precision, recall, and F1-score values of 93.39%, 93.25%, and 93.28% for the 1D CNN model, and 93.80%, 93.75%, and 93.76% for the 1D ResNet model. The precision, recall, and F1-score improvement on bagging ensemble averaging soft voting (BEA) is 3.25, 3.25, and 3.28 for the 1D CNN model and 3.49, 3.5, and 3.52 for the 1D ResNet model.

TABLE 8. Precision, recall, f1-score values for bagging ensemble deep learning model CNN.

Model	Precision		Recall		F1-Score	
	1D CNN	1D ResNet	1D CNN	1D ResNet	1D CNN	1D ResNet
Bagging Subset1	**90.14**	**90.31**	**90.00**	**90.25**	**90.00**	**90.24**
Bagging Subset 2	88.33	89.52	87.75	89.50	87.87	89.50
Bagging Subset 3	89.27	89.81	89.25	89.50	89.24	89.55
BEA	**93.39**	**93.80**	**93.25**	**93.75**	**93.28**	**93.76**
BEM	92.31	93.06	92.25	93.00	92.27	93.02

BEA = Bagging Ensemble Averaging Soft Voting, BEM = Bagging Ensemble Maximum Soft Voting

TABLE 9. Accuracy values for bagging ensemble deep learning models 1D CNN, 1D resnet, and hybrid.

Model	Testing Accuracy	
	Averaging Soft Voting	Maximum Sof Voting
Bagging Ensemble 1D CNN	93.25	92.25
Bagging Ensemble 1D ResNet	93.75	93.00
Bagging Ensemble Hybrid (1D CNN + 1D ResNet)	**94.25**	**93.25**

The increase in precision, recall, and F1-score values is statistically significant. This is confirmed by T-Test results, where the p-value for the 1D CNN model is 2.06E-10 and 0.0001 for the 1D ResNet model, both indicating a p-value < alpha 0.5.

In addition to ensemble bagging with each deep learning model (1D CNN and 1D ResNet), further experiments were conducted with a hybrid bagging ensemble process, combining decisions from the bagging ensemble of 1D CNN and 1D ResNet. Table 9 reveals that hybrid ensemble bagging produces higher accuracy results compared to bagging ensembles of 1D CNN and 1D ResNet deep learning models. Hybrid ensemble bagging attains the highest accuracy of 94.25%, marking a 1% improvement for the 1D CNN model and a 0.5% increase for the 1D ResNet model compared to deep learning ensemble bagging.

The experimental results highlight variations in bias and variance among each deep learning model in different experiments. However, post-bagging ensemble learning, prediction results exhibit improvements over individual models.

Each deep learning model may have a higher bias if the model is trained on the original training data (without ensemble). However, the bias and variance can be reduced when the prediction results from all the individual deep learning models are combined (ensemble). This means that by combining predictions from many models that may have different biases and variances, ensemble bagging can create more accurate and stable predictions.

TABLE 10. Example of bagging ensemble of 1D CNN deep learning model that produces correct prediction (condition 1).

Target Class	Model Probability				
	Bagging subset 1	Bagging subset 2	Bagging subset 3	Bagging ensemble averaging soft voting	Bagging ensemble maximum soft voting
0	0.1723×10^{-2}	0.9059×10^{-4}	0.3908×10^{-6}	0.6046×10^{-3}	0.1723×10^{-2}
1	0.6308×10^{-3}	0.2779×10^{-4}	0.4149×10^{-4}	0.2519×10^{-4}	0.4149×10^{-4}
2	0.3182×10^{-7}	0.4852×10^{-6}	0.2808×10^{-11}	0.1723×10^{-6}	0.4852×10^{-6}
3	**0.9983**	**0.99988**	**0.99996**	**0.99937**	**0.99996**

TABLE 11. Example of bagging ensemble of 1D CNN deep learning model that produces correct prediction (condition 2).

