Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-025-04997-x

ORIGINAL RESEARCH

Students' engagement prediction in online learning context via face emotion and data features with improved LinkNet and Bi-LSTM architecture

Rama Bhadra Rao Maddu[1] · **Murugappan S**[2]

Received: 25 December 2024 / Accepted: 18 July 2025

Abstract
Predicting student engagement is crucial for identifying the potential mental health challenges that may hinder academic performance and overall well-being. Early detection and intervention are essential to prevent detrimental effects on academic performance and overall quality of life. Effective strategies for predicting student engagement may involve analyzing various factors, this study employs a student engagement prediction model based on Improved LinkNet and Bidirectional Long Short-Term Memory (ImLN-Bi-LSTM), which considers face expression image and data features. Preprocessing, feature extraction, classification, and an engagement prediction mechanism are all included in the system. Face Expression image and data inputs perform individual pre-processing and feature extraction via distinctive approaches. The resultant features are then given to a hybrid classification model, utilizing Improved LinkNet and Bi-LSTM (Bidirectional Long Short-Term Memory) classifiers. The outcomes of ImLN-Bi-LSTM are the prediction results of student engagement in online learning. Comprehensive analyses including simulation and experimental assessments are conducted to validate the suggested ImLN-Bi-LSTM method. Moreover, at 80% of training data, the ImLN-Bi-LSTM model achieved a superior prediction accuracy of 0.951, and an F-measure of 0.905 which surpasses the result of traditional methods. The ImLN-Bi-LSTM model has the potential for use in online learning applications, and this study provides a solid and proven method for predicting student engagement.

Keywords Student engagement prediction · SLBT · I-EF · Improved LinkNet · Bidirectional long short-term memory

1 Introduction

The rise of computer and network technology has led to the widespread adoption of online learning in higher education (Sashank et al. 2023), presenting a novel approach to learning (Wang et al. 2022) (Ayouni et al. 2021). Research efforts focusing on the development of ITS show promise in enhancing educational experiences (Al Mamun and Lawrie 2023). Education remains a timeless subject of global significance (Liao et al. 2021). Within higher education institutions (Sobnath et al. 2020), students play a crucial role as stakeholders, and their achievements are of utmost significance (Oladipupo and Samuel 2024). The landscape of education has undergone substantial changes over time, particularly with a significant shift occurring due to the emergence of the COVID-19 pandemic (Thomas et al. 2022).

The integration of next-generation educational technologies, such as AIEd has expanded the scope of computer applications in education (Ouyang et al. 2023) (Deo et al. 2020). Recognizing the significance of student engagement as a cornerstone of effective learning, efforts have been directed towards leveraging real-time signals from students to enhance their learning experiences (Miller et al. 2021) (Yue et al. 2019). Predictive analytics within ITS can anticipate student challenges and provide timely interventions, such as hints or encouragement to improve learning

✉ Rama Bhadra Rao Maddu
ramamaddu7288@gmail.com

[1] Research scholar, Department of Computer and Information Science, Faculty of Science, Annamalai University, Annamalainagar, Chennai, Tamilnadu 608002, India

[2] Research Supervisor, Department of Computer and Information Science, Faculty of Science, Annamalai University, Annamalainagar, Chennai, Tamilnadu 608002, India

Published online: 04 August 2025

outcomes (Ruiz et al. 2022). The experience of human emotion is subjective and conscious, occurring when an individual aspect is either internal or external stimuli (Ngai et al. 2022).

While OL is increasingly prevalent in school education systems, its effectiveness is influenced by factors such as student attention spans and interactivity, which pose challenges in traditional classroom settings (Song et al. 2020). Despite difficulties in precisely capturing student engagement and interactivity in OL contexts, online education offers significant advantages in facilitating resource sharing and promoting educational equity (Xue and Niu 2023) (Wan et al. 2019) (Figueroa-Cañas and Sancho-Vinuesa 2020). These factors present challenges, particularly when compared to traditional classroom environments where direct interaction between students and teachers is more readily achievable. One of the primary challenges in OL is maintaining student engagement over extended periods (Savchenko and Makarov 2022) (Aydoğdu 2020). Additionally, the level of interactivity, or lack thereof, in online learning platforms, can hinder effective communication and collaboration among students and between students and instructors. This reduced interaction can impact learning outcomes and the overall effectiveness of the educational experience.

Additionally, advancements in ML enable the recognition of emotions through bio-signals, although real-time analysis remains challenging with typical equipment. DL technology has emerged as a prominent research area, particularly in image feature classification using advancements in computer software and hardware, AI technologies, and digital image processing (Abdulkader et al. 2023). The current emphasis on DL covers a wide range of areas, such as text, pronunciation, and visual components, and presents opportunities for enhancing teaching strategies. An automated approach to engagement prediction is essential, however, current image-based techniques face a key limitation in that they rely solely on spatial data from single images, which can lead to unstable and inconsistent results over time. Moreover, most existing engagement recognition datasets are limited in size, restricting progress and making it difficult to compare different methods (Selim et al. 2022). Although DL approaches like CNN-based image feature classification have shown promise, their application in student engagement detection is hindered by challenges such as poor generalizability across diverse datasets and data imbalance, which can negatively impact model performance (Flanagan et al. 2022). To overcome these challenges, this study proposes the ImLN-Bi-LSTM model, designed to improve student engagement prediction in OL environments. By combining enhanced feature extraction methods with advanced deep learning techniques, the model effectively captures subtle patterns in facial expressions and behavioral data, offering improved accuracy and robustness across diverse datasets. Thus, the model that has been introduced offers three unique contributions, which are briefly explained as follows.

- Improved SLBT include refining the feature extraction process to retrieve relevant features more effectively. This enhancement focuses on improving the LBP feature by incorporating binary patterns exclusively derived from neighboring pixel intensity values.
- Extracting the improved entropy-based feature efficiently during the feature extraction phase. Through the utilization of this Improved Deng entropy measure, the influence of alterations in data distribution on entropy calculations can be reduced, enabling accurate quantification of uncertainty or disorder within datasets.
- Employing a hybrid model in the classification phase with improved LinkNet and Bi-LSTM models. In the Improved LinkNet model, an improved loss function is employed in the decoder block. This improvement can capture intricate details within the input data. Furthermore, the utilization of an improved loss function within the decoder block aids in refining the model"s training process which leads to more precise predictions.

The remaining structure of the paper is organized as follows: Sect. 2 provides a comprehensive review of existing works. Section 3 then explores the ImLN-Bi-LSTM model"s architecture procedures. Sect. 4 then conducts experimental assessments of the suggested model. Lastly, Sect. 6 offers a summary of the entire procedure along with a closing observation.

2 Literature review

In 2023, (Ruiz et al. 2022) has proposed a video-based transfer learning approach to predict problem outcomes for students working with an intelligent tutoring system (ITS). In 2023, (Abdulkader et al. 2023) has presented an EBSAAS model. With this combination, the system uses the VGG-19 training structure for the extraction of features and to detect the attentiveness levels of the students. In 2023, (Ouyang et al. 2023) proposed an AI performance prediction method was integrated with learning analytics methods to increase student learning effects in a collaborative learning context. In 2023, (Al Mamun and Lawrie 2023) investigated an instructional design"s influence on the student-content interaction process within inquiry-based learning activities. In 2022, (Wang et al. 2022) suggested the CRRNN model for short-period activity characteristics and long-term changing patterns to predict potential at-risk students.

In 2023, (Xue and Niu 2023) has introduced a multi-output hybrid ensemble model from the SLCP to predict grades and student performance across various subjects. In 2021, (Ngai et al. 2022) developed a multimodal approach that utilized 2-channel EEG signals and the eye modality along with the face modality to improve recognition performance. In 2022, (Savchenko and Makarov 2022) employed an NN model for recognizing the students" emotions based on video images of their faces. In 2023, (Buono et al. 2023) has used a Long Short-Term Memory (LSTM) network for predicting student involvement levels using facial action units, gaze, and head positions. In 2023, (Gupta et al. 2023) has proposed a deep learning-based method that employs facial emotions to assess online learners" attention in real-time. In 2023, (Hossen and Uddin 2023) has developed a specialized system aimed at identifying and understanding student behaviors using XGBoost during online learning sessions. In 2025, (Naveen et al. 2025) has proposed a novel real-time detection framework that leveraged Transformer-enhanced Feature Pyramid Networks (FPN) with Channel-Spatial Attention (CSA), referred to as BiusFPN_CSA. However, the model is limited by its lack of integration of multimodal data inputs to provide a more comprehensive understanding of student behaviors.

3 Proposed framework for student engagement prediction via face emotion & data features

Predicting student engagement in online learning involves employing various data analysis techniques and models to anticipate how actively students will participate in their educational activities. For this purpose, this research proposed an ImLN-Bi-LSTM model for predicting student engagement in OL, which offers a more accurate, and reliable system. The model integrates two benchmark datasets namely the CKPLUS dataset (https:,, www.kaggle.com, datasets, shawon10, ckplus. xxxx), which provides facial expression images, and the OULAD dataset (https:,, www.kaggle.com, datasets, anlgrbz, student-demographics-online-education-dataoulad. xxxx), which contains student performance data. These datasets are treated as a unified input to support a multimodal analysis combining face expression images and cognitive data. These two types of inputs including facial expression images and student data are used to capture engagement. The proposed approach begins with a pre-processing step, which can be applied separately to each modality, where Gaussian filtering is used to enhance the quality of facial images, while min–max normalization standardizes the data. Then, the feature extraction is also performed independently, where the facial features are derived using I-SLBT and LGXP, while data features are extracted using I-EF and statistical features. The extracted features from both modalities are combined and then fed into a hybrid classification model that combines Improved LinkNet and Bi-LSTM for spatial feature analysis and temporal pattern recognition. The output from this process is used to predict student engagement levels, offering a comprehensive assessment based on both face expression images and performance-based indicators. Specifically, Fig. 1 shows the overall architecture of the proposed ImLN-Bi-LSTM approach for the prediction of student"s engagement in OL.

