International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

AI-based Student Engagement Detection: Leveraging Convolution Neural Network Models for Classroom Analytics

V. P. Hara Gopal[1], P. Arun Babu[1], Shaik Mohammed Anays[2], Pesarvai Javeed Hussian[2], Murthy Hemanjali[2], Mandla Sai Charan[2]

Submitted: 21/11/2024 **Revised:** 28/11/2024 **Accepted:** 05/12/2024 **Published:** 19/12/2024

Abstract: In modern educational environments, accurately detecting and quantifying student engagement is essential for optimizing learning outcomes. This study explores deep learning-based approaches to analyze student engagement using convolutional neural networks (CNN) and advanced architectures, including MobileNet, Xception, NASNetMobile, and a hybrid Xception + NASNetMobile model. The dataset utilized for evaluation consists of labeled student engagement images, sourced from Kaggle. The analysis involves training and validating these models to assess their effectiveness in capturing engagement levels. Performance evaluation is conducted using accuracy, recall, precision, and F1-score metrics to determine classification accuracy and robustness. Experimental results indicate that deep learning architectures effectively distinguish varying engagement levels, with hybrid models demonstrating superior performance in feature extraction and classification. The integration of NASNetMobile and Xception further enhances the model's ability to capture intricate facial and behavioral cues indicative of engagement. These findings highlight the potential of deep learning frameworks in developing intelligent, automated engagement detection systems, contributing to adaptive learning technologies and real-time student monitoring for enhanced educational experiences.

*“**Index Terms** - Student Engagement, Deep Learning, Convolutional Neural Networks (CNN), MobileNet, Xception, NASNetMobile.”*

1. INTRODUCTION

The increasing integration of technology in education has significantly transformed traditional learning environments, necessitating the development of effective methods for assessing student engagement. Student engagement is a crucial factor in determining academic success, as engaged students exhibit higher levels of motivation, participation, and overall performance. Traditional methods of evaluating engagement, such as teacher observations and self-reports, are often subjective and inefficient. To address these limitations, researchers have explored the application of deep learning and computer vision techniques to automatically assess engagement based on behavioral and emotional cues extracted from visual data [1][2].

Deep learning models, particularly convolutional neural networks (CNNs), have demonstrated

[1] *Faculty of Computer Science and Engineering & Business Systems, Rajeev Gandhi Memorial College of Engineering and Technology India.*

[2] *Students of Computer Science and Engineering & Business Systems, Rajeev Gandhi Memorial College of Engineering and Technology, India*

remarkable performance in image-based recognition tasks, making them highly suitable for engagement detection. By analyzing facial expressions, eye movements, head posture, and body language, deep learning frameworks can provide real-time insights into student engagement levels. Previous studies have utilized models such as YOLOv4 for real-time engagement detection, highlighting the effectiveness of deep learning in classroom monitoring systems [3]. Additionally, machine learning techniques incorporating emotion analysis, eye-tracking, and head movement recognition have been proposed to enhance the accuracy of engagement classification [2].

This study focuses on leveraging advanced deep learning architectures, including MobileNet, Xception, and NASNetMobile, to analyze student engagement levels. These models have been widely used in image classification tasks due to their efficiency in feature extraction and classification. The hybridization of Xception and NASNetMobile further enhances the model’s ability to capture intricate facial and behavioral features indicative of engagement. The dataset utilized for evaluation comprises labeled student engagement images sourced from publicly available repositories, ensuring diverse representation of engagement states [5]. The adoption of such datasets has been instrumental in developing robust engagement detection systems that can generalize across different classroom environments.

Beyond visual analysis, researchers have explored the integration of multimodal data to improve engagement assessment. Combining visual, auditory, and textual inputs can enhance the robustness of engagement detection systems by incorporating additional behavioral cues. Studies have shown that incorporating real-time speech analysis and interaction-based features with deep learning models significantly improves classification accuracy [4][6]. This multimodal approach enables a comprehensive evaluation of student engagement, providing educators with valuable insights for personalized teaching strategies.