Target class	Model Probability				
	Bagging subset 1	Bagging subset 2	Bagging subset 3	Bagging ensemble averaging soft voting	Bagging ensemble maximum soft voting
0	**0.9999**	0.4384×10^{-6}	**0.9999**	**0.6667**	**0.9999**
1	0.17339×10^{-6}	**0.9999**	0.7981×10^{-7}	0.3333	0.9996
2	0.5534×10^{-26}	0.1173×10^{-6}	0.1124×10^{-12}	0.3911×10^{-9}	0.1173×10^{-8}
3	0.2319×10^{-17}	0.1702×10^{-12}	0.1073×10^{-16}	0.5674×10^{-13}	0.1702×10^{-12}

TABLE 12. Example of bagging ensemble of 1D CNN deep learning model that produces correct prediction (condition 3).

Target class	Model probability				
	Bagging subset 1	Bagging subset 2	Bagging subset 3	Bagging ensemble averaging soft voting	Bagging ensemble maximum soft voting
0	0.3797×10^{-2}	0.3938×10^{-5}	**0.8962**	0.3112	0.8962
1	0.2659×10^{-7}	**0.7447**	0.0214	0.2554	0.7447
2	**0.9620**	0.2553	0.0824	**0.4332**	**0.9620**
3	0.3513×10^{-11}	0.1532×10^{-23}	0.1971×10^{-15}	0.1171×10^{-10}	0.3513×10^{-10}

VI. DISCUSSION

Bagging ensemble learning improves the value of test data evaluation, in which the data was unseen in forming the training model. Among the decision-making techniques, bagging ensemble averaging soft voting stands out, delivering higher value evaluation results to bagging ensemble maximum soft voting. Specifically, the averaging soft voting method achieves the highest accuracy rates: 94.25% for the hybrid bagging ensemble model, 93.75% for the 1D ResNet bagging ensemble model, and 93.25% for the 1D CNN bagging ensemble model. In contrast, the maximum soft voting technique records the highest accuracy rates of 93.25% for the hybrid bagging ensemble model, 93.00% for the 1D ResNet bagging ensemble model, and 92.25% for the 1D CNN bagging ensemble model.

The increase in accuracy after ensemble bagging depends on the combined probability value of each deep learning model. The prediction results of bagging ensemble deep learning models can provide correct prediction results if they experience the following conditions:

- Condition 1: The maximum probability of each deep learning model combined is equally correctly identified in the target class so that the identification results, when combined, will also give correct prediction results. For example, Table 10 shows that the target class is class 3. In the 1D CNN deep learning model, bagging 1, 2, and 3 are equally identified as belonging to class 3 so that the ensemble bagging results will be correctly identified in the target class.
- Condition 2: The maximum probability in one of the deep learning models is mispredicted. However, the error probability value does not affect the combined probability value because the other two deep learning models have higher probabilities, so the correct prediction can be maintained in the ensemble bagging model. For example, in Table 11, the target class is class 0. Although the 1D CNN deep learning model in bagging 2 has a wrong prediction, the ensemble bagging results still provide correct prediction results on the target class.
- Condition 3: Almost the same as in condition 2, although there is a maximum probability of two deep learning models having wrong predictions, the error probability does not affect the combined probability value because the correct prediction probability value is higher so that the correct prediction can still be maintained in the ensemble bagging model. For example, in Table 12, the target class is class 2. Although only the 1D CNN deep learning model in bagging 1 has a correct prediction, but because the probability value is higher than bagging 2 and 3, the ensemble bagging results still provide correct prediction results for the target class.

Meanwhile, prediction failure in bagging ensemble learning can occur when experiencing the following conditions:

TABLE 13. Example of bagging ensemble of 1D CNN deep learning model that produces wrong prediction (condition 4).