3.1 Pre-processing

The suggested model initiates with a crucial pre-processing phase, designed to optimize two distinct inputs such as face images and data. For face images, Gaussian filtering is employed to enhance clarity, while min–max normalization is applied to the data input to standardize its values within a specified range. Consider the input face image be $i^{face\ image}$ and data be i^{data}.

3.1.1 Pre-processing on face image via gaussian filtering

Gaussian filtering is a widely used method for noise reduction in image processing. Since the conventional filtering methods often distort subtle facial textures and perform inconsistently under varying lighting conditions or image compression. To mitigate these issues, Gaussian filtering is employed as a preprocessing step. This technique effectively attenuates high-frequency noise, resulting in a smoother image with reduced artifacts. The improved image quality facilitates more accurate subsequent processes such as feature extraction and engagement prediction. Gaussian filtering operates by using a Gaussian kernel, a two-dimensional distribution function. During the convolution process, the Gaussian kernel is slid over each pixel in the image, and the results of the Gaussian distribution across those locations are used to determine the weighted average of the nearby pixels. The input face image $i^{face\ image}$ is given as an input for this Gaussian filtering technique (Sekehravani et al. 2020) and its mathematical calculation for noise elimination is described in Eq. (1).

$$Gaussian_{filter}(a,b) = \frac{1}{2\pi Std^2} \exp\left(-\frac{a^2+b^2}{2Std^2}\right) \tag{1}$$

From Eq. (1), a and b denotes the horizontal and vertical axis distance and Std implies the standard deviation. Therefore, the input face image $i^{face\ image}$ is pre-processed by Gaussian filtering and its outcome is denoted as, $p^{face\ image}$.

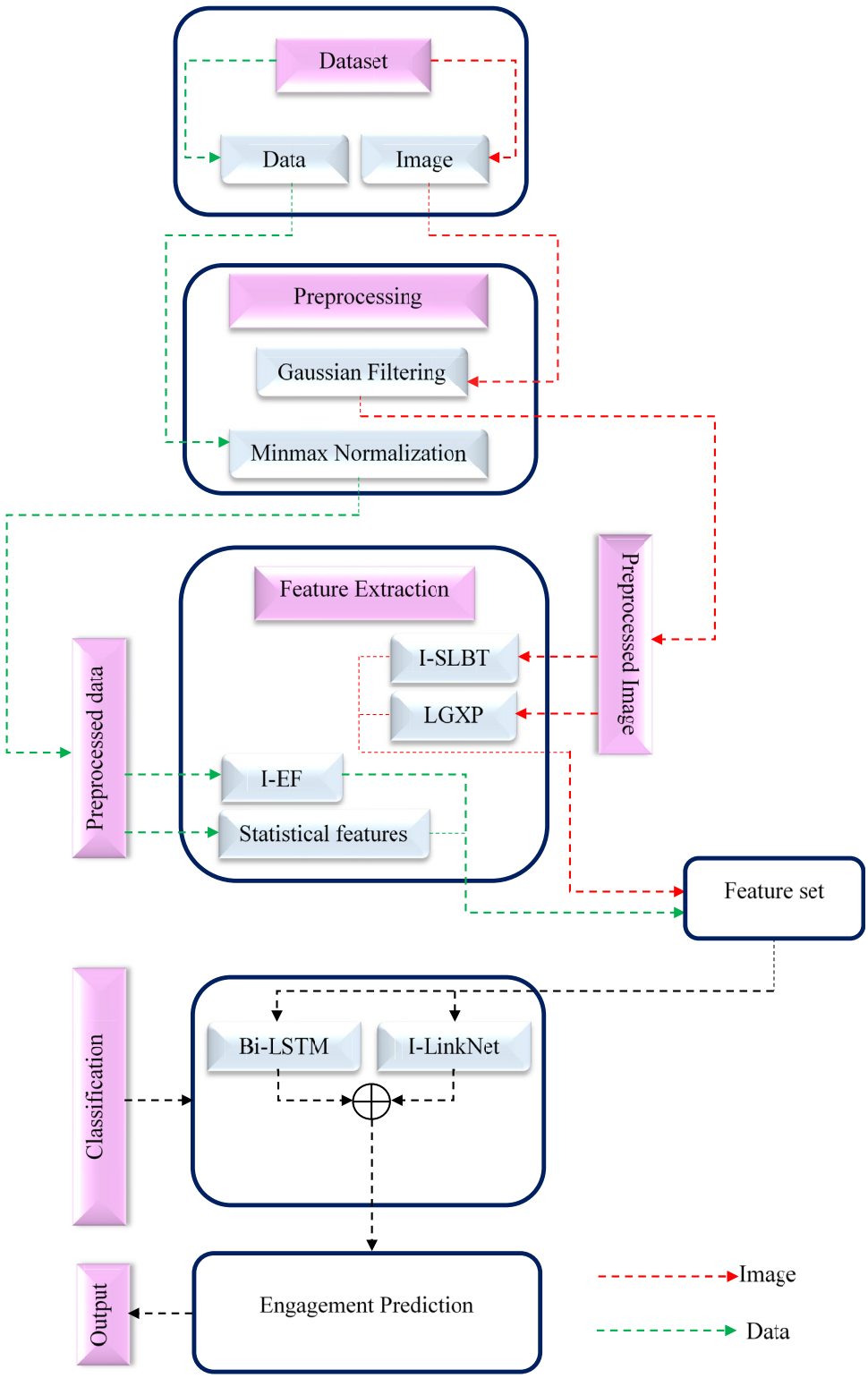

Fig. 1 Overall framework of suggested ImLN-Bi-LSTM approach for predicting student engagement

3.1.2 Pre-processing on input data via minmax normalization technique

Min–max normalization is a technique of data preprocessing usually applied to rescale mathematical features to an exact range, typically between 0 and 1. It is also known as min–max scaling. The traditional normalization techniques often assume that the input data follows a normal distribution. However, in cases where the data is skewed or contains outliers, standardization may yield inaccurate results.

To overcome these limitations, Min–Max normalization is employed. This method is particularly advantageous when dealing with features that vary in scale, as it ensures that all features contribute equally to the analysis. Additionally, it preserves the original distribution and shape of the data, thereby maintaining the intrinsic relationships among features and contributing to the precise extraction of features. The process involved in this normalization technique is calculating the minimum and maximum values of individual features in the dataset. By the way, the input data i^{data} is provided as an input for this technique (Henderi et al. 2021). Then, the following formula is applied for each data point in the feature as given in Eq. (2).

$$A_{new} = \frac{A - \min(A)}{\max(A) - \min(A)} \tag{2}$$

From Eq. (2), normalized results from the new value are implied as A_{new}, the old value is denoted as A, dataset maximum and the minimum value is represented as $\max(A)$ and $\min(A)$. By using min–max normalization, the range of values for each feature is constrained to the interval [0, 1], with the minimum value mapped to 0 and the maximum value mapped to 1. Consequently, the input data i^{data} is pre-processed by min–max normalization and its outcome is indicated as, p^{data}.

3.2 Feature extraction

During this stage, four relevant features are extracted independently from pre-processed inputs, such as pre-processed face images $p^{face\ image}$ and pre-processed data, p^{data}.

3.2.1 Feature extraction on pre-processed face image

At this phase, the pre-processed face image $p^{face\ image}$ undergoes feature extraction, where appropriate features like I-SLBT and LGXP are retrieved.

