Undergraduate Conference on Intelligent Computing and Systems (UCICS 2025)
26-27 February, 2025; Varendra University, Rajshahi, Bangladesh

Enhancing Online Learning- Distraction Detection and Engagement Monitoring with Transfer Learning CNNs

Salman Farsi
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
salmanfarsi214@gmail.com

Kazi Mahmudul Hasan
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
shafi16221@gmail.com

Hafijur Rahman Hemel
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
hemelhasan97@gmail.com

Md. Fatin Nibbrash Nakib
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
fatinnibbrash@gmail.com

Md. Taufiq Khan
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
khantaufiq2001@gmail.com

Md. Arafat Ibna Mizan
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
arafat.cse.ruet18@gmail.com

***Abstract*—It is now more important than ever to measure student attentiveness precisely and spot distractions in the midst of the massive change from offline to online learning. Effective remote student monitoring is quite difficult for many universities. This paper provides a thorough examination of deep learning methods to address these issues, filling in the gaps in previous studies. This study uses state-of-the-art convolutional neural networks (CNNs) as feature extractors, such as DenseNet121, Xception, VGG16, VGG19, ResNet50, and EfficientNet-B2, through transfer learning methodologies, using a newly curated dataset created for the online learning environment. Tests were carried out in 15 different classes, and the astounding accuracy of 97.67% was obtained. This study contributes significantly to educational technology by providing insightful information about how to increase student focus and engagement in the digital age.By showcasing the effectiveness of these advanced CNN models, this study paves the way for the development of more robust monitoring and support systems in online education, ensuring that students remain focused and engaged in virtual learning environments.**

***Index Terms*—Convolutional Neural Network (CNN), Students Distracted State Recognition, Transfer Learning**

I. Introduction

Global conventions have been fundamentally altered by the COVID-19 pandemic, which has also acted as a catalyst for already-occurring technical developments. Education is one of the many industries affected, and it has changed significantly. Due to the quick development of information and communication technology (ICT), online learning has become a vital part of educational institutions all over the world. These advancements have improved accessibility and quality for students by redefining the way educational services are provided [1].Many educational institutions have adopted online platforms to hold lessons, increase their course offerings, and administer exams as a result of the transition to remote learning.With 49 percent of students worldwide engaging in some kind of online learning, this shift has normalized virtual education. It is anticipated that there will be 57 million online learners by 2027 [2], underscoring the pressing need to improve online learning methods.Along with its advantages, this new learning approach also presents certain difficulties, especially when it comes to keeping students' attention. In virtual environments, distractions are more common, and it gets harder to keep an eye on each student's behavior as class sizes grow. Furthermore, problems like academic dishonesty during online tests are serious and compromise the validity of assessments. To improve the online learning experience, these issues necessitate creative solutions that allow for efficient student involvement monitoring [3]. This paper is organized as follows: Previous studies on student participation and monitoring systems are reviewed in Section 2. The curated dataset and the methodology used in this investigation are described in detail in Section 3. The experimental design and findings are presented in Section 4. Section 5 brings the work to a close and offers ideas for further research.

II. Literature Review

Education has been transformed by the explosive expansion of online learning platforms, which provide students convenient and easily accessible ways to learn. Distractions among students, which can result from a number of internal and external variables, frequently compromise the efficacy of online learning.

Alruwais et al. [4] used machine learning to study how to predict student participation in a virtual learning environment (VLE). In order to deal with missing values and normalize features, they preprocessed the Open University Learning Analytics Dataset (OULAD). Metrics like accuracy, precision, recall, and AUC were used to train and assess many classification algorithms (CATBoost, XGBoost, Random Forest, and MLP). With an accuracy of roughly 92.23%, CATBoost outperformed both an AISAR model and earlier studies.A system for detecting student activity is presented by Ali et al. [5] and makes use of the deep learning object detection algorithm YOLOv3. In their newly developed dataset, "SUST-S-Act," they included 150 photos of students doing seven different tasks: reading, making phone calls, using a laptop, taking books, smiling, staring, and sleeping. With a mean average precision (mAP) of 97%, the YOLOv3 model outperformed a ninety-five percent YOLOv3 and Faster R-CNN technique.

