International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V73I3P133

Volume 73 Issue 3, 473-486, March 2025
© 2025 Seventh Sense Research Group®

Original Article

EDUANOM: Deep Learning Approach for Anomaly Detection in the Classroom Environment

Dhatri Pandya[1], Keyur Rana[2]

[1]*Computer Engineering Department, Gujarat Technological University, Gujarat, India.*
[2]*Computer Engineering Department, Sarvajanik College of Engineering and Technology, Gujarat, India.*

[1]*Corresponding Author : dhatri.pandya@scet.ac.in*

Received: 25 November 2024 Revised: 17 February 2025 Accepted: 27 February 2025 Published: 28 March 2025

Abstract *- Anomaly detection is a prominent area of research in the field of deep learning. With the substantial increase in the number of CCTV located at various places, monitoring abnormal behavior is required to ensure safety and prevent the violation of rules at particular places. Anomaly detection in the classroom environment is important to ensure the safety of the students. Early detection of abnormal behavior will help educational institutions to take preventive measures against the students. Deep learning requires a huge amount of labeled data to achieve good accuracy. The data obtained from CCTV footage is currently unannotated, requiring manual annotation to facilitate the application of supervised deep learning architecture application. In addition, there is urge need for deep learning architecture trained on real-time datasets for anomaly detection in the classroom environment. To address this, we proposed a methodology based on unsupervised deep learning architecture by proposing a custom convolutional auto-encoder for anomaly detection in the classroom environment. Anomaly detection is accomplished through the utilization of a convolutional encoder, which evaluates the reconstruction loss between testing samples and training samples. This approach is effective for accurately identifying anomalies within the dataset. To implement this, similarity measures such as Mean Square Error (MSE), Kernel Density Estimation (KDE), and Structured Similarity Index Measure (SSIM) are applied. We evaluate the trained model on real-time data collected from the classroom environment with a comparative analysis of different similarity metrics. In this research work, we achieved 98% accuracy for anomaly detection in the classroom environment using the proposed methodology with similarity metrics SSIM. This research helps identify the role of unsupervised deep learning architecture and various similarity measures to identify anomalies in the classroom environment.*

Keywords *- Anomaly detection, Auto encoders, Education, Structured Similarity Index Measure, Deep Learning.*

1. Introduction

Abnormal event detection [1] has become an important area of research in machine vision and pattern recognition in recent years. The primary challenge lies in the fact that the scenarios depicting anomalous happenings are varied. Defining an interface that encompasses the limits of numerous potential abnormal occurrences is a challenging task. One approach is to define an abnormal event as an occurrence with a low probability compared to a normal event. This allows for statistical analysis of abnormal events that depart from expectations. Abnormal events are those that are not consistent with normal samples. Basically, two approaches are considered for abnormal event detection:

- Supervised Learning: It requires a large amount of labeled data for abnormal samples, which is time-consuming and expensive.
- Unsupervised learning: It requires a huge collection of normal samples. Testing samples that deviate from this normal are considered abnormal samples.

Unsupervised Deep Learning Based Anomaly Detection is critically important in foundational machine learning research and practical industrial applications [2]. Various deep learning frameworks have been developed to address the challenges in unsupervised anomaly detection. Autoencoders, a key category of unsupervised deep learning architectures, are employed in anomaly detection, as noted in references [3, 4]. The proposed deep unsupervised models for anomaly identification are based on specific assumptions that are highly effective in recognizing outliers [5]. It is vital to accurately distinguish the 'normal' and 'anomalous' regions within either the original or the latent feature space. The bulk of the data instances in the data set are classified as normal, in contrast to the other occurrences. The unsupervised anomaly detection technique assesses each data instance by determining an outlier score based on the underlying characteristics of the dataset, including distances and densities. This approach facilitates the identification of anomalies within the data [6]. The benefits of unsupervised deep anomaly detection techniques are as follows [7].

- Labelled data is not required for training the model.
- Reduces data dimensions providing compressed representation of data.
- Learning data representation of normal samples helps to detect outliers.

Anomaly Detection in video generally consists of two phases [8].

- Event representation: It entails gathering data from the video scene and extracting various features using an appropriate model depending on the type of anomaly.
- Identification and detection of anomalies: Anomaly detection often involves using predefined rules that define the characteristics of a normal occurrence.

Features that deviate from the model's learned pattern are considered anomalies or abnormal events. Auto encoders are state-of-the-art unsupervised deep learning architecture. Several variants of autoencoder are available as per [1]. One of the variants is the Convolutional Auto Encoder (CAE) [9]. The key component of the Convolutional Auto Encoder (CAE) [9] is the encoder and decoder, which consists of convolution layers as opposed to fully interconnected layers. The encoder consists of convolution layers, ReLU and max pooling layers to create a condensed representation from input images, while the decoder uses deconvolution layers and a ReLU layer for image reconstruction. CAE is effective in capturing spatial features from images for anomaly detection. Using auto encoders for anomaly detection follows the assumption that a trained autoencoder would learn the latent subspace of standard samples. Once trained, it would result in a low reconstruction error for standard samples and a high reconstruction error for anomalies.

Despite significant advancements in Artificial Intelligence (AI) and computer vision, the application of Abnormal Event Detection (AED) in educational environments remains substantially underexplored. Addressing the existing research gaps in this area is essential for ensuring the safety and well-being of students. i) Limited Contextual Understanding Current AED models predominantly utilize generic anomaly detection techniques, often neglecting the unique behavioral and environmental contexts found in the educational environment. The lack of contextualized datasets severely limits our ability to differentiate between normal variations in student behavior and genuinely abnormal events, such as violent outbursts or unauthorized access. ii) Scarcity of Robust Dataset: There is a lack of publicly available, large-scale datasets specifically tailored to abnormal behavior in educational contexts. Most existing datasets are focused on general surveillance applications and do not adequately address scenarios unique to educational environments, including student gathering in the classroom, bullying and classroom disruptions. The methodology is proposed based on a convolutional autoencoder to handle this type of anomaly detection. This paper is structured as follows. Section 2 discusses the recent anomaly detection approaches using deep learning architectures and motivation, along with major contributions to the paper. Section 3 consists of the detailed architecture of the proposed methodology along with the anomaly detection process. In Section 4, we discussed experimental results stating the effectiveness of the proposed model and analysis with different scenarios of anomaly detection using various similarity metrics. Section 5 summarizes the main findings, contributions, limitations, and future scope.

2. Related Work

Numerous approaches have been developed for anomaly detection as per [10-12]. Hasan et al. [13] introduced a novel framework for convolutional auto-encoder to reconstruct complex scenes. This framework involved using reconstruction costs to effectively identify anomalies within the reconstructed scenes. Similarly, Zhou et al. [14] proposed the utilization of spatio-temporal Convolutional Neural Networks (CNNs) to simultaneously capture and analyze joint appearance and motion characteristics comprehensively. Furthermore, Sultani et al. [15] presented an advanced approach by integrating deep neural networks with multiple instances learning techniques to accurately classify real-world anomalies such as accidents, explosions, fights, abuse, and arson. This fusion enabled a more comprehensive and nuanced understanding of various anomaly types, enhancing classification accuracy and real-world applicability. In [17], the method for detecting anomalies uses two approaches to understand pattern appearances and their motions. The first approach uses an auto-encoder architecture to reconstruct the appearance, while the second approach employs a U-Net structure to predict immediate motion from a video frame. Additionally, a patch-based method is used to estimate anomaly scores, reducing the impact of model output noise. The VidAnomalyNet [18] deep learning architecture is designed to identify anomalies in the video using novel CNN.