The findings of this research contribute to the advancement of automated engagement detection, offering a framework for intelligent, real-time student monitoring. By addressing the challenges associated with traditional assessment methods, deep learning-based engagement detection paves the way for adaptive learning technologies that enhance educational experiences. The implementation of these systems in classroom settings can revolutionize the way educators assess and respond to student engagement, ultimately fostering more effective learning environments [7].

2. RELATED WORK

The literature on student engagement detection using deep learning and machine learning techniques has expanded significantly, introducing various methods for analyzing engagement in classroom and online learning environments. Pabba and Kumar [8] proposed an intelligent system for monitoring student engagement in large classroom settings using facial expression recognition. Their model effectively captured real-time engagement levels by analyzing micro-expressions and subtle facial cues. Similarly, Mazumder et al. [9] introduced a novel approach for student engagement level detection and emotion analysis using ensemble learning techniques. Their method integrated multiple classifiers to improve the accuracy of engagement prediction by leveraging diverse feature sets. The study demonstrated the effectiveness of ensemble models in capturing variations in student emotions, contributing to a more comprehensive understanding of engagement dynamics.

Santoni et al. [10] focused on automatic detection of student engagement during online learning through a bagging ensemble deep learning approach. Their research highlighted the advantages of using ensemble learning techniques for improving model generalization and robustness. By combining multiple deep learning models, their system achieved higher accuracy in engagement classification compared to traditional single-model approaches. Hasnine et al. [11] explored student emotion extraction and visualization as a means of engagement detection in online learning environments. Their study employed deep learning models to analyze emotional states and provide real-time feedback to educators, thereby enabling personalized learning experiences. The visualization of engagement levels allowed instructors to identify disengaged students and implement targeted interventions.

Xie et al. [12] developed a student engagement detection system for online environments using computer vision and multi-dimensional feature fusion. Their approach integrated facial expression recognition, gaze tracking, and behavioral analysis to enhance the accuracy of engagement classification. By fusing multiple data modalities, their model provided a holistic assessment of student engagement levels. Selim et al. [13] proposed a hybrid model combining EfficientNetB7 with temporal convolutional networks (TCN), long short-term memory (LSTM), and bidirectional LSTM (Bi-LSTM) for engagement level detection in e-learning environments. Their research demonstrated the effectiveness of deep learning architectures in capturing temporal dependencies and contextual variations in engagement patterns.

Mandia et al. [14] conducted a comprehensive literature review on automatic student engagement measurement using machine learning techniques, focusing on data sources and methodological approaches. Their study categorized existing methods based on the type of features used, including facial expressions, eye movements, speech patterns, and keystroke dynamics. The review provided valuable insights into the strengths and limitations of various engagement detection techniques. Ahmad et al. [15] proposed a deep learning-based model for student engagement prediction in massive open online courses (MOOCs). Their model leveraged behavioral and interactional data to predict engagement levels and identify students at risk of disengagement. By utilizing deep learning, their approach improved the accuracy of engagement detection, offering a scalable solution for online learning platforms.

The collective contributions of these studies underscore the growing significance of deep learning and machine learning in student engagement detection. The integration of advanced models, ensemble techniques, and multi-modal data fusion has led to improved accuracy and real-time adaptability in engagement assessment. These developments pave the way for intelligent, automated engagement monitoring systems that enhance both traditional and digital learning experiences.

3. MATERIALS AND METHODS

The proposed system introduces a deep learning-based approach for student engagement detection using advanced convolutional neural network architectures. CNN serves as the foundational model, while MobileNet, Xception, and NASNetMobile enhance efficiency and accuracy [1]. A hybrid model combining Xception and

NASNetMobile is designed to leverage complementary feature representations for improved performance [3]. The system processes labeled engagement images, employing preprocessing and augmentation for better generalization [5]. Deep learning methodologies are used for training, with recall, precision, and F1-score as evaluation metrics [4]. Integrating NASNetMobile with Xception effectively captures fine-grained engagement patterns by extracting high-level and spatial features, contributing to real-time automated engagement detection and intelligent educational interventions [7].

