Enhancing anomaly detection performance using ResNet50 and BiLSTM networks on benchmark datasets

Dipak Ramoliya[1], Amit Ganatra[2]
[1]Department of Computer Science and Engineering, Devang Patel Institute of Advance Technology and Research (DEPSTAR),
Faculty of Technology and Engineering, Charotar University of Science and Technology (CHARUSAT), Anand, India
[2]Department of Computer Science and Engineering, Faculty of Engineering and Technology, Parul University (PU), Vadodara, India

Article Info	ABSTRACT
	Detection of abnormal activity from large video sequences is one of the biggest challenges because of ambiguity in different activities. Over the last many years, several cameras have been placed to cover the public and private sectors to monitor abnormal human activity and surveillance. In recent years, deep learning and computer vision have significantly impacted this kind of surveillance. Intelligent systems that can automatically identify unusual events in video streams are currently in high demand. A deep learning-based combinational model has been proposed to detect abnormal activity from input video streams. The proposed study uses a combination of convolution and sequential models. A ResNet50 network with a residual connection was used for initial feature extraction. The proposed bidirectional long short-term memory (BiLSTM) network has improved the extracted ResNet50 features. Simulation of the proposed model was experimented on two benchmark datasets for anomaly detection UCF Crime and ShanghaiTech. Simulation of proposed architecture has achieved 97.55% and 91.94% remarkable accuracy for UCF Crime and ShanghaiTech datasets respectively. *This is an open access article under the CC BY-SA license.*

Corresponding Author:

Dipak Ramoliya
Department of Computer Science and Engineering, Devang Patel Institute of Advance Technology and
Research (DEPSTAR), Faculty of Technology and Engineering, Charotar University of Science and
Technology (CHARUSAT)
Changa, Anand - 388421, India
Email: dipakramoliya.ce@charusat.ac.in

1. INTRODUCTION

In surveillance applications, anomalous activity detection plays a crucial role. It is possible to use automatic video capture to record anomalous human behavior without the need for system intervention [1]. Due to its complexity, the countless potential for aberrant cases, and the scarcity of available normal ones, it remains one of the most challenging and well-known research areas [2]. Surveillance cameras are almost everywhere these days, from security-sensitive locations—such as borders and military installations used to monitor terrorist activity—to private residences meant to deter burglaries. Due to weariness and monotony, manual workers were unable to maintain diligent monitoring over an extended length of time due to the large volume of data involved. Traditionally, the monitoring function is handled by human operators. These operators have to simultaneously visually evaluate many camera feeds. Technological advancements and falling costs have led to an acceleration of the installation of surveillance cameras in both public and private spaces [3]. The main problem with this approach is that the detection rate of anomalies decreases rapidly when used in crowded environments when occlusions and clutter arise.

Anomaly detection, a critical task in various domains such as finance, cybersecurity, industrial monitoring and healthcare involves recognizing patterns in data that deviate meaningfully from the norm. These anomalies can indicate significant but rare events like fraud, system failures, or unusual medical conditions, necessitating timely and accurate detection to mitigate potential risks and impacts. Traditional methods for anomaly detection often struggle with the complexity and high dimensionality of modern data, making advanced techniques essential for effective analysis [4].

Long short-term memory (LSTM) networks, a form of recurrent neural network (RNN), have demonstrated extraordinary success in sequential data processing due to their capacity to capture long-term dependencies. However, standard LSTM models process data in a single temporal direction, potentially missing contextual information that could enhance anomaly detection accuracy [4]. Bidirectional LSTM (BiLSTM) networks overcome this problem by processing data in both forward and backward directions, thus leveraging the complete temporal context [5].

Anomaly detection using deep learning presents several significant challenges, despite its potential to outperform traditional methods. One of the primary challenges is the requirement for large amounts of labelled training data, which is often scarce in real-world scenarios [6]. Anomalies, by nature, are rare and diverse, making it difficult to gather a representative and sufficiently large dataset for training deep learning models. Additionally, deep learning models, especially those with complex architectures, are computationally intensive and require substantial resources for both training and inference [4], [5]. This can be a barrier to deployment in environments with limited computational capabilities. Moreover, deep learning models can sometimes struggle with interpretability, making it hard to understand why a particular instance was classified as an anomaly, which is crucial in many applications where transparency and explainability are important. Furthermore, these models are sensitive to hyperparameter tuning and architecture design, requiring extensive experimentation and expertise to achieve optimal performance [4].

In this article, the authors have proposed an advanced architecture for anomaly detection. The major contributions of the article are i) proposed advanced convolution and BILSTM-based sequential architecture for anomaly detection, ii) simulation of the proposed study explore the effectiveness of the study by comparing the SOTA deep learning model for anomaly detection. The proposed study contributes to improving video surveillance in sensitive sectors.