Furthermore, [18] proposed a deep learning architecture based on separable 2D convolution to reduce computation. Transfer learning is applied to benchmark datasets for anomaly detection. A deep learning approach using a hybrid architecture consisting of the convolutional autoencoder and sequence-to-sequence long short-term memory auto encoder has been proposed in [19] to monitor surveillance videos continuously and detect anomalies. The proposed unsupervised learning method leverages a one-class classification framework to identify video anomalies effectively. This approach aims to enhance the accuracy and reliability of anomaly detection in various video contexts. The model's effectiveness has been demonstrated on benchmarked anomaly detection datasets, achieving significant results regarding equal error rate, area under the curve, and time required for detection [19]. In [20], anomaly detection based on the auto-encoder and deep Generative adversarial network has been proposed for the images. In addition, finding the optimal threshold for anomaly detection is also discussed. The taxonomy of autoencoders plays an important role, along with its applications in various fields, as per the literature [21]. However, the limitations of the auto-encoder variants are also discussed in [21]. Several hyperparameters contribute to the accuracy of the autoencoder in detecting anomalies. One of the important hyperparameters is the loss function, which is used to measure the reconstruction error [21].

2.1. Motivation

Implementing anomaly detection systems in classrooms and on campus is driven by the critical need to prevent violence, such as grouping students or fights, before they escalate. In educational settings, where large groups of students congregate, manually monitoring and assessing crowd behaviour can be challenging. Anomaly detection technologies can automatically identify irregular patterns, such as sudden clustering of students or heightened agitation, which may signal the onset of violence. By detecting these anomalies early, school authorities can promptly intervene, thereby preventing potential harm and maintaining a safe and conducive environment for learning. This proactive approach ensures students' physical well-being and fosters a sense of security and trust within the school community.

In existing literature discussed in related work, anomaly detection is carried out using various approaches. However, the parameter computation is very high, and the data set used is in a constraint environment. There is a need for custom lightweight, unsupervised deep learning architecture for anomaly detection in the classroom environment. The proposed research aims to leverage deep learning techniques, particularly convolutional auto encoders, to uncover anomalies or outliers in data. The focus of this study is to identify unexpected patterns within the context of students in a classroom environment. Unsupervised deep learning architecture has the potential to be valuable in identifying pupils who may be disinterested, absent, or exhibiting abnormal behavior, allowing for prompt interventions to enhance the educational setting.

2.2. Contribution

- Collection of real-time data samples consisting of normal and abnormal behavior of the students in the classroom.
- Defining a custom lightweight convolutional auto encoder for identifying anomalies in the classroom.
- Analysis of anomaly detection using similarity metrics viz. mean square error, kernel density estimation, structured similarity index measure.

3. Proposed Methodology

We propose using a custom unsupervised deep learning architecture to detect anomalies in the classroom environment. We use convolutional auto encoders, a type of artificial neural network, to recognize irregularities. These auto-encoders condense input data and then reconstruct it. Exposing the autoencoders to regular data, they learn to identify typical patterns and characteristics. The primary objective of minimizing reconstruction error is to reduce the disparity between the original data and its reconstructed representation. When the model is used on new data, any substantial divergence from the learned patterns (referred to as a high reconstruction error) signals an abnormality. Convolutional autoencoders are particularly valuable for identifying unusual occurrences or anomalies in different contexts, such as identifying fraudulent activities, safeguarding network security, and detecting system faults when abnormal instances significantly deviate from regular patterns. In addition, the proposed methodology emphasizes lightweight deep learning architecture. Deployment of trained deep learning architecture to alert anomaly detection on resource constraint devices requires a trained model with less parameter computation.

In an educational environment, an anomaly is considered when the students sit in a crowd / or stand in a crowd during classroom activities. This behavior is considered abnormal during the teaching hours in the classroom. In the proposed methodology, we have collected CCTV footage of the classroom environment of various scenarios, as listed in Table 1.

The dataset is stated as EDUANOM. The dataset samples used for training the model are stated in Figure 1. Unsupervised deep learning architecture is emphasized in the research work. Due to this, the dataset samples consist of students studying in the classroom environment, such as normal behavior. The dataset samples are collected from various classroom environments to cover various viewpoints. The diverse dataset is collected by considering different students, different seating arrangements, and empty classrooms during holidays or when there are no scheduled classes. In addition, data samples are recorded during different time zones, such as sunlight in the classroom and normal daylight conditions.

Table 1. Details of the dataset

Data Description	Details
Category of videos considered for normal behavior	Proper sitting in the classroom with different viewpoints of the classroom
The total length of the video	36 minutes of total video
Length of each video	7-11 seconds
Number of students appearing in the video	5-40
Data Source	CCTV footage of the classroom
Number of frames considered for the training	445
Resolution	460 x700

3.1. Steps of the Proposed Methodology

The key components of the proposed methodology are the encoder and decoder, along with the reconstruction loss measure. The encoder part consists of two convolutional layers with varying 3x3 filter sizes and varying numbers of filters. Each convolutional layer is followed by a pooling layer that is responsible for reducing the dimension of the feature map. The encoder plays a role in creating the latent representation of all the training samples, as stated in Figure 1. The latent representation, also known as a latent space or latent vector, is a fundamental block of auto encoders. In an auto encoder's encoder-decoder architecture, the network's encoder component condenses the input data into a condensed form called the latent representation. This representation effectively captures the incoming data's fundamental characteristics and underlying structures by reducing its dimensionality. The aim is to retain the most pertinent information while discarding unnecessary details, allowing for tasks such as data compression, dimensionality reduction, and anomaly identification. The condensed representation serves as a succinct summary of the input data, enabling the decoder to reconstruct the original data from this latent representation. The architecture of the custom convolutional autoencoder is provided in Figure 2. The hyperparameter details are stated in Table 2 concise format. The feature benefits unsupervised learning. The decoder is an essential component in the design of an auto encoder since it is responsible for reconstructing the input data using the latent representation produced by the encoder. To accomplish this, the decoder frequently utilizes up-sampling layers. Upsampling layers, also known as "deconvolutional" layers [22], enlarge the spatial dimensions of the input data by employing techniques such as nearest-neighbour interpolation or bilinear interpolation [22], as stated in Figure 2. This methodology resembles the inverse of down-sampling performed by the pooling layers within the encoder, allowing the model to efficiently and progressively reconstruct feature maps to align with the dimensions of the original input size. The upsampling method enables the decoder to improve and augment the smaller elements, rebuilding the output with higher resolution based on the lower-dimensional latent representation. Through attentive layer design, the decoder can precisely reconstruct the input data, guaranteeing that the auto-encoder successfully acquires and expresses the vital features of the data in the latent space.

Table 2. Hyper parameters of the proposed methodology

Hyper Parameters	**Value**
Epochs	40
Batch size	20
Optimiser	Rmsprop
Filter size in Convolution Layer	3 x 3
Filter Size in Max Pooling Layer	2 x 2
Stride in the Convolution layer	1
Stride in Max Pooling Layer	2
Filter size in the upsampling layer	2 x 2
Activation Function	Sigmoid, ReLU