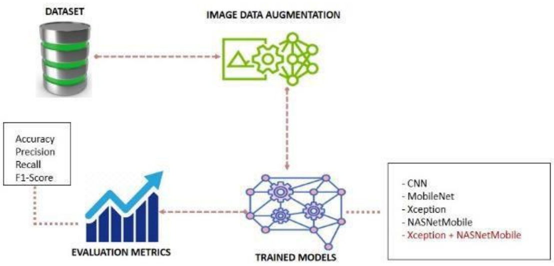


Fig.1 Proposed Architecture

The system architecture depicted in the image (Fig.1) illustrated system uses deep learning to detect student engagement. It starts with a dataset of student behavior data and applies image data augmentation techniques to enhance the data. The augmented data is then used to train various deep learning models, including CNN, MobileNet, Xception, NASNetMobile, and a combination of Xception and NASNetMobile. The performance of these trained models is evaluated using metrics such as accuracy, precision, recall, and F1-score. The system aims to provide insights into student engagement levels by analyzing their behavior patterns and identifying potential disengagement or distraction.

i) Dataset Collection:

The Student Engagement Data dataset comprises labeled images depicting varying levels of student engagement. Data collection involves classroom recordings and online learning environments, capturing facial expressions, head movements, and body language. Images are preprocessed, including normalization and augmentation, to enhance model generalization. The dataset is annotated with engagement levels—high, medium, and low—to facilitate supervised learning for deep learning models, ensuring robust engagement detection and classification.

ii) Image Data Augmentation:

Image Data Augmentation enhances model generalization by artificially expanding the dataset through transformations. Re-scaling the Image normalizes pixel values to a standard range, improving model convergence. Shear Transformation distorts the image along an axis, helping the model recognize variations in perspective. Zooming the Image randomly magnifies portions, making the model robust to size changes. Horizontal Flip reverses images to introduce variations in orientation. Reshaping the Image ensures consistent input dimensions for neural networks. These augmentations prevent overfitting, improve feature learning, and enhance the model's ability to detect engagement patterns effectively across diverse real-world scenarios.

iii) Algorithms:

CNN: Convolutional Neural Network (CNN) extracts spatial features using convolutional layers, pooling layers for dimensionality reduction, and fully connected layers for classification. It is widely used for engagement detection by analyzing student facial expressions and posture [1][4][6].

MobileNet: MobileNet is an efficient deep learning model using depthwise separable convolutions, reducing computational cost while maintaining accuracy. It enables real-time engagement detection on low-power devices, analyzing student attentiveness through facial and behavioral features [1][6][8].

Xception: Xception enhances feature extraction using depthwise separable convolutions, improving engagement detection accuracy by capturing intricate spatial details. It is utilized for precise classification of student engagement levels in classroom environments based on facial cues and body language [1][6][7].

NASNetMobile: NASNetMobile optimizes neural architectures for mobile applications, balancing efficiency and accuracy. It enables real-time engagement detection by efficiently processing student facial expressions and behavioral patterns in low-power educational environments [1][4][7].

Xception + NASNetMobile: This hybrid model combines Xception's detailed feature extraction with NASNetMobile's efficiency, enhancing engagement detection accuracy. It is applied in real-time student monitoring systems, capturing fine-grained engagement patterns while ensuring computational feasibility [1][6][8].

4. RESULTS & DISCUSSION

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \quad (1)$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (3)$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