Target class	Model probability				
	Bagging subset 1	Bagging subset 2	Bagging subset 1	Bagging ensemble averaging soft voting	Bagging subset 1
0	0.2467×10^{-4}	0.9091×10^{-4}	0.7604×10^{-2}	0.0026	0.0076
1	**0.9925**	**0.9903**	**0.9723**	**0.9849**	**0.9925**
2	0.7512×10^{-2}	0.00964	0.0148	0.0106	0.0148
3	0.1469×10^{-9}	0.4432×10^{-6}	0.0054	0.0018	0.0054

TABLE 14. Example of bagging ensemble of 1D CNN deep learning model that produces wrong prediction (condition 5).

Target class	Model probability				
	Bagging subset 1	Bagging subset 2	Bagging subset 3	Bagging ensemble averaging soft voting	Bagging ensemble maximum soft voting
0	0.0083	0.1214	0.0048	0.0448	0.1214
1	0.4339	**0.4393**	0.4207	**0.4313**	0.4393
2	0.0440	0.2795	**0.5142**	0.2792	**0.5142**
3	**0.5139**	0.1598	0.0603	0.2447	0.5139

TABLE 15. Execution time training process.

Model	Time (minutes)
1D CNN - Bagging subset 1	18.28
1D CNN - Bagging subset 2	14.17
1D CNN - Bagging subset 3	21.10
1D CNN – Bagging ensemble	53.55
1D ResNet – Bagging subset 1	36.32
1D ResNet – Bagging subset 2	44.10
1D ResNet – Bagging subset 3	43.68
1D ResNet – Bagging ensemble	124.10
Bagging ensemble hybrid	177.65

- Condition 4: In contrast to condition 1, when the maximum probability of each combined deep learning model is identified incorrectly in the target class, the identification results, when combined, will also give incorrect prediction results. For example, in Table 14, the target class is class 2. However, 1D CNN bagging deep learning models 1, 2, and 3 are equally identified as class 1, so the ensemble bagging results will be incorrectly identified in the target class.
- Condition 5: Unlike in conditions 2 and 3, although a 1D CNN deep learning model correctly predicts the target class, the ensemble learning bagging results cannot maintain the correct prediction. This is because the value of the probability of correct prediction in the model is not higher than the probability of incorrect prediction. While bagging 1 with the 1D CNN deep learning model yields a correct prediction, the maximum probability value is approximately 0.513. This suggests that the data's probability for prediction in the target class is uncertain. Consequently, when combined with this probability value, it diminishes, making it challenging to sustain the probability of correct predictions in the target class.

In terms of time usage, it is noteworthy that bagging ensemble learning requires more time compared to individual learning models (1D CNN or 1D ResNet alone). Referencing Table 15, the bagging ensemble of the 1D CNN deep learning model extends to approximately 53 minutes and 33 seconds, marking a threefold increase in duration compared to the 1D CNN deep learning individual model. Similarly, the 1D ResNet deep learning model ensemble bagging consumes about 124 minutes and 6 seconds (equivalent to 2 hours and 4 minutes and 6 seconds). Conversely, for the hybrid ensemble bagging, the execution time is twice as long as 1D CNN or 1D ResNet ensemble bagging alone, totaling a training process execution time of 2 hours, 57 minutes, and 39 seconds. Despite the extended duration of the training process, it's crucial to note that this process occurs only once.

Furthermore, these results have practical applications in online learning. The deep learning ensemble bagging model proposed in our study can serve as a continuous learning evaluation framework, providing rapid feedback to enhance students' engagement and foster a conducive learning environment. For example, this model empowers online learning platforms to discern student engagement patterns, indicating comprehensive levels during online lecture. Real-time insight into these engagement patterns enables educators to adapt teaching approaches promptly, thus enhancing student engagement. Overall, our research provides valuable insights into how a deep learning ensemble bagging approach can illuminate teacher-student interactions during the learning process.