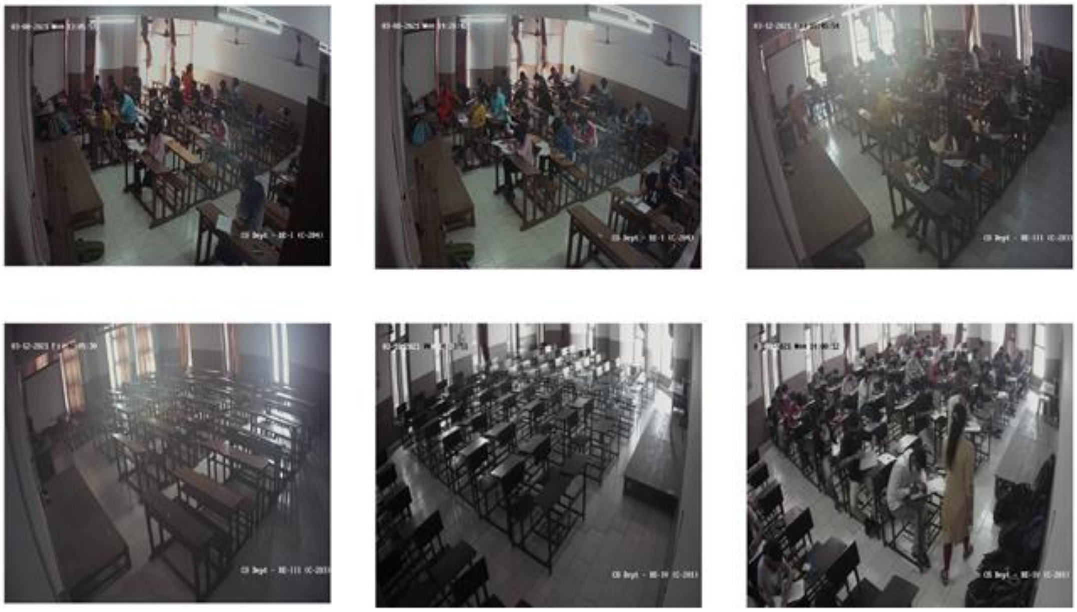


Fig. 1 Samples of the EDUANOM dataset

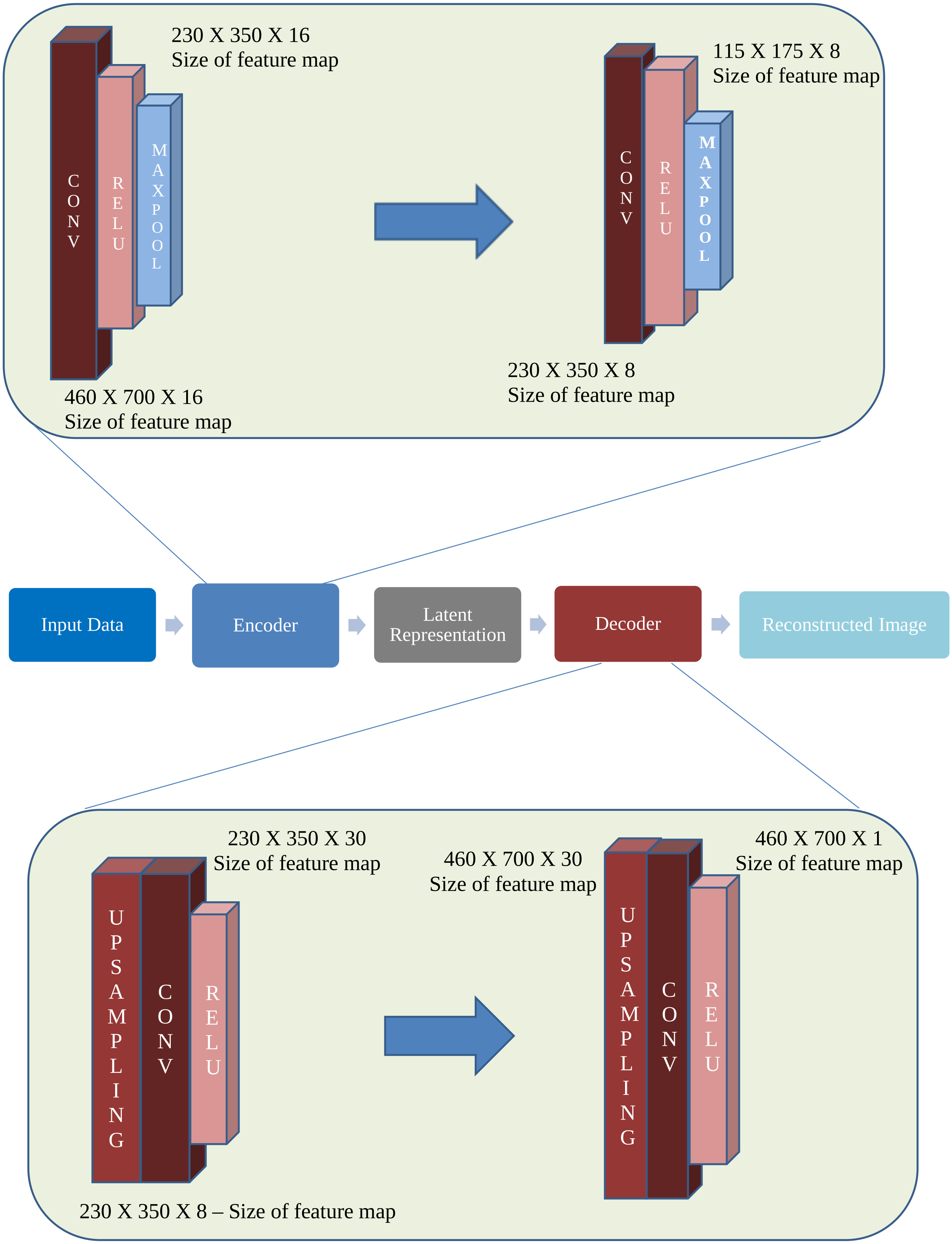


Fig. 2 Architecture of the proposed methodology

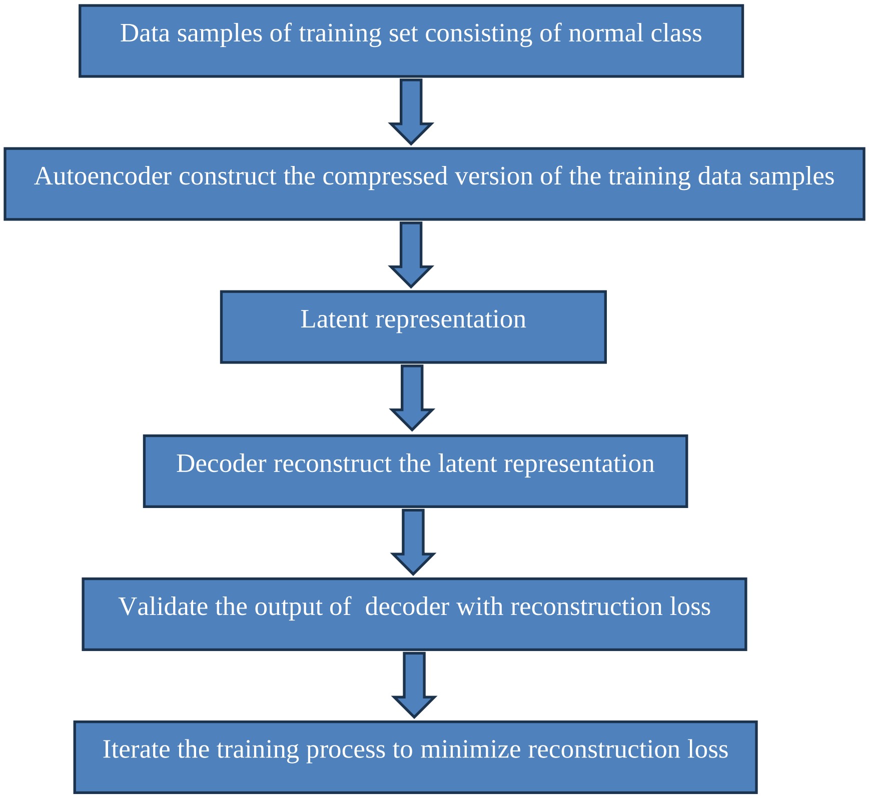

Data samples of training set consisting of normal class
Autoencoder construct the compressed version of the training data samples
Latent representation
Decoder reconstruct the latent representation
Validate the output of decoder with reconstruction loss
Iterate the training process to minimize reconstruction loss