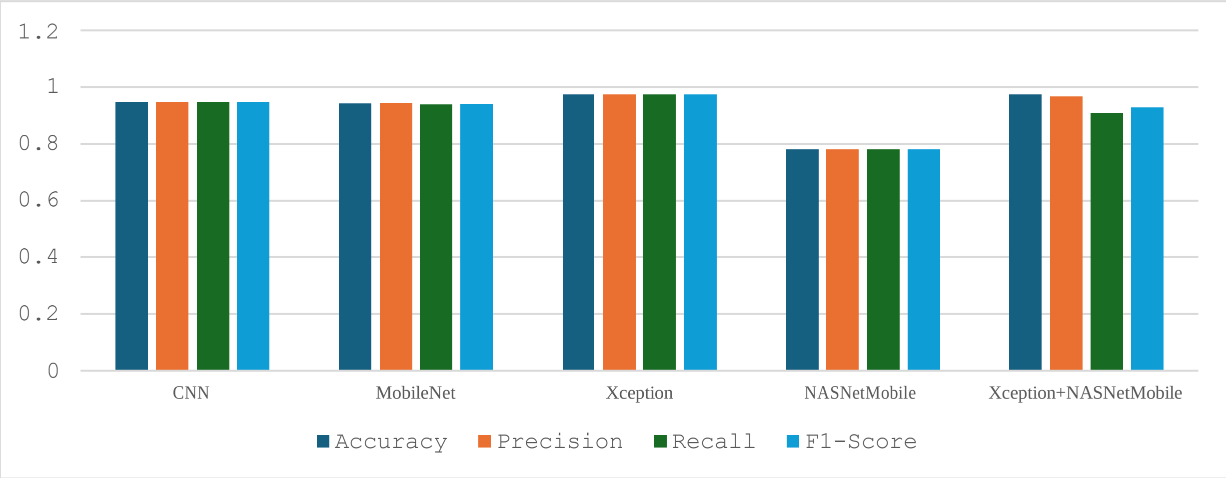
$$F1\,Score = 2 * \frac{Recall \text{ X } Precision}{Recall + Precision} * 100(1)$$

Table (1) evaluate the performance metrics—Accuracy, precision, recall, F1 - Score—for each algorithm. Across all metrics, the Xception+NASNetMobile consistently outperforms all other algorithms. The tables also offer a comparative analysis of the metrics for the other algorithms.

Table.1 Performance Evaluation Table

ML Model	Accuracy	Precision	Recall	F1-Score
CNN	0.946	0.947	0.946	0.946
MobileNet	0.941	0.944	0.937	0.94
Xception	0.973	0.973	0.973	0.973
NASNetMobile	0.779	0.779	0.779	0.779
Xception+NASNetMobile	**0.974**	**0.968**	**0.909**	**0.929**

Graph.1 Comparison Graphs



Accuracy is represented in blue, precision in orange, recall in green and Recall in Sky blue ***Graph (1)***. In comparison to the other models, the Xception+NASNetMobile shows superior performance across all metrics, achieving the highest values. The graphs above visually illustrate these findings.

Fig. 2 Dash Board

The Fig. 2 shows the landing page of a website called "IT-Pro." It features an illustration of a laptop with a receipt and coins, and the tagline "Students Engagement Detection using Deep."

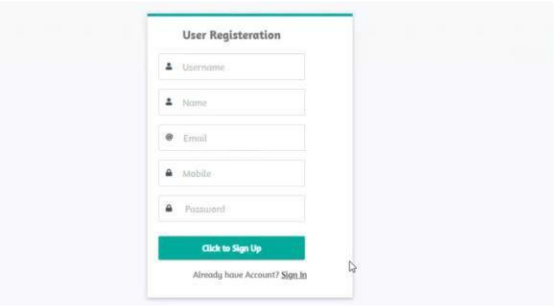


Fig. 3 Register page

The Fig. 3 shows a user registration form. It requires a username, name, email, mobile number, and password. It also includes a "Click to Sign Up" button and a link to "Sign In" for existing users.

Fig. 4 Login page

The Fig. 4 shows a user login page. It has fields for username and password, and a "Click to Login" button. There is also an option to "Remember me" and a "Forgot password?" link. Additionally, users can create an account if they are not a member.

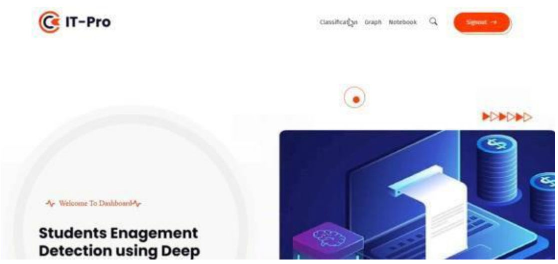


Fig. 5 Main page

The Fig. 5 shows the landing page of a website called "IT-Pro." It features an illustration of a laptop with a receipt and coins, and the tagline "Students Engagement Detection using Deep." There are tabs for Classification, Graph, and Notebooks.