VII. CONCLUSION

This research has successfully developed a model for discerning students' engagement through video recordings in online learning, utilizing the DAiSEE dataset. Employing a deep learning ensemble approach—specifically bagging ensemble learning—three models are proposed: bagging ensemble learning for the 1D CNN deep learning model, bagging ensemble for the 1D ResNet deep learning model, and a hybrid bagging ensemble of 1D CNN and 1D ResNet deep learning models. Experimental results highlight the efficacy of combining deep learning models with bagging ensemble learning, demonstrating an enhancement in the evaluation of student engagement prediction. Individual deep learning models achieve peak accuracy, reaching 90% for 1D CNN and 90.25% for 1D ResNet.

Bagging ensemble learning employs two decision-making techniques: averaging soft voting and maximum soft voting. Averaging soft voting achieves accuracies of 93.25% for 1D CNN and 93.75% for 1D ResNet, outperforming maximum soft voting, which achieves accuracies of 92.25% for 1D CNN and 93% for 1D ResNet.

Furthermore, the hybrid ensemble bagging yields superior accuracy outcomes compared to individual deep learning models, reaching the highest accuracy value of 94.25%. This represents a 1% increase for the 1D CNN model and a 0.5% increase for the 1D ResNet model.

In conclusion, individual deep learning models may exhibit higher bias when trained on original data. However, when their predictions are combined through ensemble bagging, bias and variance may decrease. This suggests that combining predictions from models with differing biases and variances can lead to more accurate and stable predictions. However, the accuracy improvement post-ensemble bagging depends on the probability values of each combined model.

Nonetheless, it's important to note that bagging ensemble learning may extend training execution time compared to individual models.

REFERENCES

[1] K. V. Elumalai, J. P. Sankar, J. A. John, N. Menon, M. S. M. Alqahtani, and M. A. Abumelha, "Factors affecting the quality of e-learning during the COVID-19 pandemic from the perspective of higher education students," *J. Inf. Technol. Educ., Res.*, vol. 19, pp. 731–753, Sep. 2020, doi: 10.28945/4628.

[2] V. J. García-Morales, A. Garrido-Moreno, and R. Martín-Rojas, "The transformation of higher education after the COVID disruption: Emerging challenges in an online learning scenario," *Frontiers Psychol.*, vol. 12, pp. 1–6, Feb. 2021, doi: 10.3389/fpsyg.2021.616059.

[3] O. B. Adedoyin and E. Soykan, "COVID-19 pandemic and online learning: The challenges and opportunities," *Interact. Learn. Environments*, vol. 31, no. 2, pp. 863–875, Feb. 2023, doi: 10.1080/10494820.2020.1813180.

[4] A. Haleem, M. Javaid, M. A. Qadri, and R. Suman, "Understanding the role of digital technologies in education: A review," *Sustain. Operations Comput.*, vol. 3, pp. 275–285, Jan. 2022, doi: 10.1016/j.susoc.2022.05.004.

[5] C. Greenhow, C. R. Graham, and M. J. Koehler, "Foundations of online learning: Challenges and opportunities," *Educ. Psychologist*, vol. 57, no. 3, pp. 131–147, Jul. 2022, doi: 10.1080/00461520.2022.2090364.

[6] K. Aldrup, B. Carstensen, and U. Klusmann, "Is empathy the key to effective teaching? A systematic review of its association with teacher–student interactions and student outcomes," *Educ. Psychol. Rev.*, vol. 34, no. 3, pp. 1177–1216, Sep. 2022, doi: 10.1007/s10648-021-09649-y.

[7] F. D'Errico, M. Paciello, and L. Cerniglia, "When emotions enhance students' engagement in e-learning processes," *J. E-Learn. Knowl. Soc.*, vol. 12, no. 4, pp. 9–23, 2016, doi: 10.20368/1971-8829/1144.

[8] Z. Zhang and K. Hyland, "Fostering student engagement with feedback: An integrated approach," *Assessing Writing*, vol. 51, Jan. 2022, Art. no. 100586, doi: 10.1016/j.asw.2021.100586.