By examining lecture footage, Hasnine et al. [6] created a clever application to identify student participation in online learning. Using a CNN that has already been trained, the system employs computer vision to identify six basic emotions from pupils' faces: anger, contempt, fear, happiness, sadness, surprise, and neutrality. Students are categorized as extremely interested, engaged, or disengaged based on the concentration index that is computed using these emotions and eye gaze data.Slyman et al. [7] address the increasing need for automated solutions in educational analysis by presenting a novel method for identifying classroom activities from audio recordings. Their study uses audio from common webcams to classify nine different classroom activities, such as "lecture," "group work," and "student question," using a variety of neural network designs, including fully connected, convolutional, and recurrent networks. In 2022, Pillai [8] presents a new computer vision method for tracking classroom participation. At random intervals, the system analyzes photos taken by a high-resolution camera using the real-time object detection method YOLOv4. By classifying these photos as either "engaged" or "not engaged" kids, weekly engagement scores are produced for each individual student. YOLOv4 was selected because of its accuracy and speed. Metrics including mAP50, IoU, precision, recall, and F1-score significantly improved with YOLOv4, as evidenced by a comparison with YOLOv3 and the effects of data augmentation using Generative Adversarial Networks (GANs). To ensure robustness, the author emphasizes the need for more validation and improvement but also highlights ethical considerations with ongoing monitoring and the possibility of bias in the image recognition system.

III. Materials and Methods

In this section, the data collection process is described first. Following this, the details of the proposed convolutional neural network are outlined. Finally, a brief explanation of transfer learning is provided.

A. Dataset

The dataset introduced in [9], includes 5118 photos of different types of distraction. The dataset was initially split into an 80:20 ratio, creating separate training and testing sets. Subsequently, 80% of the training data was further divided into training and validation subsets using an 80:20 ratio.

B. Sample Data

TABLE I presents a selection of sample images from the collected dataset, comprising 5,118 images. These images, captured without any augmentation, provide a realistic representation of the data. The dataset has been divided into three segments for effective model evaluation: training, validation, and testing, in a ratio of 6:2:2. This distribution ensures a thorough assessment of the model's performance by allocating ample data for training while reserving sufficient samples for validation and testing.

C. Proposed Approach

The suggested methodology improves student distraction detection and concentration analysis by utilizing sophisticated transfer learning models and a specially created Convolutional Neural Network (CNN). Dataset preparation is the first phase in the process, which includes preprocessing techniques like scaling, normalization, and cropping to maximize images for model input as well as data augmentation to correct class imbalances. While a custom CNN with convolutional, pooling, dropout, and fully connected layers was created for comparison, transfer learning models, including VGG16, VGG19, EfficientNetB0, EfficientNetB2, EfficientNetB6, DenseNet121, ResNet50, ConvNeXtTiny, ConvNeXtXLarge, and InceptionV3, were refined to extract spatial and semantic characteristics from the dataset. VGG16 outperformed the other models in terms of capturing pertinent features for this job, as evidenced by its greatest classification accuracy. Using a categorical cross-entropy loss function, an Adam optimizer, and GPU acceleration, all models were trained. To guarantee reliable training, strategies like learning rate scheduling, early halting, and model checkpointing were used. Metrics including accuracy, precision, recall, and F1-score were used to assess performance. A comparative study revealed that VGG16 was the best model for striking a compromise between classification accuracy and computational efficiency for real-world implementation. The workflow diagram of the research methodology is shown in Figure 1.

IV. Experimental Analysis

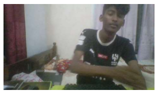
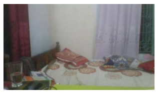
A. Experimental Setup

The study was conducted on a Kaggle environment utilizing a GPU accelerator with 30GB of RAM, and Python was the primary programming language employed for coding. To determine which combination of hyperparameters produces the greatest results on a validation set, we methodically explored a variety of values in our study.The chosen hyperparameters used in our investigation are displayed in Table II.

B. Experimental Results

TABLE III presents the results of the applied transfer learning approaches, along with the prediction time required. Notably, the VGG16 architecture attained the highest overall accuracy of 97.65 or 98% among the others transfer learning approaches considered. TABLE IV presentas the measures of accuracy,

TABLE I: Sample images of students taken through webcam while attaining online class for 15 different classes

Serial	Class Name	Sample Image
1	Class 0 – Typing Only	
2	Class 1 – Browsing Laptop	
3	Class 2 – Makeup	
4	Class 3 – Talking With Phone	
5	Class 4 – Browsing Phone	
6	Class 5 – Drinking	
7	Class 6 – Writing	
8	Class 7 – Looking Away	
9	Class 8 – Reading	
10	Class 9 – Eating	
11	Class 10 – Drowsy	
12	Class 11 – Towards Screen	
13	Class 12 – Talking With Friends	
14	Class 13 – Related Stuff	
15	Class 14 – Empty Background	