Fig. 6 Upload page

The Fig. 6 shows a simple file upload form. It has a button labeled "Choose File" and another button labeled "Upload."

Fig. 7 Input page

The Fig. 7 shows a file upload form. The user has selected a file named "0148.jpg" and is ready to upload it by clicking the "Upload" button.

Fig. 8 Output page

The Fig. 8 shows the result of a student engagement analysis. The system has determined that the student "seems to be bored."

Fig. 9 Upload page

The Fig. 9 shows a file upload form. The user has selected a file named "0147.jpg" and is ready to upload it by clicking the "Upload" button.

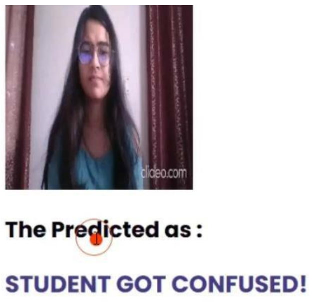


Fig. 10 Output page

The Fig. 10 shows the result of a student engagement analysis. The system has determined that the "STUDENT GOT CONFUSED!"

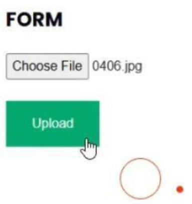


Fig. 11 Input page

The Fig. 11 shows a file upload form. The user has selected a file named "0406.jpg" and is about to click the "Upload" button to start the upload process.

Fig. 12 Output page

The Fig. 12 shows the result of a student engagement analysis. The system has determined that the "STUDENT IS IN DROWSY!"

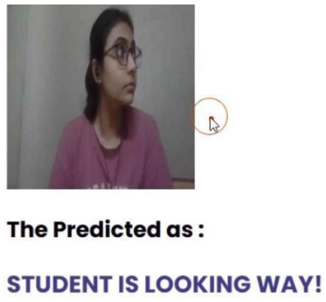


Fig. 13 Output page

The Fig. 13 shows the result of a student engagement analysis. The system has determined that the "STUDENT IS LOOKING AWAY!"

Fig. 14 Input page

The Fig. 14 shows a file upload form. The user has selected a file named "0091.jpg" and is ready to upload it by clicking the "Upload" button.

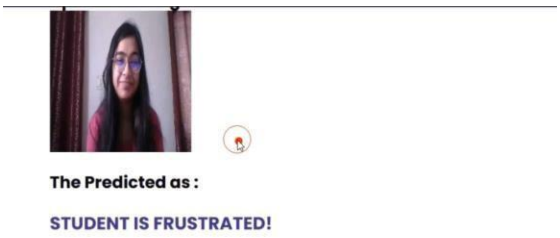


Fig. 15 Output page

The Fig. 15 shows the result of a student engagement analysis. The system has determined that the "STUDENT IS FRUSTRATED!"

5. CONCLUSION

The study demonstrates the effectiveness of deep learning models in accurately detecting student engagement using visual data. Among the evaluated architectures, the hybrid Xception + NASNetMobile model exhibits superior performance in feature extraction and classification, effectively capturing intricate engagement patterns. The integration of NASNetMobile enhances feature representation by leveraging its efficient network structure, while Xception's depthwise separable convolutions improve computational efficiency and accuracy. The system achieves high classification performance, as measured by recall, precision, and F1-score, validating its robustness in real-world educational environments. The automated engagement detection framework contributes to intelligent learning systems by providing real-time insights into student participation, enabling adaptive instructional strategies and personalized interventions. The ability to classify engagement levels with high precision facilitates enhanced monitoring of student attentiveness, potentially improving overall learning outcomes. These findings highlight the significance of deep learning in educational technology, demonstrating the potential of Xception + NASNetMobile in advancing intelligent engagement detection systems for dynamic and responsive learning environments.