[9] I. Dubovi, "Cognitive and emotional engagement while learning with VR: The perspective of multimodal methodology," *Comput. Educ.*, vol. 183, Jul. 2022, Art. no. 104495, doi: 10.1016/j.compedu.2022.104495.

[10] E. Di Lascio, S. Gashi, and S. Santini, "Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 1–21, Sep. 2018, doi: 10.1145/3264913.

[11] J. Liao, Y. Liang, and J. Pan, "Deep facial spatiotemporal network for engagement prediction in online learning," *Appl. Intell.*, vol. 51, no. 10, pp. 6609–6621, Oct. 2021, doi: 10.1007/s10489-020-02139-8.

[12] A. Nurrahma Rosanti Paidja and F. A. Bachtiar, "Engagement emotion classification through facial landmark using convolutional neural network," in *Proc. 2nd Int. Conf. Inf. Technol. Educ. (ICITE)*, Jan. 2022, pp. 234–239, doi: 10.1109/ICITE54466.2022.9759546.

[13] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image Vis. Comput.*, vol. 27, no. 12, pp. 1743–1759, Nov. 2009, doi: 10.1016/j.imavis.2008.11.007.

[14] Z. Zhang, Z. Li, H. Liu, T. Cao, and S. Liu, "Data-driven online learning engagement detection via facial expression and mouse behavior recognition technology," *J. Educ. Comput. Res.*, vol. 58, no. 1, pp. 63–86, Mar. 2020, doi: 10.1177/0735633119825575.

[15] M. U. Abdullah and A. Alkan, "A comparative approach for facial expression recognition in higher education using hybrid-deep learning from students' facial images," *Traitement du Signal*, vol. 39, no. 6, pp. 1929–1941, Dec. 2022, doi: 10.18280/ts.390605.

[16] P. Sharma, S. Joshi, S. Gautam, S. Maharjan, V. Filipe, and M. C. Reis, "Student engagement detection using emotion analysis, eye tracking and head movement with machine learning," 2023, *arXiv:1909.12913*.

[17] S. Akshay and P. Vasanth, "A CNN based model for identification of the level of participation in virtual classrooms using eye movement features," in *Proc. IEEE Int. Conf. Electron., Comput. Commun. Technol. (CONECCT)*, Jul. 2022, pp. 1–6, doi: 10.1109/CONECCT55679.2022.9865694.

[18] S. Khenkar and S. Kammoun Jarraya, "Engagement detection based on analyzing micro body gestures using 3D CNN," *Comput., Mater. Continua*, vol. 70, no. 2, pp. 2655–2677, 2022, doi: 10.32604/cmc.2022.019152.

[19] Z. Fang, W. Li, J. Zou, and Q. Du, "Using CNN-based high-level features for remote sensing scene classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 2610–2613, doi: 10.1109/IGARSS.2016.7729674.

[20] J. Yu, Z. Cai, P. He, G. Xie, and Q. Ling, "Multi-model ensemble learning method for human expression recognition," 2022, *arXiv:2203.14466*.

[21] A. Gupta, A. D'Cunha, K. Awasthi, and V. Balasubramanian, "DAiSEE: Towards user engagement recognition in the wild," 2016, *arXiv:1609.01885*.

[22] M. M. Santoni, T. Basaruddin, and K. Junus, "Convolutional neural network model based students' engagement detection in imbalanced DAiSEE dataset," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 3, pp. 617–626, 2023. [Online]. Available: https://www.ijacsa.thesai.org

[23] F. Anowar, S. Sadaoui, and B. Selim, "Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)," *Comput. Sci. Rev.*, vol. 40, May 2021, Art. no. 100378, doi: 10.1016/j.cosrev.2021.100378.