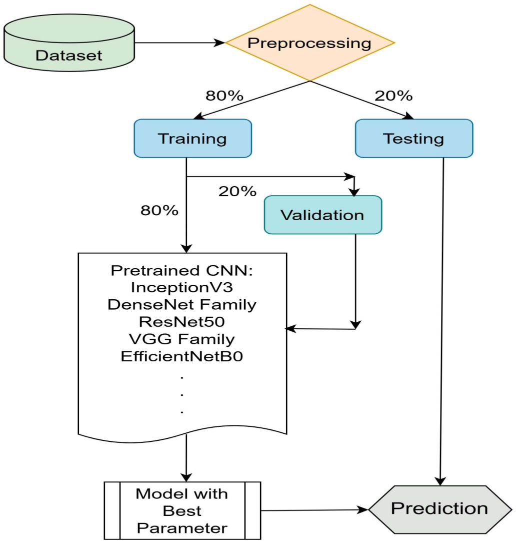


Fig. 1: Workflow of the Research Methodology

TABLE II: Parameters used in the pre-trained deep learning models

Parameter	Value
batch size	24
number of epochs	30
optimizer	adam
learning rate	0.0001
output classifier layer	softmax
activation function	relu
loss function	Categorical Cross Entropy

precision, recall, f1-score, support, and other class-specific evaluations for every class utilizing transfer learning using VGG16.

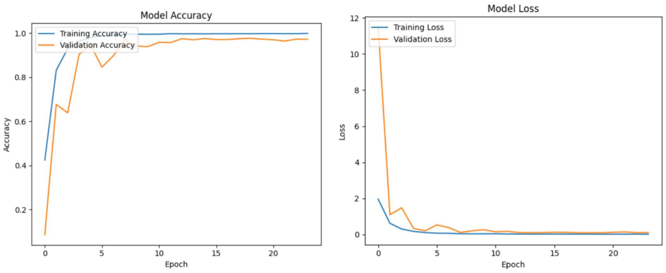
TABLE III: Analysis of CNN Architectures: Accuracy, Model Parameters, and Prediction Time on the Collected Dataset

Approaches	**Accuracy (%)**
InceptionV3	89.25
Densenet121	95.31
EfficientNetB0	94.92
EfficientNetB2	96.67
EfficientNetB6	95.80
InceptionResNetV2	90.72
ResNet50	95.80
VGG19	97.46
VGG16	97.65
ConvNeXtTiny	94.92
ConvNeXtXLarge	96.58

The training and validation accuracy and loss curves for the fusion model are shown in Fig. 2a and Fig. 2b.

TABLE IV: Measures of accuracy, precision, recall, f1-score, support, and other class-specific evaluations for every class utilizing transfer learning using VGG16, with a scale of 1.00 for 100% and 0.00 for 0%

Classes	Precision	Recall	F1-Score	Support
Typing only	1.00	0.96	0.98	52
Browsing Laptop	0.98	1.00	0.99	64
Makeup	0.95	0.97	0.96	36
Talking with Phone	0.97	0.97	0.97	73
Browsing Phone	0.97	0.99	0.98	74
Drinking	1.00	0.98	0.99	55
Writing	0.99	0.99	0.99	102
Looking Away	0.97	0.89	0.93	36
Reading	0.97	0.99	0.98	133
Eating	1.00	0.98	0.99	128
Drowsy	0.94	0.94	0.94	52
Towards Screen	0.96	0.98	0.97	56
Talking with Friends	0.97	1.00	0.99	75
Related Stuff	0.92	0.92	0.92	49
Empty Background	1.00	0.97	0.99	39
accuracy			0.98	1024
macro avg	0.97	0.97	0.97	1024
weighted avg	0.98	0.98	0.98	1024



(a) Accuracy Curve

(b) Loss Curve

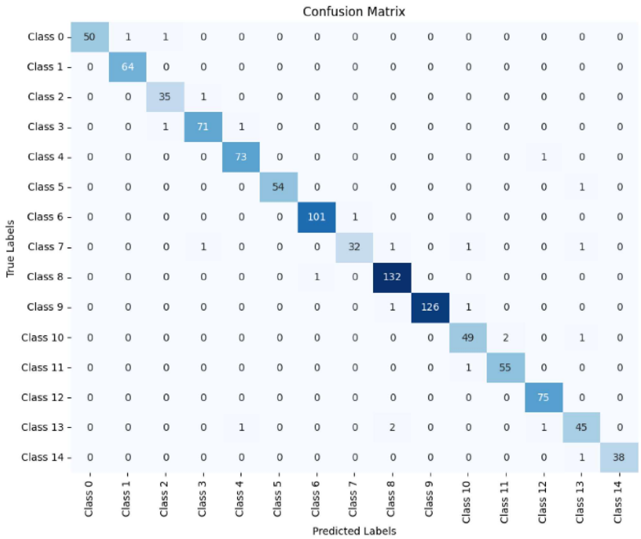
Fig. 2: Accuracy and Loss Curve for Proposed Model.