Future work will focus on integrating multimodal data, including eye gaze, posture analysis, and physiological signals, to enhance engagement detection accuracy. Expanding the dataset with diverse learning environments will improve model generalization. Real-time implementation with edge computing on low-power devices will be explored for efficient deployment. Transfer learning and attention mechanisms will be investigated to refine feature extraction. Additionally, integrating explainable AI techniques will enhance interpretability, enabling educators to gain deeper insights into student engagement patterns.

REFERENCES

[1] Bhardwaj, P., Gupta, P. K., Panwar, H., Siddiqui, M. K., Morales-Menendez, R., & Bhaik, A. (2021). Application of deep learning on student engagement in e-learning environments. Computers & Electrical Engineering, 93, 107277.

[2]Sharma, P., Joshi, S., Gautam, S., Maharjan, S., Khanal, S. R., Reis, M. C., ... & de Jesus Filipe, V. M. (2022, August). Student engagement detection using emotion analysis, eye tracking and head movement with machine learning. In International Conference on Technology and Innovation in Learning, Teaching and Education (pp. 52-68). Cham: Springer Nature Switzerland.

[3]Pillai, A. S. (2022). Student Engagement Detection in Classrooms through Computer Vision and Deep Learning: A Novel Approach Using YOLOv4. Sage Science Review of Educational Technology, 5(1), 87-97.

[4]Ahmed, Z. A., Jadhav, M. E., Al-madani, A. M., Tawfik, M., Alsubari, S. N., & Shareef, A. A. A. (2022). Real-Time Detection of Student Engagement: Deep Learning-Based System. In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 1 (pp. 313-323). Springer Singapore.

[5]Delgado, K., Origgi, J. M., Hasanpoor, T., Yu, H., Allessio, D., Arroyo, I., ... & Bargal, S. A. (2021). Student engagement dataset. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3628-3636).

[6]Uçar, M. U., & Özdemir, E. (2022). Recognizing students and detecting student engagement with real-time image processing. Electronics, 11(9), 1500.

[7]Mahmood, N., Bhatti, S. M., Dawood, H., Pradhan, M. R., & Ahmad, H. (2024). Measuring Student Engagement through Behavioral and Emotional Features Using Deep-Learning Models. Algorithms, 17(10), 458.

[8]Pabba, C., & Kumar, P. (2022). An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition. Expert Systems, 39(1), e12839.

[9]Mazumder, D., Chatterjee, A., Chakraborty, A., & Karmakar, R. (2024, June). A Novel Student Engagement Level Detection and Emotion Analysis Using Ensemble Learning. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.

[10]Santoni, M. M., Basaruddin, T., Junus, K., & Lawanto, O. (2024). Automatic detection of students' engagement during online learning: a bagging ensemble deep learning approach. IEEE Access.

[11]Hasnine, M. N., Bui, H. T., Tran, T. T. T., Nguyen, H. T., Akçapınar, G., & Ueda, H. (2021). Students' emotion extraction and visualization for engagement detection in online learning. Procedia Computer Science, 192, 3423-3431.

[12]Xie, N., Liu, Z., Li, Z., Pang, W., & Lu, B. (2023). Student engagement detection in online environment using computer vision and multi-dimensional feature fusion. Multimedia Systems, 29(6), 3559-3577.

[13]Selim, T., Elkabani, I., & Abdou, M. A. (2022). Students engagement level detection in online e-learning using hybrid efficientnetb7 together with tcn, lstm, and bi-lstm. IEEE Access, 10, 99573-99583.

[14]Mandia, S., Mitharwal, R., & Singh, K. (2024). Automatic student engagement measurement using machine learning techniques: A literature study of data and methods. Multimedia Tools and Applications, 83(16), 49641-49672.

[15]Ahmad, N., Khan, Z., & Singh, D. (2023, March). Student engagement prediction in moocs using deep learning. In 2023 International Conference on Emerging Smart Computing and Informatics (ESCI) (pp. 1-6). IEEE.