[24] Y. Chen, J. Zhou, Q. Gao, J. Gao, and W. Zhang, "MDNN: Predicting student engagement via gaze direction and facial expression in collaborative learning," *Comput. Model. Eng. Sci.*, vol. 136, no. 1, pp. 381–401, 2023, doi: 10.32604/cmes.2023.023234.

[25] P. Buono, B. De Carolis, F. D'Errico, N. Macchiarulo, and G. Palestra, "Assessing student engagement from facial behavior in on-line learning," *Multimedia Tools Appl.*, vol. 82, no. 9, pp. 12859–12877, Apr. 2023, doi: 10.1007/s11042-022-14048-8.

[26] S. Ikram, H. Ahmad, N. Mahmood, C. M. N. Faisal, Q. Abbas, I. Qureshi, and A. Hussain, "Recognition of student engagement state in a classroom environment using deep and efficient transfer learning algorithm," *Appl. Sci.*, vol. 13, no. 15, p. 8637, Jul. 2023, doi: 10.3390/app13158637.

[27] O. M. Nezami, M. Dras, L. Hamey, D. Richards, S. Wan, and C. Paris, "Automatic recognition of student engagement using deep learning and facial expression," 2018, *arXiv:1808.02324*.

[28] C. Pabba and P. Kumar, "An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition," *Expert Syst.*, vol. 39, no. 1, pp. 1–28, Jan. 2022, doi: 10.1111/exsy.12839.

[29] T. Selim, I. Elkabani, and M. A. Abdou, ''Students engagement level detection in online e-learning using hybrid EfficientNetB7 together with TCN, LSTM, and bi-LSTM,'' *IEEE Access*, vol. 10, pp. 99573–99583, 2022, doi: 10.1109/ACCESS.2022.3206779.

[30] A. Abedi and S. S. Khan, ''Improving state-of-the-art in detecting student engagement with resnet and TCN hybrid network,'' 2021, *arXiv:2104.10122*.

[31] N. K. Mehta, S. S. Prasad, S. Saurav, R. Saini, and S. Singh, ''Three-dimensional DenseNet self-attention neural network for automatic detection of student's engagement,'' *Appl. Intell.*, vol. 52, no. 12, pp. 13803–13823, Sep. 2022, doi: 10.1007/s10489-022-03200-4.

[32] A. Abedi, C. Thomas, D. B. Jayagopi, and S. S. Khan, ''Bag of states: A non-sequential approach to video-based engagement measurement,'' 2023, *arXiv:2301.06730*.

[33] H. Li, Y. Li, and F. Porikli, ''Convolutional neural net bagging for online visual tracking,'' *Comput. Vis. Image Understand.*, vol. 153, pp. 120–129, Dec. 2016, doi: 10.1016/j.cviu.2016.07.002.

[34] M. Mohammed, H. Mwambi, I. B. Mboya, M. K. Elbashir, and B. Omolo, ''A stacking ensemble deep learning approach to cancer type classification based on TCGA data,'' *Sci. Rep.*, vol. 11, no. 1, pp. 1–22, Aug. 2021, doi: 10.1038/s41598-021-95128-x.

[35] E. K. Jadoon, F. G. Khan, S. Shah, A. Khan, and M. ElAffendi, ''Deep learning-based multi-modal ensemble classification approach for human breast cancer prognosis,'' *IEEE Access*, vol. 11, pp. 85760–85769, 2023, doi: 10.1109/access.2023.3304242.

[36] W. Jiang, Y. Wu, F. Qiao, L. Meng, Y. Deng, and C. Liu, ''Model level ensemble for facial action unit recognition at the 3rd ABAW challenge,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 2336–2343, doi: 10.1109/CVPRW56347.2022.00260.

[37] Z. Zhang, B. Chen, J. Sun, and Y. Luo, ''A bagging dynamic deep learning network for diagnosing COVID-19,'' *Sci. Rep.*, vol. 11, no. 1, pp. 1–15, Aug. 2021, doi: 10.1038/s41598-021-95537-y.