The confusion matrices in Fig. 3 illustrate the classification performance of fusion models

This result is a testament to the remarkable contributions of researchers who have worked on the renowned VGG16 model.

V. Conclusion

To sum up, this research provides a thorough analysis of tracking student involvement and identifying distractions in online learning settings. The urgent demand for precise student concentration monitoring was met by utilizing transfer learning strategies and a carefully selected dataset. Extensive testing and analysis produced remarkable accuracy rates; VGG16 performed best, with an accuracy of 97.65 or 98%. These results open the door for the creation of automated monitoring systems for online courses in addition to providing insightful information about how to improve student focus and engagement in the digital age.In order to provide a path for future research and development in this important area, this paper compares prominent transfer learning architectures and provides comprehensive performance data. The work creates new avenues for investigation, such as improving monitoring systems, incorporating real-time feedback systems, and expanding to suit other learning environments. The ultimate goal of these initiatives is to raise the standard of online education and give students everywhere a positive learning environment. Ultimately, this is only the start; there are still a lot of obstacles to overcome. Much progress has been made thus far with CNN's potent weapon. The virtual world and online education are about to enter a new age. There is yet the best to come.



	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	Class 11	Class 12	Class 13	Class 14
Class 0	50	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Class 1	0	64	0	0	0	0	0	0	0	0	0	0	0	0	0
Class 2	0	0	35	1	0	0	0	0	0	0	0	0	0	0	0
Class 3	0	0	1	71	1	0	0	0	0	0	0	0	0	0	0
Class 4	0	0	0	0	73	0	0	0	0	0	0	0	1	0	0
Class 5	0	0	0	0	0	54	0	0	0	0	0	0	0	1	0
Class 6	0	0	0	0	0	0	101	1	0	0	0	0	0	0	0
Class 7	0	0	0	1	0	0	0	32	1	0	1	0	0	1	0
Class 8	0	0	0	0	0	0	1	0	132	0	0	0	0	0	0
Class 9	0	0	0	0	0	0	0	0	1	126	1	0	0	0	0
Class 10	0	0	0	0	0	0	0	0	0	0	49	2	0	1	0
Class 11	0	0	0	0	0	0	0	0	0	0	1	55	0	0	0
Class 12	0	0	0	0	0	0	0	0	0	0	0	0	75	0	0
Class 13	0	0	0	0	1	0	0	0	2	0	0	0	1	45	0
Class 14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	38

Fig. 3: Confusion matrix for VGG16 on our dataset.

References

[1] T. AL Mseiedein, "Exploring factors affecting undergraduate students' acceptance of e-learning environment," vol. 22, p. 513, 11 2022.

[2] D. Peck, "Online learning statistics: The ultimate list in 2024," 1 2024.

[3] S. Ong and G. Quek, "Enhancing teacher–student interactions and student online engagement in an online learning environment," *Learning Environments Research*, vol. 26, 01 2023.

[4] N. Alruwais and M. Zakariah, "Student-engagement detection in classroom using machine learning algorithm," *Electronics*, vol. 12, no. 3, p. 731, 2023.

[5] M. Ali, X. De Zhang, and M. Harun-Ar-rashid, "Student activity detection using deep learning with yolov3," *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, vol. 8, pp. 757–769, 2020.

[6] M. N. Hasnine, H. T. Bui, T. T. T. Tran, H. T. Nguyen, G. Akçapınar, and H. Ueda, "Students' emotion extraction and visualization for engagement detection in online learning," *Procedia Computer Science*, vol. 192, pp. 3423–3431, 2021.

[7] E. Slyman, C. Daw, M. Skrabut, A. Usenko, and B. Hutchinson, "Fine-grained classroom activity detection from audio with neural networks," *arXiv preprint arXiv:2107.14369*, 2021.

[8] A. S. Pillai, "Student engagement detection in classrooms through computer vision and deep learning: A novel approach using yolov4," *Sage Science Review of Educational Technology*, vol. 5, no. 1, pp. 87–97, 2022.

[9] S. R. Kabir, A. Y. Srizon, N. T. Esha, M. F. Faruk, S. M. Hasan, M. N. Khansur, and M. M. Islam, "Student distraction detection and concentration analysis via transfer learned convolutional neural networks," in *2024 IEEE International Conference on Power, Electrical, Electronics and Industrial Applications (PEEIACON)*. IEEE, 2024, pp. 1–6.