3.2.1.1 I-SLBT In this work, an improved SLBT method is proposed, which focuses on local binary patterns to effectively capture local texture and shape features in images. In SLBT, each pixel"s encoding depends on its intensity relative to neighboring pixels (Lakshmiprabha and Majumder xxxx). Here, the pre-processed face image $p^{face\ image}$ is provided as input for this phase. Assume $p^{face\ image}$ denote N set of training images with W landmark points of the shape. Shape variations are acquired through the alignment of landmark points, followed by PCA on these points. In the training set, any shape vector can be expressed in Eq. (3).

$$W \approx \overline{W} + E_s^v F_s \tag{3}$$

$$F_s = E_s^{v^T}\left(W - \overline{W}\right) \tag{4}$$

From Eq. (3), the mean of the shape is represented by $\overline{W}$, E_s^v includes the vectors of eigen for largest eigenvalues (κ_s) and the parameter model of the shape is implied as F_s(s means shape in F_s). It is likely to evaluate the model parameter shape related to an image by rewriting Eq. (4). In texture modeling, a shape-free patch is created by transforming every set of training images into the mean shape. Texture modeling in AAM is done using the shape-free patch direct intensity values. On the other hand, SLBT uses LBP on the shape-free patch to extract features that are noise-invariant and illuminated. Assume a 3×3 window with a central pixel (r_c, s_c), gr_c denotes the value of intensity and $Q = q(gr_i)$ denotes the local texture here $gr_i\ (i = 0, 1, 2, 3, 4, 5, 6, 7)$ related to grey values for nearby pixels. These pixels are thresholded via center value gr_c as $q(a(gr_0 - gr_c), ..., a(gr_7 - gr_c))$ and the function is described in Eq. (5).

$$a(n) = \begin{cases} 1 & , n > 0 \\ 0 & , n \le 0 \end{cases} \tag{5}$$

$$LBP_{(r_c,s_c)} = \sum_{i=0}^{7} a(gr_i - gr_c)2^i \tag{6}$$

By the way, the above Eqs. (5) and (6) are the conventional LBP equation. The conventional LBP may not adequately represent the richness and diversity of texture information in the data. Therefore, an Improved LBP (Bavkar et al. 2022) is proposed to overcome these issues, and its expression is defined in Eq. (7).

$$ILBP = \left[\frac{\sum_{i=0}^{7} a(gr_i - gr_c)\,2^i + t.2^{gr_i - 1}}{\left(\sum_{i=0}^{7} gr^{i-1}\right) * gr_c}\right] \tag{7}$$

From Eq. (7), the pixel value of the neighbor and center is denoted as gr_i and gr_c, the function t is defined in Eq. (8).

$$t = \begin{cases} 1 & if\ TM \ge gr_i\ and\ AM \le gr_c \\ 0 & else\ if\ AM \ge gr_i\ and\ AM > gr_c \\ 1 & else\ if\ AM < gr_i\ and\ gr_i \le gr_c \\ 0 & else\ if\ TM \le gr_i\ and\ gr_i \ge gr_c \end{cases} \tag{8}$$

From Eq. (8), AM represents the arithmetic mean and TM denotes the Trimmed Mean. Following that, the TM expression is given in Eq. (9).

$$TM = \frac{1}{x-2y}\sum_{i=y+1}^{x-y} f\left(TM_{(i)}\right) \tag{9}$$

Using Eq. (7), the center pixel gr_c of the LBP pattern can be attained. Consider K be the histogram feature of LBP for the images training set. Subsequently, texture modelling is applied via PCA similar to shape modelling given in Eq. (10).

$$F_t = E_t^T\left(K - \overline{K}\right) \tag{10}$$

From Eq. (10), the parameter of the texture model is implied as F_t, eigenvectors are denoted as E_t, and the vector of the mean is notated as $\overline{K}$. After that, the parameter of shape and texture vector is attained via Eq. (11). The weight calculation for each shape parameter is represented by the diagonal matrix D_s takes into account the differing units of shape and texture values. To derive the shape texture parameter, which governs texture, global, and local shape characteristics which is employed in PCA on the merged parameter vector as described in Eq. (12).

$$F_{st} = \begin{pmatrix} D_s E_s \\ E_t \end{pmatrix} \tag{11}$$

$$c = E_{st}^T\left(F_{st} - \overline{F}_{st}\right) \tag{12}$$

From Eq. (12), the parameter of the shape texture is signified as c. Thus, the I-SLBT feature is implied as $I_{SLBT}^{face\ image}$.

3.2.1.2 LGXP Utilizing the LXP operator to encode the Gabor stage into LGXPs makes it easier to retrieve texture information at different scales and orientations. Incidentally, the pre-processed image is given as input for this LGXP feature. To model an LGXP pattern, typically the phases are first quantized into a diverse range. Then, the phases of the central pixel and those of its neighbors are quantized via the LXP operator (Shanthi and Koppu 2023). Finally, the resulting binary labels are concatenated to form a local pattern of the central pixel which are in decimal and binary form, as described in Eq. (13) and (14).

$$LGXP_{\eta,o}^{feat}\left(gr_c\right) = \left[LGXP_{\eta,o}^{s}, LGXP_{\eta,o}^{s-1}, ..., LGXP_{\eta,o}^{1}\right]_{binary} \tag{13}$$

$$LGXP_{\eta,o} = \left[\sum_{i=1}^{s} 2^{i-1}.LGXP_{\eta,o}^{i}\right]_{decimal} \tag{14}$$

From Eqs. (13) & (14), the central pixel"s location in the gabor phase is denoted as pi_c, orientation and scale is implied as o and η, the neighborhood size is represented as s and the pattern evaluated between center gr_c and neighbor pixel gr_i is expressed as $LGXP_{\eta,o}^{i}\left(i = 1, 2, ..., s\right)$, also its calculation is defined in Eq. (15).

$$LGXP_{\eta,o}^{i} = Q\left(\phi_{\eta,o}\left(gr_c\right)\right) \otimes Q\left(\phi_{\eta,o}\left(gr_i\right)\right), i = 1, 2, ..., s \tag{15}$$

From Eq. (15), phase is implied as $\phi_{\eta,o}$, operator of LXP is signified as $\otimes$ which is depend on the operator of XOR as given in Eq. (16). Furthermore, the operator of quantization is denoted as Q and the phase quantized code is calculated based on the specified quantity of phase ranges as outlined in Eq. (17).

$$m \otimes n = \begin{cases} 0 & , if\ m = n \\ 1 & , else \end{cases} \quad \text{i.e.,} \quad Q\left(\phi_{\eta,o}\right) = i; \tag{16}$$

$$if \frac{360 * i}{p^r} \leq \phi_{\eta,o} < \frac{360 * (i+1)}{p^r}, i = 0, 1, ..., p^r - 1 \tag{17}$$

Here, the amount of phase range is implied as p^r. Following the definition of the pattern, a pattern map is calculated for every Gabor kernel. Subsequently, these pattern maps are partitioned into sub-blocks of non-overlapping. The histograms of sub-blocks at different scales and orientations are aggregated to generate the LGXP descriptor for the input face image as outlined in Eq. (18).

$$h^i = \left[h_{\eta_0,o_{0,1}}^i, ..., h_{\eta_0,o_{0,n}}^i; ...; h_{\eta_{0-1},o_{r-1},1}^i, ..., h_{\eta_{0-1},o_{r-1},n}^i\right] \tag{18}$$

From Eq. (18), the i^{th} sub-block of the LGXP map of the histogram with o orientation and scale η is given as $h_{\eta,o,i}^i\left(i = 1, 2, ..., n\right)$. Thereby, the LGXP feature is signified as $I_{LGXP}^{face\ image}$. Finally, the facial features including I-SLBT and LGXP obtained from the preprocessed face image is represented as F_{img}.

3.2.2 Feature extraction from pre-processed data

In this phase, pertinent features like statistical features and I-EF are retrieved from the pre-processed data p^{data}.

3.2.2.1 Statistical features Statistical features such as mean, median, and standard deviation are derived from the pre-processed data p^{data}.

Mean: It evaluates the average of the values in the dataset (Toptaş and Hanbay 2021) as described in Eq. (19).

$$M = \frac{1}{a}\sum_{i=1}^{a} p^{data} \tag{19}$$

Median: It is a measure of central tendency that represents the middle value of a dataset when it is ordered from smallest to largest (Abu et al. 2020). As indicated by Eq. (20), it splits the dataset in half evenly, with 50% of the values lying below and the other half above the median.

$$Median = \frac{j}{2} \tag{20}$$

Standard deviation: It measures the distribution of values in a dataset, indicating how much the single values differ from the mean as defined as per Eq. (21).

$$S_{deviation} = \sqrt{\frac{1}{a}\sum_{i=1}^{a}\left(p^{data} - M\right)^2} \tag{21}$$

From above Equation, M represents the mean and a implies the overall number of values within the dataset. Hence, the statistical features are obtained and its result is denoted as $I_{ST_f}^{data}$.

3.2.2.2 I-EF An I-EF entropy is proposed which can capture complex relationships and dependencies within engagement data more effectively. Entropy is a metric for quantifying uncertainty in a dataset with a randomly generated variable having a probability mass function. Equation (22) outlines the conventional formulation of entropy (Yan and Deng 2020) which provides a mathematical expression to capture the degree of disorder within the dataset in which the pre-processed data p^{data} is given as an input.

$$E_i = -\sum_{i=1}^{Z} \mathrm{P}\left(p^{data}\right) \log_B \mathrm{P}\left(p^{data}\right) \tag{22}$$

In order to address the problem of reliance on the spread of data within a dataset, Eq. (24) suggests an enhanced Deng entropy. Here, the I-EF is performed via two steps.

Step 1: This alternative formulation aims to address the variability in entropy values resulting from alterations in the data distribution.

$$IE = \left(\sum w_x(i) \times E_i\right) - \left[-\sum_{G\subseteq\varphi} m(G)\log_2\left(\frac{m(G)}{2^{|G|}-1}e^{\frac{|G|-1}{|s|}}\right)\right] \tag{23}$$

$$IE = \left(\sum w_x(i) * \sum_{G\subseteq\varphi} m(G)\log_2\left(\frac{m(G)}{2^{|G|}-1}e^{\frac{|G|-1}{|s|}}\right)\right) \tag{24}$$

From Eqs. (23) and (24), $w_x(i) = 2\left(1 - \frac{1}{1+\exp(-E_i)}\right)$ implies the weight function (Munagala and Kodati 2021), the function of mass is implied as m on the frame of φ discernment, the focal element of m is implied as G, the cardinality of G is signified as $|G|$.