Fig. 3 Training Phase of the proposed Methodology

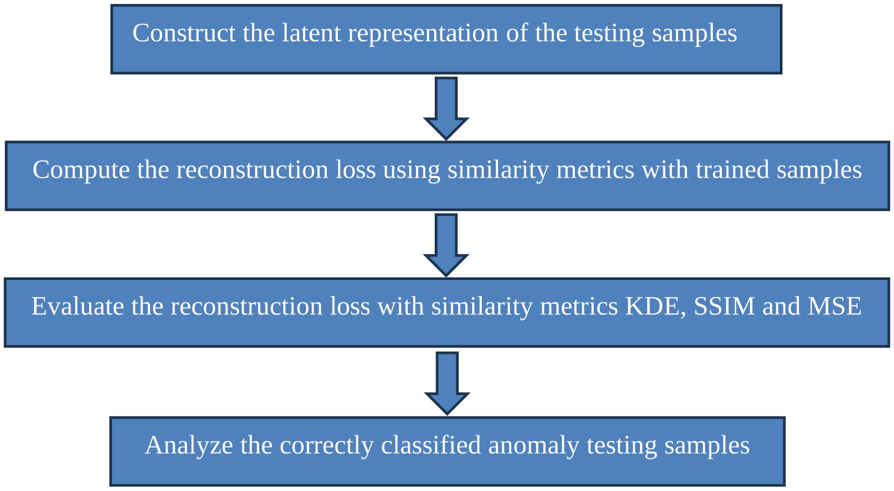

Construct the latent representation of the testing samples
Compute the reconstruction loss using similarity metrics with trained samples
Evaluate the reconstruction loss with similarity metrics KDE, SSIM and MSE
Analyze the correctly classified anomaly testing samples

Fig. 4 Testing Phase of the proposed Methodology

3.2. Working of the Training Phase and Testing Phase

As indicated in Figures 3 and 4, the steps of the proposed methodology for the training and testing phase using custom CAE are depicted. The key component is the encoder and decoder. The encoder creates the latent representation of all the training samples. The training process is iterated to minimize the reconstruction loss calculated for the original training samples and reconstructed samples. This is done to prevent the loss of information in the reconstructed image as the output of the decoder. The hyperparameters used for training the proposed methodology are stated in Table 2. In the testing phase of the proposed methodology, the testing sample is first applied to the proposed custom convolutional autoencoder. This will create the reconstructed image with a compressed representation of the image. The reconstructed image is compared with training data samples with three similarity metrics. The anomaly sample has a high reconstruction error as compared to the original data samples. The testing samples with low reconstruction error are considered normal behavior. However, the threshold is set for reconstruction error in the testing phase. Anomaly is considered if the reconstruction error exceeds the threshold.

Table 3. Architectural details of the proposed methodology

Type of Layer	(Height, Width, Channels)	Parameter Computation
Input Layer	(460,700,1)	-
Encoder		
Convolution Layer	(460, 700, 16)	160
Max Pooling Layer	(230, 350, 16)	-
Convolution Layer	(230, 350, 8)	1160
Max Pooling Layer	(115, 175, 8)	-
Decoder		
Up sampling Layer	(230, 350, 8)	-
Convolutional Layer	(230, 350, 30)	584
Up sampling Layer	(460, 700, 30)	-
Convolutional Layer	(460, 700, 1)	2190
Total Parameter Computation		4365

Table 3 indicates the dimension of the feature map produced by each layer, parameter computation, and the number of feature maps or channels. The number of feature maps depends on the filters used in each layer. Three similarity measures are used in the proposed methodology to evaluate the reconstructed data. i) Mean Squared Error [23] ii) Kernel Density Estimation [24, 25] iii) Structured Similarity Index measure [26, 27] discussed in the next section. Auto encoders are neural network structures utilized for unsupervised learning, primarily focusing on compressing and rebuilding input data.

3.3. Similarity Metrics

The Mean Squared Error (MSE) [23] is an essential metric for evaluating the difference between the original input Y_i and its reconstructed output P_i by the autoencoder, as stated in Equation (1). This measure offers valuable insights into the accuracy of the reconstruction processes. ith data sample represents the total number of samples ranging from 1 to n. More precisely, it quantifies the average square deviation between the pixel or feature values in the original data and their corresponding values in the reconstruction. A reduced MSE implies that the auto encoder successfully captures the fundamental characteristics of the input data and recreates them with minimum loss. This demonstrates a more precise and efficient encoding-decoding process. Therefore, by monitoring the MSE, one may adjust the settings of the autoencoder and evaluate its effectiveness in tasks such as reducing dimensions, removing noise, or detecting anomalies.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{i=n}(Y_i - P_i)^2 \tag{1}$$

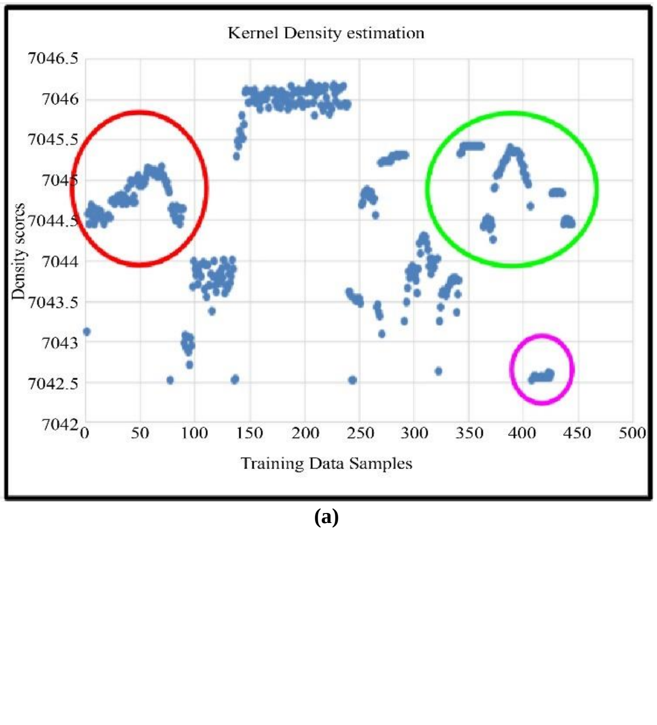
Kernel Density Estimation (KDE) [24, 25] is a method employed in auto encoders to approximate the probability density function of the data inside the latent space. An auto encoder is a model that compresses input into a lower-dimensional representation using an encoder. The encoded characteristics may then be modelled using KDE in the latent space. In anomaly detection, Kernel Density Estimation Kernel Density Estimation (KDE) serves as an effective method for estimating the probability density function of normally distributed data points. Data points that are located in regions of low density can be identified as anomalies. The calculation of a single kernel data point is stated by Equation (2). x represents the neighbouring point, xi represents data points, and h represents the bandwidth. i represents the total number of samples ranging from 1 to n. The parameter h defines the sample window utilized for estimating the probability of a new data point. This specification is crucial for ensuring accurate assessments in our analytical processes.

$$\text{K(x)} = \frac{1}{\text{h}\sqrt{2\pi}}\ \text{e}^{-0.5\left(\frac{\text{x}-\text{x}_\text{i}}{\text{h}}\right)^2} \tag{2}$$