The rest of the article is organized as section 2 demonstrates literature analysis and comparative analysis of recent work on anomaly detection using deep learning methods. Section 3 presents the proposed BiLSTM architecture with residual connections. Section 4 illustrates the simulation of the proposed architecture and statistical analysis with SOTA deep learning models over different benchmark datasets.

2. LITERATURE SURVEY

There are two types of methods for identifying unusual activity: low-level anomaly detection and high-level anomaly detection. Low-level anomaly detection techniques locate local spatiotemporal regions that likely have aberrant low-level feature patterns before high-level analysis, including activity classification and object tracking, is carried out. [7] inter-activity context feature-based architecture proposed by Zhu et al. [7]. The proposed study uses a greedy search algorithm for point-based contextual anomaly detection. VIRAT ground dataset used for the simulation of the proposed study.

Nguyen and Meunier [8] proposed a U-Net-based architecture for anomaly detection from video sequences. The proposed study used convolution architecture including 4 streams of convolutions of filter sizes 1×1, 3×3, 5×5 and 7×7 for feature extraction. Simulation of the proposed study used benchmark datasets like Avenue and UCSD Ped2. The proposed study was able to achieve 0.869 and 0.962 accuracy respectively.

A faster R-CNN-based DeepAnomaly model was proposed by Christiansen et al. [9]. to identify unusual activity in films. The suggested architecture can identify people up to 90 meters away. To identify anomalies, pre-train convolution models were employed. The suggested study makes use of a proposed approach that offers between 300 and 2,000 regions per image. Subsequently, a categorization network designates a label for every region and routes it via it. DeepAnomaly is a helpful technique for region suggestion when there are few or no areas per image.

Sabokrou et al. [10] proposed an autoencoder (AE) based model for anomaly detection. The proposed architecture finds errors in abnormal parts in the reconstruction phase of sparse AE. Simulation of proposed architecture uses UCSD Ped2 and UMN benchmark datasets and patch size with 30×30 x10Proposed study marks patches as abnormal if more than 40% pixels are detected. Simulation finds 99.6 AUC with an 82% detection rate.

Khan et al. [11] proposed convolution-based architecture for monitoring of irregular human actions, and traffic surveillance. Authors have also contributed to surveillance datasets such as the vehicle accident

image dataset (VAID). The dataset contains of 1360 accidental images of vehicles. The authors have extended the simulation of the proposed architecture with different types of pooling, activation, and optimizer methods. The proposed study finds 82% accuracy using modified convolutional neural network (CNN) models.

Le and Kim [12] proposed a video anomaly detection system employing an attention-based residual autoencoder architecture. The suggested network uses both spatial and temporal information in a single, integrated network and is based on unsupervised learning. Simulation of the proposed study uses three different benchmark datasets UCSD Ped2, CUHK, and Shanghaitech. Simulation of the proposed study finds AUC values as 73.6%, 86.7% and 97.4% for ShanghaiTech, CUHK Avenue, and UCSD Ped2 datasets respectively.

Avola et al. [13] proposed anomaly detection based on generative adversarial network (GAN). The proposed study uses GAN architecture for aerial video surveillance at low altitudes. Proposed study focuses on low-altitude sequences in complex scenarios wherein a tiny object or gadget may be cause for concern or danger. Simulation of the study uses benchmark change detection (UMCD) and unmanned aerial vehicle (UAV) I mosaicking dataset. Simulation of GAN architecture finds 95.7% structural similarity index (SSIM). Table 1 presents a comparative examination of various deep learning methodologies for anomaly identification [14]–[18].

Table 1. Comparative analysis with other deep learning approaches for anomaly detection [14]–[18]

Authors	Year	Model	Dataset	Result
Zaheer et al. [14]	2020	One-Class Classifier	UCSD Ped2	98.1%
Tang et al. [15]	2020	U-Net	Avenue	85.1%
Zhou et al. [16]	2019	LSTM	Avenue	86.1%
Wan et al. [17]	2021	LSTM	UCF Crime	95.12%
Doshi and Yilmaz [18]	2023	YOLO-v4	Avenue	85.78%