Step 2: By the way, the proposed entropy is satisfied via the entropy correction (Kermani and Plett xxxx) condition which is provided in Eq. (25).

$$|IE| < E_i \tag{25}$$

From Eq. (25),IE and E_i denotes the Improved entropy and conventional entropy. Thus, the I-EF feature is achieved and its result is depicted as I_{I-EF}^{data}. Finally, the face features (SLBT and LGXP features)F_{img} are attained from $p^{face\ image}$ and the data features (Statistical features, and I-EF) F_{data} are obtained from p^{data} are merged and totally signified as fe_d^i.

3.3 Classification via hybrid ImLN-Bi-LSTMmodel

In this phase, the extracted features fe_d^i derived from input face expression images and data are fed to the proposed hybrid classification model, which integrates the Improved LinkNet and Bi-LSTM models. Traditional models like CNNs excel at processing local patterns, but they often struggle with learning long-term temporal dependencies and may fail to capture the spectral relationships within multimodal data. In contrast, the Bi-LSTM model effectively manages long-range temporal dependencies in the extracted features, allowing it to capture context from both past and future inputs. This ability significantly enhances the model"s capacity to identify and predict complex engagement patterns. On the other hand, Improved LinkNet boosts prediction performance through its exceptional capability to capture both spectral and spatial information. By combining these two models, the proposed strategy leverages their complementary strengths, resulting in more precise and robust engagement predictions.

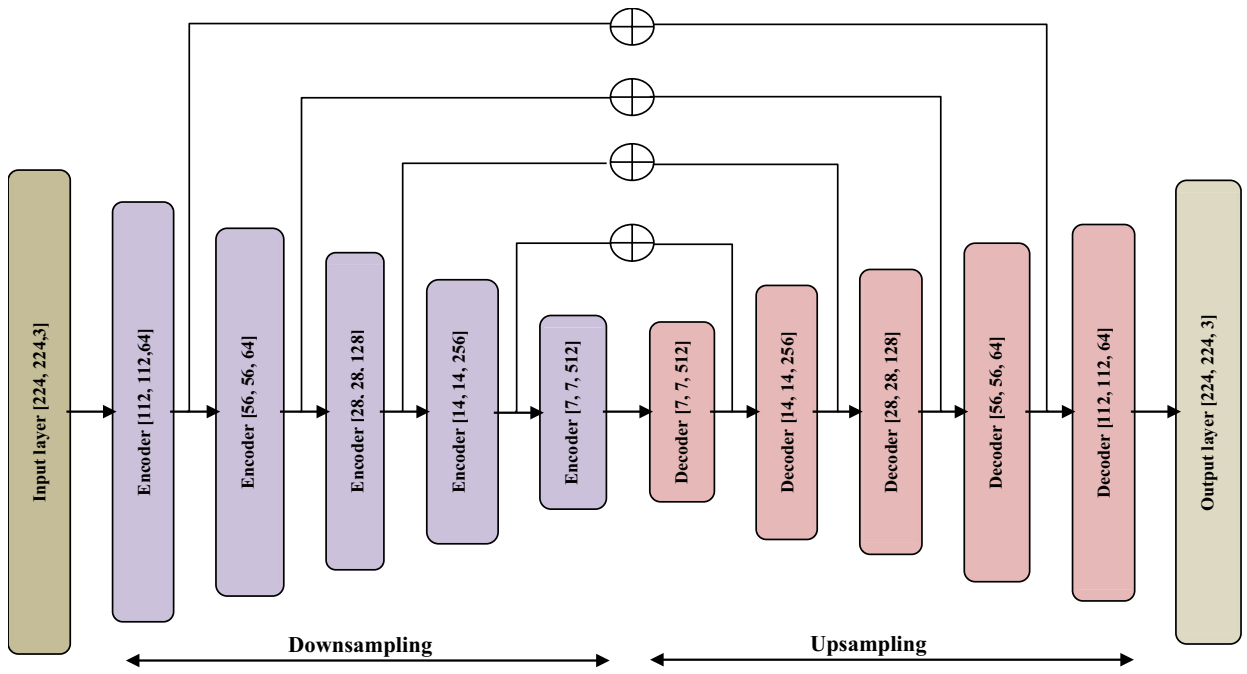

Fig. 2 Architecture of conventional LinkNet model

3.3.1 Improved LinkNet model

In computer vision, LinkNet is a kind of CNN architecture that is frequently employed for semantic classification tasks (Ramasamy et al. 2023). The architecture of the conventional LinkNet model is shown in Fig. 2**.**

Therefore, to classify the student engagement prediction, an Improved LinkNet is used. This procedure is fully explained in the section that follows. Figure 3 displays the architectural diagram of the Improved LinkNet model**.**

3.3.1.1 Encoder block The input extracted features fe_d^i is given as an input for the encoder block. After that, zero padding is a technique used to pad the input feature map with zeros along the spatial dimensions (height and width) before applying convolution. This is an additional convolutional layer that is applied to the result of the preceding phase, just like Convolution Layer 1. Convolution Layer 2"s output is subjected to a second batch normalization layer in order to stabilize and normalize the activations. Next, the batch normalization layer"s output is subjected, element by element, to the ReLU activation function. Similar to the initial zero padding, this step pads the feature map with zeros. This is the final convolutional layer (convolution layer 3) within the encoder block. From the given input feature map, more features are retrieved. After Convolution Layer 3, an additional batch normalization layer is added to stabilize and normalize the activations. Finally, Soft sign is an activation function that is similar to tanh but has a smoother gradient. This process likely involves applying the soft sign function element-wise to the output of the batch normalization layer.

3.3.1.2 Decoder block The decoder block receives the output from the encoder block as its input. The first convolutional layer within the decoder block applies convolutional operations to the input features from the encoder. Similar, to the encoder block, batch normalization is applied to the output of convolution layer 1 to normalize and stabilize activations. In the context of the decoder block, sigmoid activation might be used to confirm that the outcome values are in the range suitable for representing probabilities, especially if the task involves binary classification or pixel-wise segmentation. Then, the upsampling is performed as a process of increasing the spatial resolution of the feature map. Following upsampling, another convolutional layer (convolution layer 2) is applied to further refine the feature map. The outcome of Convolution Layer 2 is then subjected to batch normalization in order to stabilize and normalize the activations. After that, the Leaky ReLU activation function is used which permits a small, positive gradient when the input is negative. This is the final convolutional layer (convolution layer 3) within the decoder block. It further refines the features and prepares the output for the final prediction. Ultimately, an improved loss function is suggested, which may incorporate additional constraints or penalties to better address the characteristics of the problem at hand, such as class imbalance or spatial coherence. Here, an improved loss function has been employed by combining the cosh

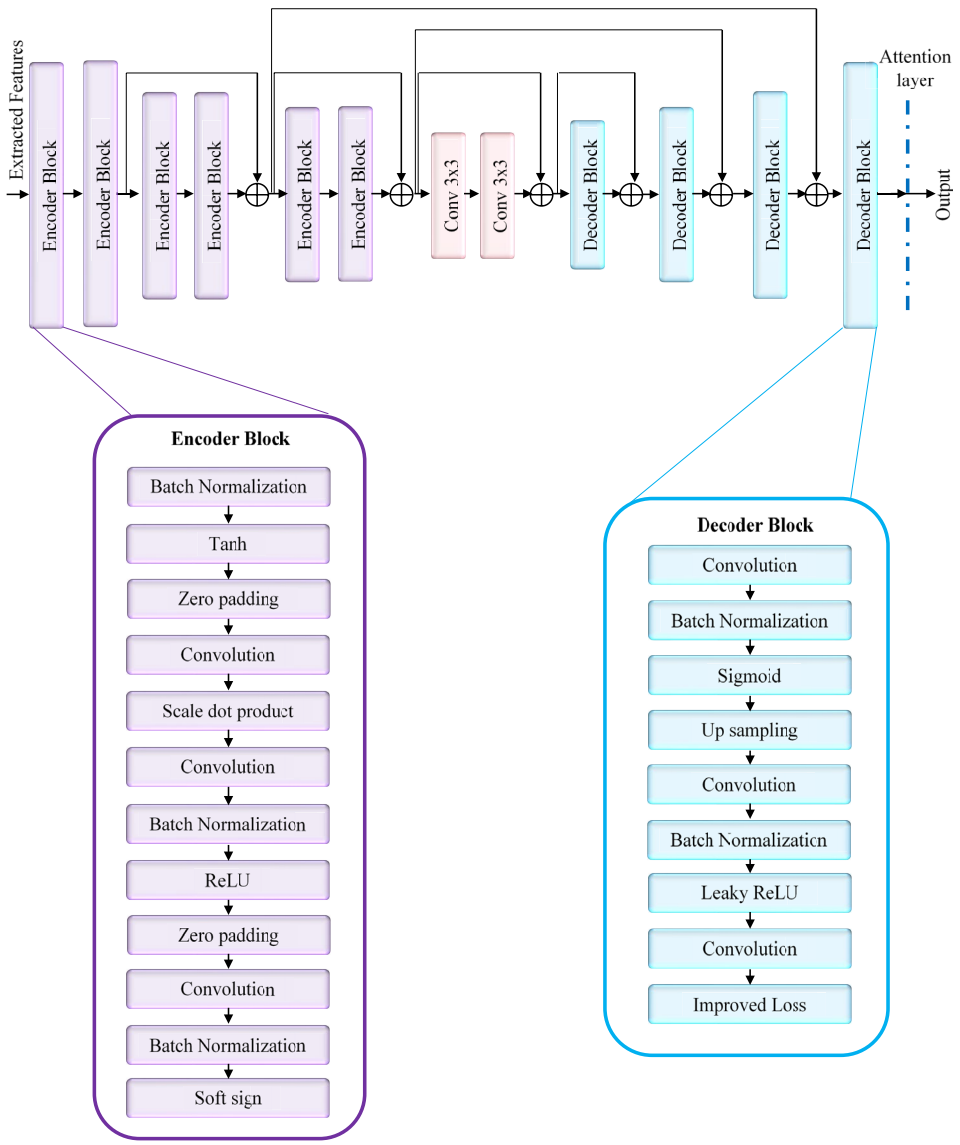

Fig. 3 Architecture model of Improved LinkNet model

loss, Dice loss, and balanced cross-entropy as defined by Eq. (26) and (27).