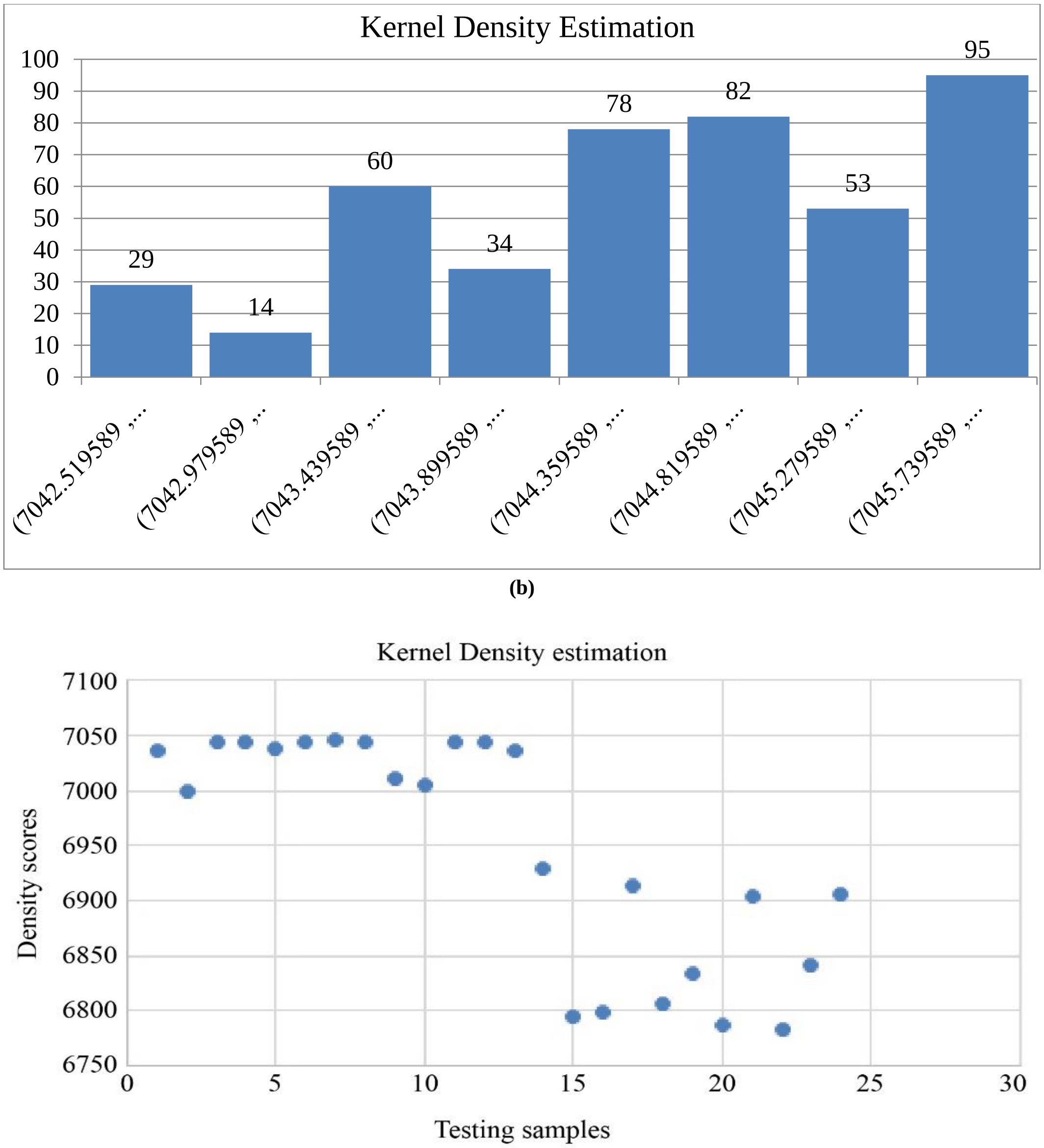
Calculation of Kernel density estimation of data points by summing the kernel of different data points is stated in Equation (3). n indicates the total number of data points.

$$KDE_j = \frac{1}{n}\sum_{\text{i=1}}^{\text{i=n}} \frac{1}{h\sqrt{2\pi}}\ e^{-0.5\left(\frac{x-x_i}{h}\right)^2} \tag{3}$$

The kernel density estimation of training samples, which consists of data points consisting of the normal behaviour of students, lies within the same range. However, data samples consisting of anomaly behaviour lie within different ranges, as stated in Figure 5(a) indicates the training data samples have data distribution consisting of neighbouring density scores and diverse density scores. Figure 5(c) indicates the testing samples consisting of density scores with dissimilar density scores compared to the training samples.



(a)



(b)

(c)

Fig. 5 Kernel Density Estimation, (a) KDE for training data samples, (b) KDE for data samples having the same range of density scores, (c) KDE density scores for testing samples.

In addition, Figure 5(b) indicates the total number of samples with the same density score from training data, which states that the features are similar to those of a particular collection of samples. In addition, Figure 5(c) depicts normal samples with high-density scores and anomaly samples with low-density scores. The Structured Similarity Index Measure (SSIM) [26, 27] is commonly used to assess picture quality.

It is often employed in the context of convolutional auto encoders to evaluate how closely the reconstructed outputs resemble the original inputs. The SSIM metric extracts three important key features from an image: Contrast, Luminance, and Structure. The Luminance Function is characterized by the mathematical representation l (x, y), wherein μ denotes the mean of a specified image, while x and y refer to the two images being compared. Constants C1 and C2 are usually set to C1=0.01 and C2=0.03 as stated in Equation (4)

$$l(x, y) = \frac{2\mu_x\ \mu_y\ + C1}{\mu_x^2 + \mu_y^2 + C2} \qquad (4)$$

The contrast function c(x, y) is a mathematical function that defines the difference or distinction between two elements x and y. It is used to quantify the degree of variation or dissimilarity between the elements, often in the context of images or signals, as stated in Equation (5). σ denotes the standard deviation of a given image. x and y are the two images being compared.

$$c(x,y) = \frac{2\sigma_x \ \sigma_y \ + C1}{\sigma_x^2 + \sigma_y^2 + C2} \quad (5)$$

The structure is defined by the function s(x, y) as stated in Equation (6). σ indicates the standard deviation of a given image. The two images being compared are x and y.

$$s(x,y) = \frac{\sigma_{xy} + C_2}{\sigma_x\sigma_y + C_2} \quad (6)$$

$$\sigma_{xy} \ = \frac{1}{N-1}\sum_{i=1}^{i=N}(x_i - \mu_x)\left(y_i - \mu_y\right) \quad (7)$$

The SSIM metric calculates the structural similarity of two pictures by considering factors such as brightness, contrast, and structural information. This provides a more thorough evaluation of image quality compared to just comparing individual pixels. When employed in auto encoders, SSIM can operate as both a loss function and an evaluation measure to guarantee that the reconstruction preserves crucial structural elements and visual consistency. This is especially advantageous in applications like image denoising, inpainting, or super-resolution, where maintaining the perceived quality of the rebuilt pictures is vital. By prioritizing the optimization for the Structural Similarity Index (SSIM), auto encoders may more effectively capture and replicate the delicate nuances of the input data, resulting in reconstructions of superior quality. where $\alpha > 0$, $\beta > 0$, $\gamma > 0$ denote the relative importance of each of the metrics as stated [26, 27] in Equation (8).

$$\text{SSIM}(x, y) = [\ l(x,y)]^{\alpha} \cdot [\ c(x,y)]^{\beta} \cdot [\ s(x,y)]^{\gamma} \quad (8)$$

4. Experimental Results and Analysis

We evaluated the auto encoder's performance using three metrics: Mean Squared Error (MSE) [23], Kernel Density Estimation (KDE) [24, 25], and the Structured Similarity Index Measure (SSIM) [26, 27]. The details regarding the accuracy of the trained model are stated in Figure 8. The experimental results are evaluated using the confusion matrix, as stated in Figure 7. The results of MSE are stated in Figure 6. which quantifies the difference in the values of each corresponding pixel between the sample and the reference images. The figure x-axis indicates the number of testing samples, and the y-axis represents the means square error. A higher mean square error value indicates that the testing image is anomalous.

A total of 24 testing samples are considered, out of which the ground truth of the first 14 samples is an anomaly, whereas the remaining are normal. Only a few samples are correctly identified as an anomaly based on the MSE, as depicted in Figure 6. The Mean Squared Error (MSE) [23] fails to account for the contextual or structural relationships between data points, as the autoencoder output is a blurred reconstruction. In scenarios where the interplay of features is significant, MSE may inadequately maintain these relationships, potentially resulting in reconstructions close to numerical value but contextually or structurally inaccurate. Figure 6 indicates the mean square error for 24 testing samples consisting of the first 10 samples of anomaly and the remaining 14 normal samples. Higher MSE indicates the pattern deviates from normal samples, indicating anomaly. Low MSE indicates the pattern matches the normal samples. However, few samples are misclassified as normal samples.