[38] W. Deng, Q. Xu, J. Liu, Y. Lu, M. Fan, and X. Liu, ''Image classification method of longhorn beetles of Yunnan based on bagging and CNN,'' in *Proc. 5th Int. Conf. Pattern Recognit. Artif. Intell. (PRAI)*, Aug. 2022, pp. 198–203, doi: 10.1109/PRAI55851.2022.9904155.

[39] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, ''OpenFace 2.0: Facial behavior analysis toolkit,'' in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 59–66.

[40] S. D'Mello and A. Graesser, ''Dynamics of affective states during complex learning,'' *Learn. Instruct.*, vol. 22, no. 2, pp. 145–157, Apr. 2012, doi: 10.1016/j.learninstruc.2011.10.001.

[41] X. Zheng, W. Chen, Y. You, Y. Jiang, M. Li, and T. Zhang, ''Ensemble deep learning for automated visual classification using EEG signals,'' *Pattern Recognit.*, vol. 102, Jun. 2020, Art. no. 107147, doi: 10.1016/j.patcog.2019.107147.

[42] M. Grandini, E. Bagli, and G. Visani, ''Metrics for multi-class classification: An overview,'' 2020, *arXiv:2008.05756*.

[43] S. Ayesha, M. K. Hanif, and R. Talib, ''Overview and comparative study of dimensionality reduction techniques for high dimensional data,'' *Inf. Fusion*, vol. 59, pp. 44–58, Jul. 2020, doi: 10.1016/j.inffus.2020.01.005.

[44] Y. Kiarashinejad, S. Abdollahramezani, and A. Adibi, ''Deep learning approach based on dimensionality reduction for designing electromagnetic nanostructures,'' *NPJ Comput. Mater.*, vol. 6, no. 1, pp. 1–12, Feb. 2020, doi: 10.1038/s41524-020-0276-y.

MAYANDA MEGA SANTONI received the bachelor's degree in computer science from IPB University, in 2012, and the master's degree in computer science from Universitas Indonesia, in 2015, where she is currently pursuing the Ph.D. degree in computer science. She lectures with the Undergraduate Informatics Study Program, UPN Veteran Jakarta. During the master's study, she received the Domestic Postgraduate Education Scholarship for Lecturer Candidates. She also received the 2022 Indonesian Education Scholarship. Her research interests include computer vision, machine learning, data mining, and social signal processing.

T. BASARUDDIN received the degree (cum laude) in mathematics from Universitas Gadjah Mada, in 1984, the Master of Science and Ph.D. degrees in computing from The University of Manchester, England, U.K., in 1988 and 1990, respectively, and the Honorary Doctorate degree from Management Science University–Malaysia for Higher Education Quality and Management, in 2019. He is currently a Professor of computer science with the Faculty of Computer Science, Universitas Indonesia. He is also an Expert in numerical computing, modeling, and simulation. His expertise is in optimization, mathematical modeling, mathematical programming, distributed systems, nonlinear optimization, optimization theory, and social signal processing.

KASIYAH JUNUS received the B.S. degree in mathematics from Universitas Gadjah Mada, Yogyakarta, Indonesia, in 1986, the M.S. degree in computer science from the University of Western Ontario, London, ON, Canada, in 1991, and the Ph.D. degree in computer science from Universitas Indonesia, Depok, Indonesia, in 2017. She is currently a Professor of computer science with the Faculty of Computer Science, Universitas Indonesia. Her research interests include online learning, inquiry communities, and metacognition.

OENARDI LAWANTO received the B.S. degree in electrical engineering from Iowa State University, in 1986, the M.S. degree in electrical engineering from the University of Dayton, in 1988, and the Ph.D. degree in human resource education from the University of Illinois at Urbana, in 2008. He is currently a Professor of engineering education with Utah State University, USA. His research interests include cognition, e-learning, learning, problem-solving, and design.