$$I_{lossFun} = [\log(\cosh(dice\ loss))] + Balanced\ cross\ entropy \tag{26}$$

$$= \left[\begin{array}{l}\left\{\log\left(\frac{e^{f}+e^{-f}}{2}\right)\left(1-\frac{2q\hat{p}+1}{q+\hat{p}+1}\right)\right\} \\ +\left[-(\lambda * q\log(\hat{q}))+(1-\lambda)*(1-q)\log(1-q)\right]\end{array}\right]/2 \tag{27}$$

From Eq. (27), $\lambda = \frac{1}{2}$, $\hat{q}$ *or* $\hat{p}$ implies the predicted value, and q *or* p denotes the actual value. Thereby, the outcome attained from this decoder layer that is feature map is denoted as Fe_m.

3.3.1.3 Attention layer The output of the decoder block or the final feature maps from the decoder may serve as the input to the attention layer (Hamdi et al. 2022). The significance of various feature map components is indicated by attention weights, which are calculated at each time by the attention layer. A weighted summation of the feature maps is then computed using the attention weights, with features with more attention weights contributing more to the final representation. Thus, the scores alignment expression is described by Eq. (28).

$$Scores_t = \tanh(Fe_m.We_{Att} + Bi_{Att}) \tag{28}$$

where, the attention layer trainable weights and bias are denoted as We_{Att} and Bi_{Att}. Following that, the scores $Scores_t$ undergo normalization through the softmax function to derive attention weights δ_t as described in Eq. (29).

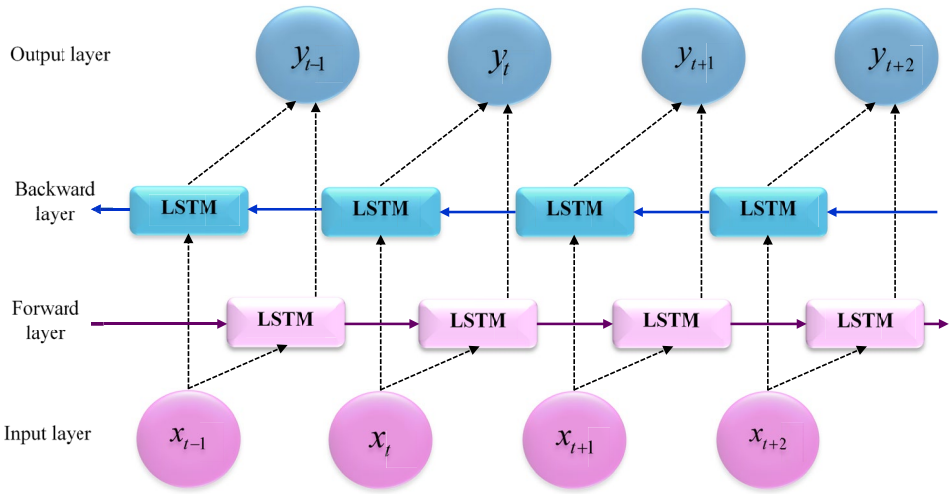

Fig. 4 Architecture of the Bi-LSTM model

$$\delta_t = Soft_{\max}\left(Scores_t\right) \tag{29}$$

Once attention weights are computed, the subsequent step involves the calculation of the context vector which is referred to as the attention vector as defined in Eq. (30). This operation entails the aggregation of T neurons via a weighted sum.

$$A_t = \sum_{i=1}^{T} \delta_t Ne_t \tag{30}$$

From Eq. (30),Ne_t denotes the neurons. Thereby, the Improved LinkNet model classifies the student engagement prediction that is denoted as $I_{linkNet}^{i,d}$.

3.3.2 Bi-LSTM model

The Bi-LSTM architecture comprises two interconnected layers with each Bi-LSTM predicting the sequence of each element based on the context of preceding and succeeding elements (Hamayel and Owda 2021). The forward function of the Bi-LSTM is characterized by l units as inputs and H_i as the number of hidden units that are computed via Eqs. (31), and (32). Figure 4 illustrates the architecture of the Bi-LSTM. Within the Bi-LSTM network, the hidden layer preserves two sets of values: one for the forward calculation (A) and another for the reverse calculation (A transpose). The outcome value y relies on both A and A transpose.

$$a_c^t = \sum_{j=1}^{l} fe_d^i we_{lc}^{lstm} + \sum_{c,t>0}^{H_i} Bi_{c\prime}^{t-1} we_{c\prime c} \tag{31}$$

$$a_c^t = \phi_c\left(a_c^t\right) \tag{32}$$

Table 1 Outcomes from the ImLN-Bi-LSTM model

(Data, image)	Target
(Pass, surprise)	0
(Withdrawn, sad)	1
(Fail, fear)	2
(Distinction, happy)	3

Table 2 Hyper-parameter setting of the classifier

Model	Hyperparameter
ImLN-Bi-LSTM	units:128, Drop-out Rate:0.2, batch size = 128, epochs = 25, verbose = 1

From Eqs. (31) and (32), c_t denotes the current input with l units, fe_d^i represents the input, H_i implies the hidden state, and we^{lstm} signifies the weight. Thus, the result obtained from this Bi-LSTM model is denoted as $Bi_{LSTM}^{i,d}$. By the way, the suggested ImLN-Bi-LSTM technique effectively classifies the student engagement prediction via face expression image and data inputs. Hence, the outcomes for the hybrid classification model that is tabulated in Table 1**.** Also, Table 2 shows the hyper-parameter settings of the classifier.

3.4 Engagement prediction

Predicting student engagement using an emotion index involves employing data-driven methods to anticipate students" emotional states and participation levels in educational tasks. In the ImLN-Bi-LSTM model, four classes of emotions are identified, each further divided into engaged and not engaged categories. By the way, the four classes are (Distinction, Happy), (Withdrawn, Sad), (Fail, Fear), and (Pass, Surprise) as well as their labels are (0,1,2,3).

Utilizing the predicted emotions as input in the engagement phase is determined and the engagement index is computed via Eq. (33).

$$Eng_{index} = E_p \times We_{emotion} \tag{33}$$

here, the probability of the emotion is signified as E_p i.e., $E_p = Emotion\ \Pr obability\left(\begin{array}{c} Emotion = (Distinction, happy), (Withdrawn, Sad), \\ (Fail, fear)\,\& \\ (Pass, Surprise) \end{array}\right)$ and its corresponding weight is denoted by $We_{emotion}$.

Subsequently, the targeted emotion with its corresponding weight value is depicted in table 3.

Thus, the predicted emotion outcome is obtained by satisfying one of three conditions i.e., (i) when the emotion index value is less than 0.3 weight value it is predicted as engaged, (ii) emotion is considered "not engaged" when its engagement index falls within the range of 0.3 to 0.6 and (iii) if the weight value is greater than 0.6 it predicts as engaged. Therefore, the employed imln-bi-lstm model efficiently predicts student engagement in ol and its result is denoted as $P^{i,d}_{Linknet-Bilstm}$.

Table 3 Emotion and its weight values

Emotion	Weight value
Pass, surprised	0.6
Withdrawn, sad	0.3
Fail, fear	0.3
Distinction, happy	0.6

Table 4 Training and testing details of the multimodal dataset

Training data (%)	Number of training input	Number of testing input
60	369	246
70	430	185
80	492	123
90	553	62

4 results and discussion

4.1 Simulation procedure

Python was used in the simulation of the suggested students" engagement prediction; the version used was "Python 3.7." "11th Gen Intel(R) Core (TM) i5-1135G7 @ 2.40 GHz 2.42 GHz" was the processor used for simulation, and the system has "16.0 GB" of RAM installed. Additionally, analysis for predicting students" engagement was carried out on CKPLUS (Facial image) (https:,, www.kaggle.com, datasets, shawon10, ckplus. xxxx) and OULAD (data) (https:,, www.kaggle.com, datasets, anlgrbz, student-demographics-online-education-dataoulad. xxxx).

4.2 Dataset description

4.2.1 Facial image data description

This dataset aims to categorize each facial expression into one of 7 emotions: "anger, contempt, disgust, fear, joy, sadness, and surprise." It consists of 981 images used for this classification task. In this work, the anger, disgust and contempt labels are neglected because they occur rarely, so this study focuses only on happy, sad, fearful, and surprised emotions.