From Table 4, it can be depicted that SSIM gives good accuracy as compared to other metrics for anomaly detection. For anomalous images, the SSIM will be lower. The Structural Similarity Index (SSIM) metric extracts 3 key features from an image: Luminance, Contrast, and Structure. The information present in the testing images contains different luminance, contrast, and structure as compared to the normal images trained with the dataset. Due to these characteristics, this metric is ideal for anomaly detection. Once the density estimate is computed, Kernel Density estimation can be used to identify anomalies by evaluating the density of new or existing data points.

Data points that fall in regions of very low density are considered anomalies, as they are unlikely under the estimated distribution of the normal data. A threshold can be set to determine which points are classified as anomalies. Points with a density below this threshold are flagged as potential outliers.

However, selecting the appropriate threshold for density is still challenging. KDE is beneficial in situations where the data does not follow a specific parametric distribution, making it a flexible tool for anomaly detection in various domains. For our research work, KDE needs to yield reasonable accuracy. We have analysed the application of various metrics used for an autoencoder for anomaly detection by conducting experiments on real-time datasets summarised further.

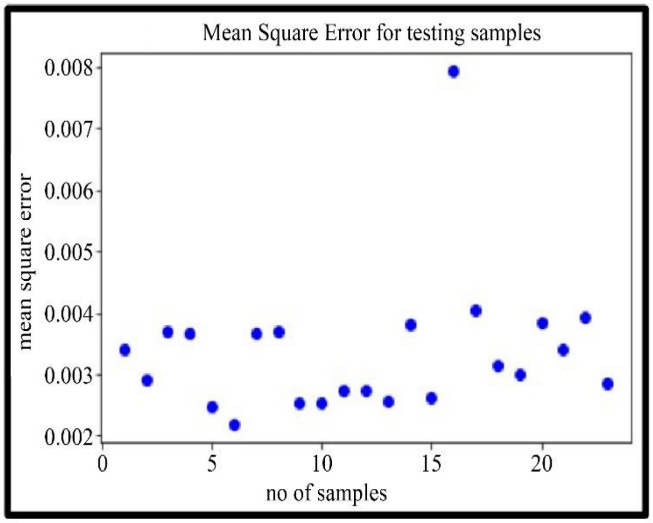
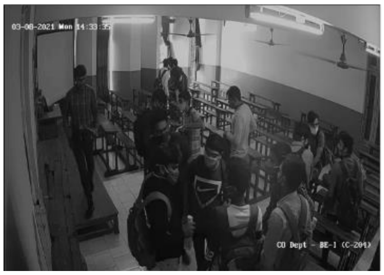
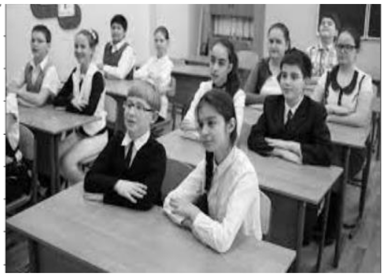
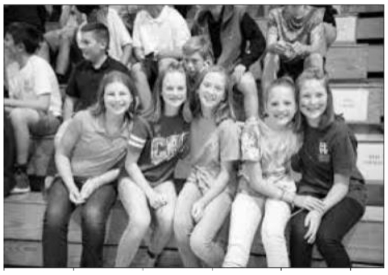
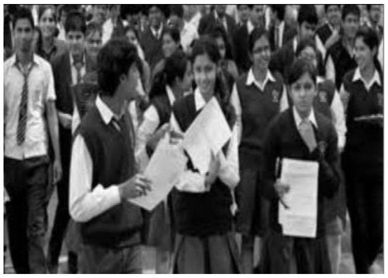
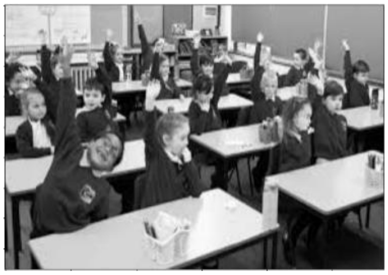
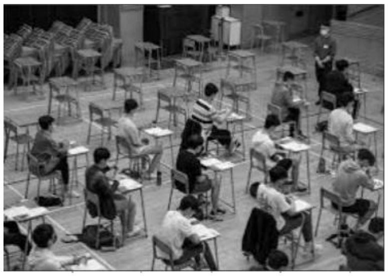


Fig. 6 Mean Squared Error for testing samples concerning training samples

Table 4. Evaluation of the testing samples on the proposed methodology

						
Ground Truth	Anomaly			Normal		
Similarity Metric	SSIM	KDE	MSE	SSIM	KDE	MSE
Predicted Class label	Anomaly	Anomaly	Anomaly	Normal	Abnormal	Normal
						
Ground Truth	Anomaly			Normal		
Similarity Metric	SSIM	KDE	MSE	SSIM	KDE	MSE
Predicted Class label	Anomaly	Normal	Normal	Anomaly	Anomaly	Anomaly
						
Ground Truth	Normal			Normal		
Similarity Metric	SSIM	KDE	MSE	SSIM	KDE	MSE
Predicted Class label	Normal	Abnormal	Normal	Normal	Abnormal	Normal

This study utilized an auto-encoder model to identify anomalies, especially detecting a student standing in a crowd during a theoretical lesson. The proposed methodology is evaluated on the collected dataset, as no such dataset is available publicly. Three main similarity measures are applied to identify anomalies: Kernel Density Estimation (KDE), Mean Square Error (MSE) reconstruction, and Structural Similarity Index Measure (SSIM). The following is an examination of the outcomes derived from these measurements. The students standing had lower density readings, which distinguished them as outliers and uniforms across several test conditions. The KDE metric demonstrated a high level of sensitivity towards slight variations in the latent space. The autoencoder successfully reconstructed non-standing students with low MSE values, showing that the model represented the basic organization of the group well.

TP	**TN**
5	**5**
9	**5**
FP	**TN**

(a) Mean Squared Error

TP	**TN**
6	**4**
12	**2**
FP	**TN**

(b) Kernel Density Estimation

TP	**TN**
10	**0**
1	**13**
FP	**TN**

(c) SSIM

TP: Anomaly Detection TN: Normal Class

Fig. 7 Confusion matrix of the experimental results for anomaly

However, the MSE values for the standing student were much higher, indicating difficulty in accurately reproducing this specific event. For blurred reconstruction, among the average students in the group, the SSIM values were high, indicating that the recreated pictures maintained similar structural features to the original ones.

However, the SSIM values for the standing students were lower compared to those of the seated students, showing a noticeable structural difference in the reconstructions. This suggests that the auto encoder did not preserve the standing student's posture well, as indicated by the SSIM measure. SSIM was effective in identifying precise structural abnormalities that may have been missed by MSE.

The lower SSIM values observed for the standing student highlight the importance of structural information in detecting abnormalities, making SSIM a valuable addition for evaluation compared to MSE and KDE for evaluating structural integrity. Each of the three measures, namely KDE, MSE, and SSIM, had a distinct role in identifying the anomaly (the standing student) in the crowd during a theoretical lecture.

The SSIM metric offered valuable information on structural variations, accurately capturing noticeable discrepancies. Using these parameters, the auto-encoder-based anomaly detection system successfully identified the standing student as an abnormality. This demonstrates the model's capacity to recognize both subtle and substantial anomalies in a classroom environment.

Figure 9 states the number of testing samples correctly predicted using the proposed methodology using various similarity metrics. It depicts that anomaly is correctly predicted for given testing samples using SSIM.

The proposed methodology is evaluated in a classroom environment that is not considered for training samples, students gathering in different crowds, and the normal arrangement of students in various classroom environments.

The testing samples related to anomalies in the classroom environment were collected from the web. However, it is considered abnormal for certain examples where students sit in groups on the staircase. This sample is evaluated to test the generalisation ability of the proposed methodology.

In addition, the proposed methodology is also evaluated on the UCSD [29] Anomaly Dataset. UCSD [29] dataset consists of samples of educational institution pedestrian walkways. The crowd's density in the walkways varied, ranging from sparse to very crowded. Under normal conditions, the video captures only pedestrians.

Abnormal events are defined as either the presence of non-pedestrian entities in the walkways or unusual pedestrian motion patterns. The proposed model is specifically designed for the scenario of the classroom environment, not the outdoor one.

However, it performed well on the UCSD Dataset for identifying abnormal behavior cases, such as non-pedestrian entities and increased crowd density, as stated in Table 5. It also states the accuracy of the UCSD [29] dataset using various techniques recently proposed by various researchers.

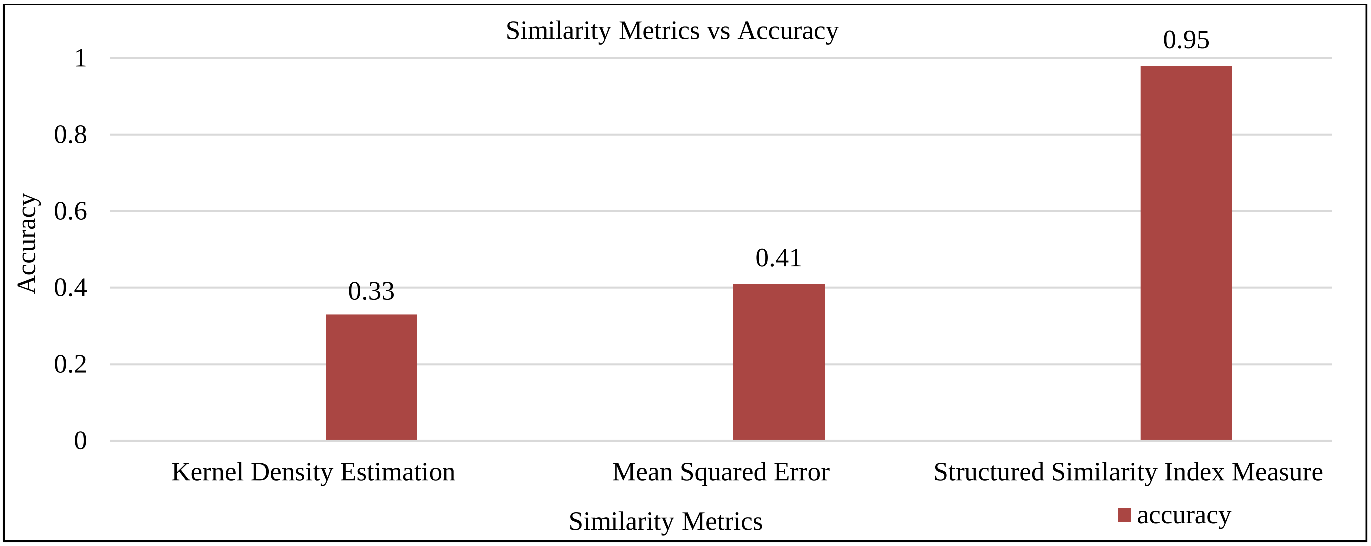


Fig. 8 Comparison of similarity metrics vs accuracy on the EDUANOM dataset

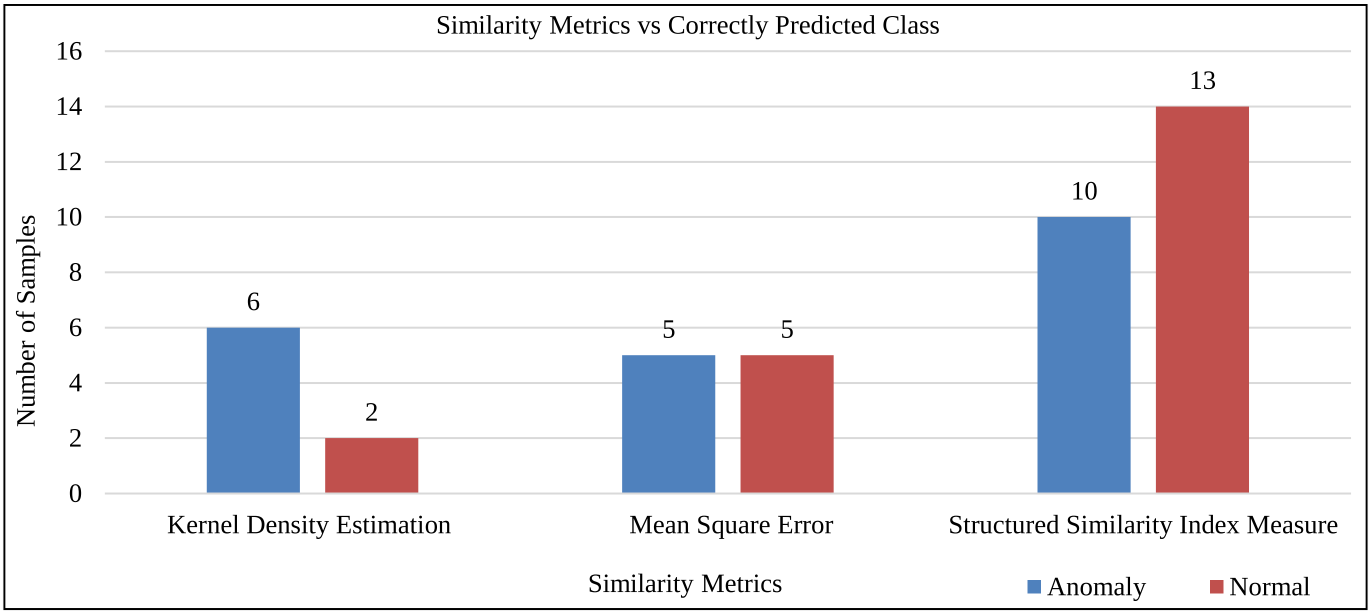


Fig. 9 Comparison of similarity metrics vs number of samples correctly predicted for EDUANOM

Table 5. Comparison of accuracy on the dataset

Methodology	Dataset	Accuracy
Two Stream Convolutional Neural Network [28]	UCSD [29]	84%
AED using GAN [30]	UCSD [29]	95%
Proposed	UCSD [29]	90%
Proposed	EDUANOM	95%

5. Discussion

A new approach has been developed to detect abnormal behaviour of students in the classroom using a custom Convolutional Autoencoder (CAE) combined with various similarity metrics. The gathering of students in the classroom environment is considered abnormal, especially in cases where social distancing needs to be ensured. This innovative method delivers exceptional accuracy and efficiency. Here are the key points: i) Role of the Convolutional Autoencoder: The CAE is essential for identifying patterns and simplifying complex data. It effectively recognizes intricate behaviours in students and classroom activities. The CAE confidently distinguishes between normal and abnormal patterns by reconstructing input data, with significant reconstruction errors indicating potential anomalies. ii) Significance of Similarity Metrics: Similarity metrics are crucial for measuring original and reconstructed data differences. Key metrics include: - Mean Squared Error (MSE): This metric highlights pixel-level differences, with higher MSE values indicating potential anomalies. - Structural Similarity Index (SSIM): This metric effectively focuses on visual characteristics, making it superior at detecting subtle behavioural changes. The CAE operates efficiently without needing extensive labelled data, cutting down on preparation time. Quick calculations ensure real-time results, which are critical for immediate responses. Challenges remain in fine-tuning the CAE to avoid overfitting and selecting the most effective metrics for various anomalies. Future research will focus on adaptive metrics and hybrid models that will further enhance the detection of temporal changes. By leveraging a custom CAE and diverse similarity metrics, there is a strong potential to significantly improve the detection of unusual activities in schools significantly, ensuring swift identification and response to any issues.