4.2.2 Cognition data description

The Open University Online Learning Platform (VLE) is the source of the dataset. Distance learners use this platform to do a variety of tasks, including reading course materials, taking part in forum discussions, turning in assignments, and checking their grades. Consisting of 7 carefully chosen courses, referred to as modules within the dataset, it distinguishes between presentations for semesters one and two. These are indicated by the letters "J" and "B" following the respective years. Moreover, the dataset encompasses demographic particulars about the students, encompassing their location, age group, educational level, gender, and any reported handicaps. This dataset consists of labels such as distinction, withdrawn, fail, and pass.

Therefore, this study incorporates these two benchmark datasets having facial image data and cognition data, which are combined as a multimodal dataset and these details are assumed to be obtained from a single student for student engagement prediction. The training and testing details of the multimodal dataset are shown in Table 4.

4.3 Performance analysis

A thorough investigation was conducted to assess the efficacy of both ImLN-Bi-LSTM and conventional strategies in predicting students" engagement. This extensive examination involved evaluating a broad range of critical metrics, including "F-measure, Precision, Matthews Correlation Coefficient (MCC), and Accuracy." Additionally, the evaluation included Statistical assessment to afford deeper insights into the models" performance. The ImLN-Bi-LSTM method"s effectiveness was further compared

against state-of-the-art techniques such as CNN (Ngai et al. 2022) and NN (Savchenko and Makarov 2022) along with the traditional classifiers like Efficient Net, Mobile Network, RNN, DenseNet, and LinkNet. Moreover, the ImLN-Bi-LSTM approach underwent comparison with IDBN+CNN (Maddu and Murugappan 2024). These comparisons were conducted using both the CKPLUS and OULAD datasets, ensuring a comprehensive assessment across different datasets to gauge the robustness and generalizability of the methods. Additionally, Fig. 5 depicts the sample images that have undergone pre-processing, which are utilized in predicting students" engagement.

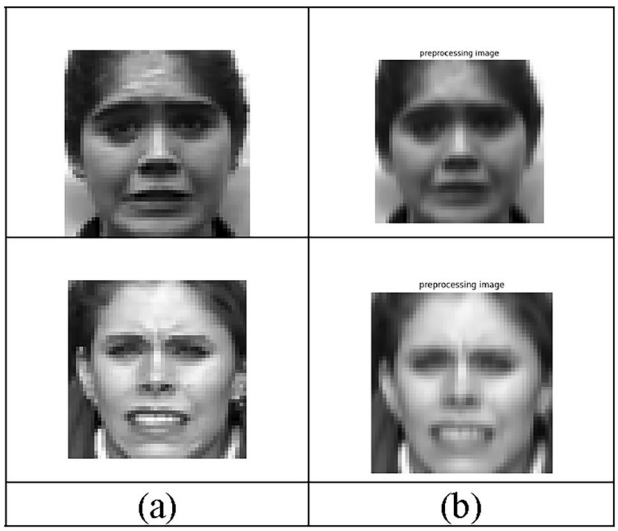

Fig. 5 Images for Students" Engagement Prediction using CKPLUS dataset **a** Sample Images and **b** Gaussian Filter-based Pre-processed images

4.4 Comparative evaluation of performance measures

The performance metric analysis of an ImLN-Bi-LSTM model is compared against established methods such as Efficient Net, NN (Savchenko and Makarov 2022), CNN (Ngai et al. 2022), XGBoost (Hossen and Uddin 2023) and FPN_CSA_Trans_EH [47], Mobile Network, RNN, DenseNet, and LinkNet, is shown in Fig. 6. The accuracy outcome of each model over various rates of the training data is shown in Fig. 6a. The accuracy measure provides a quick and broad understanding of the model"s performance by showing the proportion of correct predictions. The ImLN-Bi-LSTM model consistently exhibits higher accuracy than traditional methods. Specifically, with 60% of the training data, the ImLN-Bi-LSTM model achieves an accuracy of 0.892, surpassing other models such as EfficientNet, NN (Savchenko and Makarov 2022), CNN (Ngai et al. 2022), Mobile Network, RNN, DenseNet, XGBoost (Hossen and Uddin 2023), FPN_CSA_Trans_EH (Naveen et al. 2025), and LinkNet, which achieve accuracies around 0.732, 0.614, 0.507, 0.558, 0.627, 0.653, and 0.629 respectively. Similarly, the ImLN-Bi-LSTM model continues to lead at 90% of training data, outperforming other models with an F-measure of 0.951. This higher performance suggests that the model is more effective at capturing the complexities and patterns in student behavior that correlate with engagement.

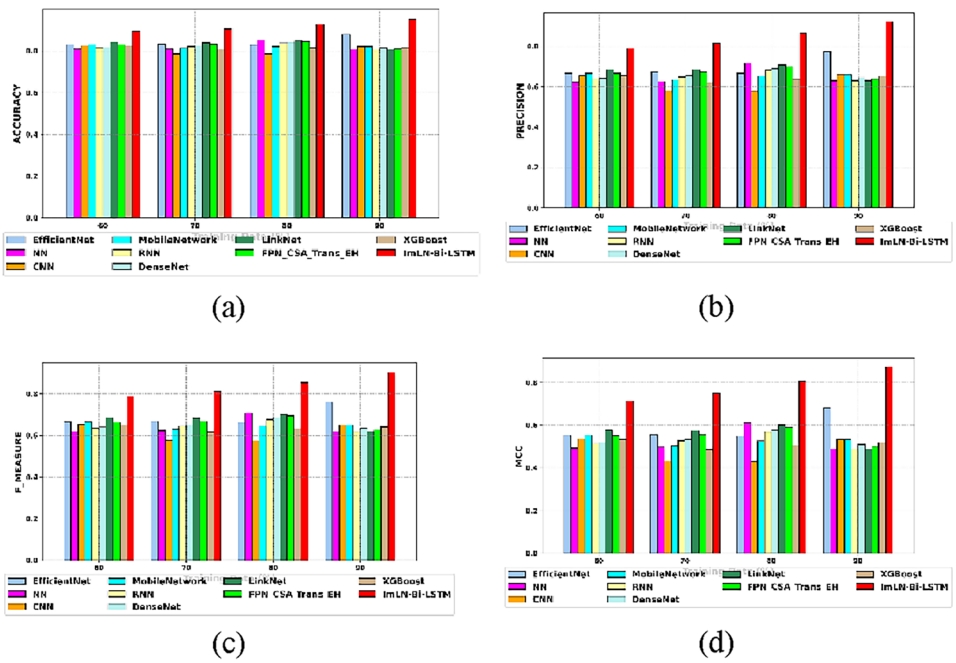

Fig. 6 Comparison of Performance Metric Analysis on ImLN-Bi-LSTM Method vs. Conventional Techniques

Table 5 Analysis of statistical test

ImLN-Bi-LSTM Vs	T-test	Wilcoxon *p*-value	Friedman *p*-value
EfficientNet	0.078	0.091	0.072
NN	0.099	0.091	0.095
CNN	0.094	0.082	0.093
MobileNetwork	0.095	0.077	0.076
RNN	0.080	0.076	0.077
DenseNet	0.098	0.089	0.075
LinkNet	0.077	0.097	0.077
FPN_CSA_Trans_EH	0.086	0.086	0.089
XGBoost	0.089	0.096	0.077

Table 6 K-fold validation of the ImLN-Bi-LSTM and the existing techniques

Methods	K = 2 (%)	K = 3 (%)	K = 4 (%)	K = 5 (%)
EfficientNet	86.70	86.71	86.93	82.84
NN	85.27	82.21	83.18	82.16
CNN	82.84	83.65	85.98	82.98
MobileNetwrok	82.39	85.09	85.46	83.65
RNN	86.52	82.05	83.67	85.33
DenseNet	83.93	83.23	86.36	82.27
Linknet	86.52	82.63	84.00	85.69
FPN_CSA_Trans_EH	86.10	83.09	84.52	85.77
XGBoost	85.23	83.67	83.50	84.12
ImLN-Bi-LSTM	93.72	91.19	92.98	91.65

4.5 Analysis of statistical test

Table 5 presents the results of statistical test analysis of the proposed ImLN-Bi-LSTM model against existing models using three tests: the T-test, Wilcoxon signed-rank test, and Friedman test, with a significance threshold set at 0.1. A *p*-value below this threshold indicates that the performance difference between models is statistically significant. When compared to MobileNetwork, the *p*-values are 0.095 (T-test), 0.077 (Wilcoxon), and 0.076 (Friedman), all below the threshold, confirming statistical significance. Similarly, comparisons with models such as EfficientNet, RNN, and XGBoost also yield *p*-values below 0.1 in the T-test, Wilcoxon signed-rank test, and Friedman test, supporting the robustness of the proposed model. These results collectively indicate that the performance improvements observed with ImLN-Bi-LSTM are not due to random variation but are statistically meaningful, affirming the model"s effectiveness.