6. Conclusion

Anomaly detection is a challenging area in the field of computer vision and deep learning. In this research work, we have collected a real-time data set consisting of samples of students in the crowd in the classroom environment. In this paper, the proposed methodology is based on unsupervised deep learning architecture based on a custom convolutional autoencoder. Anomaly is detected based on similarity metrics such as Kernel Density Estimation (KDE), Mean Square Error (MSE), and Structure Similarity Index measure (SSIM). We achieved good accuracy with SSIM as compared to KDE and MSE. SSIM is effective in identifying structural abnormalities. Our proposed model has less parameter computation and good accuracy. This research work is useful in applications for anomaly detection in the classroom environment based on unsupervised deep learning architecture. It requires various data samples for normal behavior. However, the limitation of the auto encoder is the tendency of the model to overfit, especially when subtle variations within normal data are incorrectly flagged as anomalies. This overfitting can cause the auto encoder to yield elevated reconstruction errors for minor deviations, diminishing its overall effectiveness. Another key limitation is the challenge of determining an optimal threshold to separate normal and anomalous data. Setting this threshold often demands substantial tuning and expert knowledge, as there can be overlap in the reconstruction error distributions of normal and anomalous data. In future work, the proposed work can be extended by including a huge number of data samples consisting of various normal actions for students or educational institutions. The robustness and scalability of the anomaly detection system allow the experiment with variants of Autoencoder with varied and diverse datasets for the education environment. This will help improve the system's generalization ability for anomaly detection. In addition, a hybrid approach can be explored to extract temporal features for anomaly detection.

References

[1] QinMin Ma, "Abnormal Event Detection in Videos Based on Deep Neural Networks," *Scientific Programming*, vol. 2021, pp. 1-8, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] Mahmood Yousefi-Azar et al., "Autoencoder-Based Feature Learning for Cyber Security Applications," *2017 International Joint Conference on Neural Networks*, Anchorage, AK, USA, pp. 3854-3861, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[3] Pierre Baldi, "Autoencoders, Unsupervised Learning, and Deep Architectures," *Proceedings of ICML Workshop on Unsupervised and Transfer Learning, PMLR 27*, pp. 37-50, 2012. [Google Scholar] [Publisher Link]

[4] Junhai Zhai et al., "Autoencoder and Its Various Variants," *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Miyazaki, Japan, pp. 415-419, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[5] Markus Goldstein, and Seiichi Uchida, "A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data," *PloS one*, vol. 11, no. 4, pp. 1-31, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[6] Charu C. Aggarwal, *An Introduction to Outlier Analysis*, Outlier Analysis, Springer, pp. 1-34, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[7] Durgesh Samariya, and Amit Thakkar, "A Comprehensive Survey of Anomaly Detection Algorithms," *Annals of Data Science*, vol. 10, no. 3, pp. 829-850, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Jyothi Honnegowda, Komala Mallikarjunaiah, and Mallikarjunaswamy Srikantaswamy, "An Efficient Abnormal Event Detection System in Video Surveillance using Deep Learning-Based Reconfigurable Autoencoder," *Information Systems Engineering*, vol. 29, no. 2, pp. 677-686, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Mehmet Saygın Seyfioğlu, Ahmet Murat Özbayoğlu, and Sevgi Zubeyde Gürbüz, "Deep Convolutional Autoencoder for Radar-Based Classification of Similar Aided and Unaided Human Activities," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1709-1723, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[10] Dan Xu et al., "Learning Deep Representations of Appearance and Motion for Anomalous Event Detection," *British Machine Vision Conference*, pp. 1-12, 2015. [Google Scholar] [Publisher Link]

[11] Yingying Zhu, Nandita Nayak, and Amit Roy-Chowdhury, "Context-Aware Activity Recognition and Anomaly Detection in Video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 1, pp. 91-101, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[12] Nasaruddin Nasaruddin et al., “Deep Anomaly Detection through Visual Attention in Surveillance Videos,” *Journal of Big Data*, vol. 7, no. 1, pp. 1-17, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Mahmudul Hasan et al., “Learning Temporal Regularity in Video Sequences,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 733-742, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[14] Shifu Zhou et al., “Spatial-Temporal Convolutional Neural Networks for Anomaly Detection and Localization in Crowded Scenes,” *Signal Processing: Image Communication*, vol. 47, pp. 358-368, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[15] Waqas Sultani, Chen Chen, and Mubarak Shah “Real-World Anomaly Detection in Surveillance Videos,” *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 6479-6488, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[16] Trong Nguyen Nguyen, and Jean Meunier “Anomaly Detection in Video Sequence with Appearance-Motion Correspondence,” *2019 IEEE International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), pp. 1273-1283, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[17] Yong Shean Chong, and Yong Haur Tay, “Abnormal Event Detection in Videos using Spatiotemporal Autoencoder,” *Proceedings Part II 14th International Symposium Advances in Neural Networks ISNN, Lecture Notes in Computer Science*, Sapporo, Hakodate, and Muroran, Hokkaido, Japan, pp. 189-196, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[18] K. Chidananda, and A.P. Siva Kumar, “VidAnomalyNet: An Efficient Anomaly Detection in Public Surveillance Videos through Deep Learning Architectures,” *International Journal of Safety and Security Engineering*, vol. 14, no. 3, pp. 953-966, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Karishma Pawar, and Vahida Attar, “Application of Deep Learning for Crowd Anomaly Detection from Surveillance Videos,” *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, pp. 506-511, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] Jaroslav Kopcan, Ondrej Skvarek, and Martin Klimo, “Anomaly Detection using Autoencoders and Deep Convolution Generative Adversarial Networks,” *Transportation Research Procedia*, vol. 55, pp. 1296-1303, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[21] Kamal Berahmand et al., “Autoencoders and Their Applications in Machine Learning: A Survey,” *Artificial Intelligence Review*, vol. 57, no. 2, pp. 1-52, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Xing Fang, “Understanding Deep Learning Via Backtracking and Deconvolution,” *Journal of Big Data*, vol. 4, no. 1, 1-14, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[23] Timothy O. Hodson, “Root-Mean-Square Error (RMSE) or Mean Absolute Error (MAE): when to use them or not,” *Geoscientific Model Development*, vol. 15, no. 14, pp. 5481-5487, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[24] Stanisław Węglarczyk, “Kernel Density Estimation and its Application,” *ITM Web Conference*, vol. 23, pp. 1-8, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[25] Adriano Z. Zambom, and Ronaldo Dias, “A Review of Kernel Density Estimation with Applications to Econometrics,” *International Econometric Review (IER)*, vol. 5, no. 1, pp. 20-42, 2013. [Google Scholar] [Publisher Link]

[26] Jim Nilsson, and Tomas Akenine-Möller, “Understanding SSIM,” *Image and Video Processing*, pp. 1-8, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] Paul Bergmann et al., “MVTec AD-A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection,” *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, pp. 9584-9592, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[28] Abdelhafid Berroukham et al., “Deep Learning-Based Methods for Anomaly Detection in Video Surveillance: A Review,” *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 1, pp. 314-327, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[29] Vijay Mahadevan et al., “Anomaly Detection in Crowded Scenes,” *2010 IEEE Transactions on Pattern Analysis and Machine Intelligence*, San Francisco, CA, USA, pp.1975-1981, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[30] Mahdyar Ravanbakhsh et al., “Abnormal Event Detection in Videos using Generative Adversarial Nets,” *Computer Vision and Pattern Recognition*, pp. 1-5, 2017. [CrossRef] [Google Scholar] [Publisher Link]