4.6 K-fold validation

Table 6 shows the k-fold validation of the recommended ImLN-Bi-LSTM model and the existing frameworks like Efficient Net, NN, CNN, Mobile Network, RNN, DenseNet, XGBoost and FPN_CSA_Trans_EH, and LinkNet. As a result, when k = 2, the suggested ImLN-Bi-LSTM technique has attained an accuracy of 93.72%. The value surpasses the outcomes of the existing techniques like Efficient Net (86.70%), NN (85.27%), CNN (82.84%), Mobile Network (82.39%), RNN (86.52%), DenseNet (83.93%), XGBoost (85.23%), FPN_CSA_Trans_EH (86.10%) and LinkNet (86.52%). Also, the proposed ImLN-Bi-LSTM has attained superior performance on accuracy across different k-values. Therefore, the proposed ImLN-Bi-LSTM exhibits outstanding performance on generalization compared to other existing techniques in predicting student engagement.

Table 7 Analysis of computational time

Methods	Time(s)
EfficientNet	88.2
NN	75.2
CNN	66.2
MobileNetwork	82.01
RNN	80.62
DenseNet	71.86
LinkNet	71.5
FPN_CSA_Trans_EH	96.21
XGBoost	89.21
ImLN-Bi-LSTM	55.21

Table 8 Assessment of the ImLN-Bi-LSTM model and the conventional methods for space complexity

Methods	Space complexity (KB)
EfficientNet	0.89
NN	0.45
CNN	0.45
MobileNetwork	0.21
RNN	0.36
DenseNet	0.21
LinkNet	0.33
FPN_CSA_Trans_EH	0.25
XGBoost	0.22
ImLN-Bi-LSTM	0.09

4.7 Analysis of computational time

Table 7 presents the computational time (in seconds) for various models, highlighting the ImLN-Bi-LSTM as the most efficient with a processing time of 55.21 s, outperforming other models. In comparison, models like CNN and DenseNet require 66.2 s and 71.86 s, respectively, while EfficientNet and XGBoost take 88.2 s and 89.21 s, respectively. On the other hand, the FPN_CSA_Trans_EH model requires the most time at 96.21 s. Moreover, MobileNetwork and RNN show moderate processing times of 82.01 s and 80.62 s, respectively, while LinkNet performs with a time of 71.5 s. Overall, the ImLN-Bi-LSTM is a good option for real-time applications in dynamic online educational settings because it not only performs competitively but also excels in computational efficiency.

4.8 Analysis of space complexity

Space complexity analysis refers to the amount of memory required by a model to complete a task. The analysis of space complexity across various models used for student engagement prediction is shown in Table 8. As shown in the evaluation, the ImLN-Bi-LSTM exhibits the lowest space complexity of 0.09 KB. In contrast, traditional models such as EfficientNet (0.89 KB), NN and CNN (each at 0.45 KB), DenseNet (0.21 KB), MobileNetwork (0.21 KB), RNN (0.36 KB), LinkNet (0.33 KB), FPN_CSA_Trans_EH (0.25 KB), and XGBoost (0.22 KB) also demonstrate higher space requirements than ImLN-Bi-LSTM. This substantial reduction in memory usage not only emphasizes the model"s compactness but also supports its suitability for deployment in real-time educational environments, without compromising predictive performance.

4.9 Performance comparison of ImLN-Bi-LSTM model in terms of pose, expression and color

Table 9 displays the results of an evaluation of the suggested ImLN-Bi-LSTM model"s performance using three distinct features: pose, expression, and color. Across all evaluation metrics, the color feature achieves the highest accuracy (0.912), sensitivity (0.881), and specificity (0.884), indicating its strong capability in correctly identifying both engaged and non-engaged states. The proposed model in terms of color feature achieves highest F-measure at 0.884, reflecting a balanced performance in terms of precision and recall. Although expression features show the highest precision (0.875) and NPV (0.871), their overall performance is slightly lower than that of color. The pose feature performs comparably well, with the highest MCC (0.863). Overall, the analysis confirms that the ImLN-Bi-LSTM model performs effectively across all features, with color features offering the most reliable and accurate results.

4.9.1 Analysis of ROC

Figure 7 illustrates the ROC curve analysis for the proposed ImLN-Bi-LSTM model, which is designed to predict student engagement, compared against several existing models including EfficientNet, NN, CNN, MobileNet, RNN, DenseNet, LinkNet, FPN_CSA_Trans_EH, and XGBoost. The ROC curve is constructed by plotting the average TPR against the FPR across various threshold levels. For effective student engagement prediction, the model should achieve an area under the curve of 0.95 or higher. In particular, when the TPR was set to 0.92, the proposed ensemble model recorded an FPR of 0.014 and achieved an AUC of 0.95, indicating high predictive performance. In comparison, the existing models demonstrated significantly lower AUC values, underscoring the superior performance of the ImLN-Bi-LSTM approach in predicting student engagement.

Table 9 Performance evaluation of ImLN-Bi-LSTM using pose, expression and color

Metrics	Pose (%)	Expression (%)	Color (%)
Accuracy	90.90	90.20	91.20
Sensitivity	85.10	86.20	88.10
Specificity	87.20	86.10	88.40
Precision	86.70	87.50	87.00
F-measure	86.70	87.10	88.40
MCC	86.30	85.50	85.30
NPV	85.80	87.10	87.00
FPR	12.80	13.90	11.60
FNR	14.90	13.80	11.90

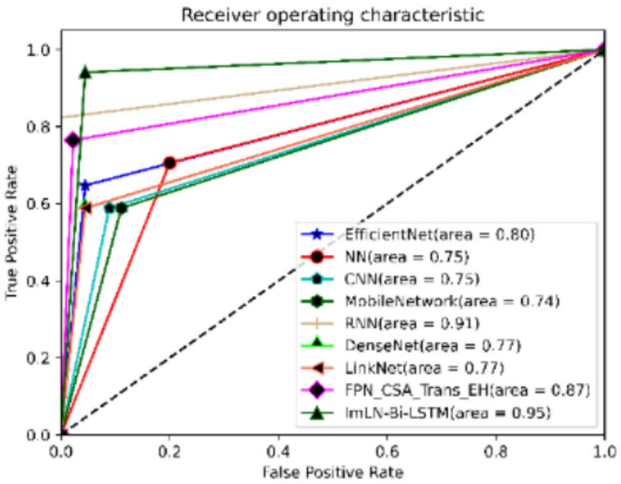

Fig. 7 ROC analysis on ImLN-Bi-LSTM Method and Conventional Methods

4.9.2 Feature analysis using SHAP

An Explainable AI (XAI) technique called Shapley Additive exPlanations (SHAP) uses game theory to understand model predictions by considering each characteristic as a player influencing the final result. It provides both local and global explanations, making it suitable for understanding model behavior in student engagement prediction. The SHAP summary plots shown in Fig. 8 highlight the most influential features. In the beeswarm plot (Fig. 8a), Feature 202 has the greatest impact on predictions, with higher values of this feature associated with increased SHAP values, indicating a stronger contribution to predicting higher engagement. The average impact of the features on the model"s output is displayed in Fig. 8b as a bar plot. Feature 35 has the highest overall significance, followed by Features 32, 143, and others. These observations highlight the

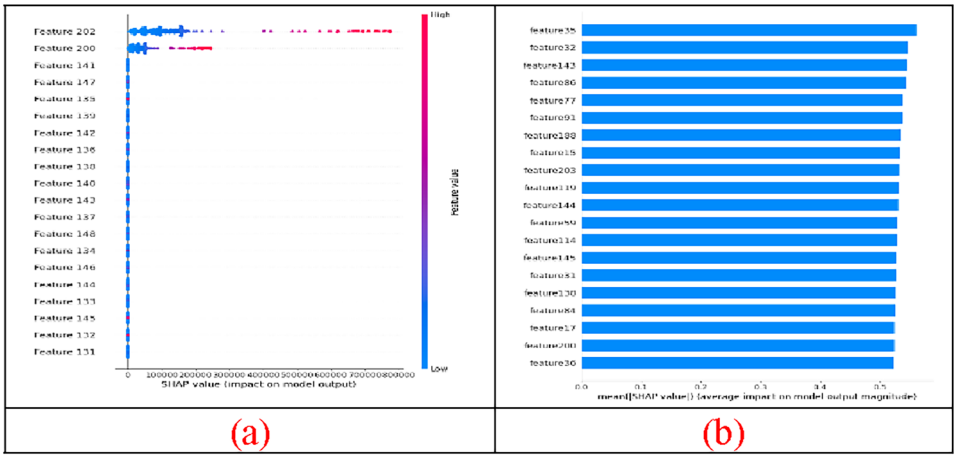

Fig. 8 Feature analysis using SHAP values **a** Beeswarm plot and **b** bar plot

significance of attributes for accurately predicting students" involvement in OL.

5 Practical implication

The proposed ImLN-Bi-LSTM model demonstrates outstanding prediction performance for student engagement in online learning by leveraging facial images and related data. This makes it possible to identify at-risk students in a timely manner, enabling teachers to properly intervene and enhance academic results. However, the lack of comprehensive investigation into data privacy (ie) the preservation of facial data over time introduces risks of data breaches and limited generalization across diverse student populations representing key limitations of the proposed study. Future work will address these issues by implementing privacy-preserving techniques and integrating with real-time online educational platforms to enhance practical deployment in further studies.

6 Conclusion

To sum up, this study involved a number of steps, including feature extraction, classification, pre-processing, and student involvement prediction. Preprocessing and extraction of features were done separately on the given input, which consisted of images and data with facial emotions on them. Within pre-processing, techniques like Gaussian Filtering and Minmax normalization were employed for image and data processing respectively. Feature extraction involved capturing relevant features from face expression images such as SLBT and LGXP-based features, along with statistical features and I-EF from data. These features were then subjected to a hybrid classification model for student engagement classification. The hybrid system, an integration of Improved LinkNet and Bi-LSTM techniques processed these features distinguishing itself from traditional classifiers. Additionally, engagement prediction was integrated into this ImLN-Bi-LSTM model ensuring precise outcomes. The efficacy of the suggested approach was confirmed by means of multiple experimental evaluations. Notably, the ImLN-Bi-LSTM model obtained a maximum accuracy of 0.952, meanwhile, EfficientNet, NN, CNN, RNN, Mobile Network, DenseNet, and LinkNet achieved minimal accuracies around 0.832, 0.784, 0.813, 0.813, 0.802, 0.809, and 0.801 respectively, suggesting potential shortcomings in their ability to accurately identify engaged students. In future work, this study will incorporate real-time data from a diverse group of students to evaluate the performance and robustness of the model.

Acknowledgements I am truly grateful to the co-authors of this work for their insightful and helpful recommendations throughout the conception and development of this study.

Author contributions Rama Bhadra Rao Maddu conceived the presented idea and designed the analysis. Also, he carried out the experiment and wrote the manuscript with support from Dr.S Murugappan. All authors discussed the results and contributed to the final manuscript. All authors read and approved the final manuscript.

Funding This research did not receive any specific funding.

Data availability The input underlying this article is available at https://www.kaggle.com/datasets/shawon10/ckplus and https://www.kaggle.com/datasets/anlgrbz/student-demographics-online-education-dataoulad.

Declarations

Conflict of interest Authors declared that they have no conflict of interest.

Clinical trial number Not applicable.

Ethical approval Not applicable.

Human Ethics and Consent to Participate declarations. Not applicable.

Informed consent Not applicable.

References

Abdulkader R, Ayasrah FT, Nallagattla VR, Hiran KK, Dadheech P, Balasubramaniam V, Sengan S (2023) Optimizing student engagement in edge-based online learning with advanced analytics. Array 19:100301

Abu MA, Rosleesham S, Suboh MZ, Yid MS, Kornain Z, Jamaluddin NF (2020) Classification of EMG signal for multiple hand gestures based on neural network. Indonesian J Electr Eng Comput Sci 17(1):256–263

Al Mamun MA, Lawrie G (2023) Student-content interactions: exploring behavioural engagement with self-regulated inquiry-based online learning modules. Smart Learn Environ 10(1):1

Aydoğdu Ş (2020) Predicting student final performance using artificial neural networks in online learning environments. Educ Inf Technol 25(3):1913–1927

Ayouni S, Hajjej F, Maddeh M, Al-Otaibi S (2021) A new ML-based approach to enhance student engagement in online environment. PLoS ONE 16(11):e0258788

Bavkar DM, Kashyap R, Khairnar V (2022) Multimodal sarcasm detection via hybrid classifier with optimistic logic. J Telecommun Inf Technol 30(3):97–114

Buono P, De Carolis B, D"Errico F, Macchiarulo N, Palestra G (2023) Assessing student engagement from facial behavior in on-line learning. Multimed Tools Appl 82(9):12859–12877

Deo RC, Yaseen ZM, Al-Ansari N, Nguyen-Huy T, Langlands TA, Galligan L (2020) Modern artificial intelligence model development for undergraduate student performance prediction: an investigation on engineering mathematics courses. IEEE Access 8:136697–136724

Figueroa-Cañas J, Sancho-Vinuesa T (2020) Early prediction of dropout and final exam performance in an online statistics course. IEEE Revista Iberoamericana De Tecnologias Del Aprendizaje 15(2):86–94

Flanagan B, Majumdar R, Ogata H (2022) Early-warning prediction of student performance and engagement in open book assessment by reading behavior analysis. Int J Educ Technol High Educ 19(1):41

Gupta S, Kumar P, Tekchandani RK (2023) Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. Multimed Tools Appl 82(8):11365–11394

Hamayel MJ, Owda AY (2021) A novel cryptocurrency price prediction model using GRU, LSTM and bi-LSTM machine learning algorithms. AI 2(4):477–496

Hamdi S, Oussalah M, Moussaoui A, Saidi M (2022) Attention-based hybrid CNN-LSTM and spectral data augmentation for COVID-19 diagnosis from cough sound. J Intell Inf Syst 59(2):367–389

Henderi H, Wahyuningsih T, Rahwanto E (2021) Comparison of min-max normalization and z-score normalization in the K-nearest neighbor (kNN) algorithm to test the accuracy of types of breast cancer. Int J Inf Inf Syst 4(1):13–20

Hossen MK, Uddin MS (2023) Attention monitoring of students during online classes using XGBoost classifier. Comput Educ Artif Intell 5:100191

https://www.kaggle.com/datasets/anlgrbz/student-demographics-online-education-dataoulad

https://www.kaggle.com/datasets/shawon10/ckplus

Kermani M, Plett E. Modified entropy correction formula for the Roe scheme. In39th Aerospace Sciences Meeting and Exhibit 2001; 83

Lakshmiprabha NS, Majumder S. Face recognition system invariant to plastic surgery. In2012 12th International conference on intelligent systems design and applications (ISDA) 2012;258–263

Liao J, Liang Y, Pan J (2021) Deep facial spatiotemporal network for engagement prediction in online learning. Appl Intell 51(10):6609–6621

Maddu RB, Murugappan S (2024) Online learners" engagement detection via facial emotion recognition in online learning context using hybrid classification model. Soc Netw Anal Min 14(1):43

Miller AL, Fassett KT, Palmer DL (2021) Achievement goal orientation: a predictor of student engagement in higher education. Motiv Emot 45(3):327–344

Munagala V, Kodati SP (2021) Enhanced holoentropy-based encoding via whale optimization for highly efficient video coding. Vis Comput 37(8):2173–2194

Naveen A, Jacob IJ, Mandava AK (2025) Detection of student engagement via transformer-enhanced feature pyramid networks on channel-spatial attention. Информатика и Автоматизация 24(2):631–656

Ngai WK, Xie H, Zou D, Chou KL (2022) Emotion recognition based on convolutional neural networks and heterogeneous bio-signal data sources. Inf Fusion 77:107–117

Oladipupo O, Samuel S (2024) A learning analytic approach to modelling student-staff interaction from students" perception of engagement practices. IEEE Access 12:10315–10333

Ouyang F, Wu M, Zheng L, Zhang L, Jiao P (2023) Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. Int J Educ Technol High Educ 20(1):4

Ramasamy G, Singh T, Yuan X (2023) Multi-modal semantic segmentation model using encoder based Link-Net architecture for BraTS 2020 challenge. Procedia Comput Sci 218:732–740

Ruiz N, Yu H, Allessio DA, Jalal M, Joshi A, Murray T, Magee JJ, Delgado KM, Ablavsky V, Sclaroff S, Arroyo I (2022) ATL-BP: a student engagement dataset and model for affect transfer learning for behavior prediction. IEEE Trans Biomet Behavior Identity Sci 5(3):411–424

Sashank YT, Kakulapati V, Bhutada S (2023) Student engagement prediction in online session. Int J Recent Innov Trends Comput Commun 11(2):43–47

Savchenko AV, Makarov IA (2022) Neural network model for video-based analysis of student's emotions in e-learning. Opt Memory Neural Netw 31(3):237–244

Sekehravani EA, Babulak E, Masoodi M (2020) Implementing canny edge detection algorithm for noisy image. Bull Electr Eng Inf 9(4):1404–1410

Selim T, Elkabani I, Abdou MA (2022) Students engagement level detection in online e-learning using hybrid efficientnetb7 together with tcn, lstm, and bi-lstm. IEEE Access 10:99573–99583

Shanthi A, Koppu S (2023) Remora namib beetle optimization enabled deep learning for severity of COVID-19 lung infection identification and classification using CT images. Sensors 23(11):5316

Sobnath D, Kaduk T, Rehman IU, Isiaq O (2020) Feature selection for UK disabled students" engagement post higher education: a machine learning approach for a predictive employment model. IEEE Access 8:159530–159541

Song X, Li J, Sun S, Yin H, Dawson P, Doss RR (2020) SEPN: a sequential engagement based academic performance prediction model. IEEE Intell Syst 36(1):46–53

Thomas C, Sarma KP, Gajula SS, Jayagopi DB (2022) Automatic prediction of presentation style and student engagement from videos. Comput Educ Artific Intell 3:100079

Toptaş B, Hanbay D (2021) Retinal blood vessel segmentation using pixel-based feature vector. Biomed Signal Process Control 70:103053

Wan H, Liu K, Yu Q, Gao X (2019) Pedagogical intervention practices: improving learning engagement based on early prediction. IEEE Trans Learn Technol 12(2):278–289

Wang X, Guo B, Shen Y (2022) Predicting the at-risk online students based on the click data distribution characteristics. Sci Program 2022(1):9938260

Xue H, Niu Y (2023) Multi-output based hybrid integrated models for student performance prediction. Appl Sci 13(9):5384

Yan H, Deng Y (2020) An improved belief entropy in evidence theory. IEEE Access 8:57505–57516

Yue J, Tian F, Chao KM, Shah N, Li L, Chen Y, Zheng Q (2019) Recognizing multidimensional engagement of e-learners based on multi-channel data in e-learning environment. IEEE Access 7:149554–149567