3. METHODOLOGY

The proposed study uses a combinational approach for anomaly detection. The proposed study used ResNet architecture for feature extraction followed by BiLSTM [4] to enhance features based on feedback learning. The architecture of the proposed study is mainly divided into three major categories as feature extraction with ResNet50, feature enhancement with BiLSTM, and classification with multi-layer perceptron (MLP) [19]. Figure 1 demonstrates the conceptual architecture of the proposed architecture. Five layers of convolution layers interconnected with max pooling for feature exactions. A stack of convolution layers is interconnected with a residual layer to reduce the issue of vanishing gradients in deep neural networks. The second phase of architecture consists of BiLSTM. LSTM architecture also helps to prevent vanishing issues for feature learning proposed study has used four layers of BiLSTM. Where a series of LSTM models interconnected in counter order with each other, generally called as forward and backward pass of the LSTM network [20].
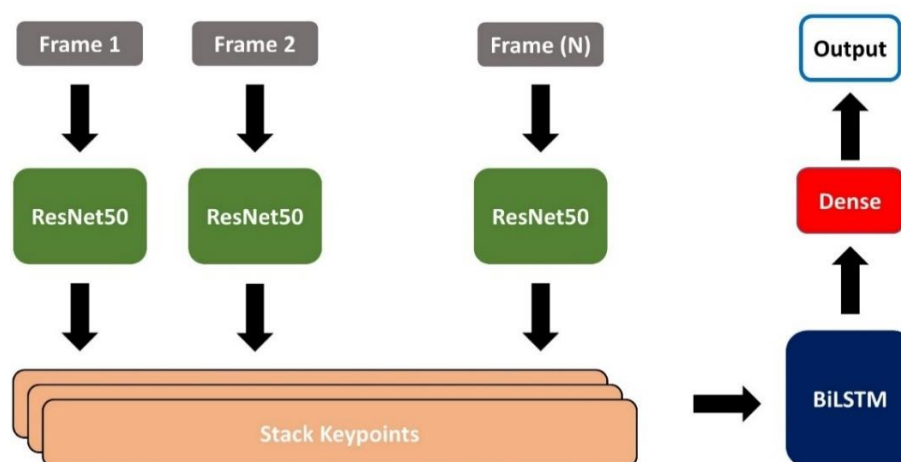

Figure 1. Residual-based BiLSTM proposed architecture for anomaly detection [20]

The proposed architecture uses five layers of convolution with skip connection with output sizes of (112× 112, 56×56, 28×28, 14×14. 7×7) followed by a polling layer. Intermediate layers of residual connections are separated with normalized layers with the same kernel size. The residual module aims to include the network-extracted features. Furthermore, the vanishing gradient problem is mitigated by the usage of short connections. With ResNet serving as the foundation, Figure 2 illustrates the fundamental workings of convolution learning. Equation (1) illustrates the normalizing procedure used to normalize features during ReLU's activation. Through the use of skip connections, which combine the outputs of stacked layers with those from previous layers, significantly deeper networks may now be trained than in the past [21].

$$f(x) = \begin{cases} x, if\ x > 0 \\ 0, rest\ of\ the\ cases \end{cases} \tag{1}$$
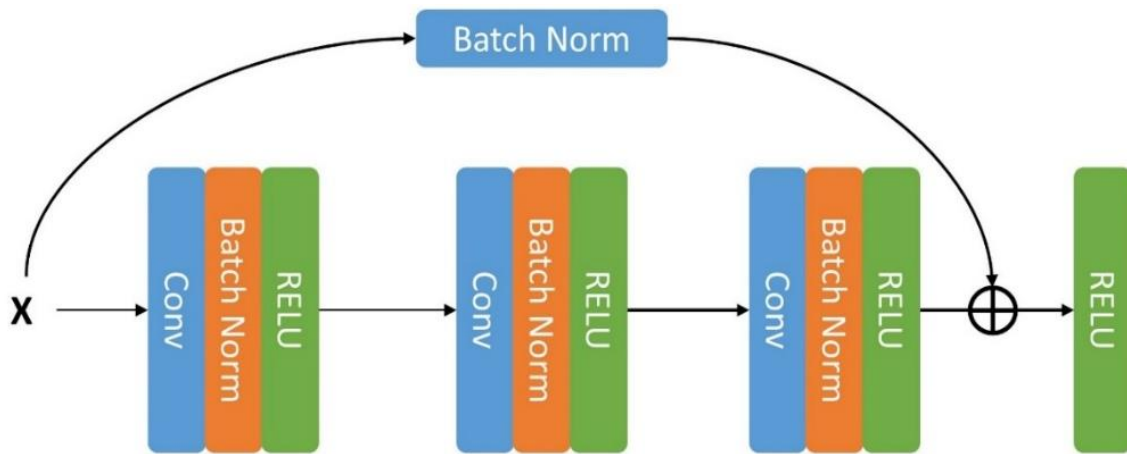

Figure 2. Conceptual design of residual connection used in the proposed study [21]

The ResNet module is composed of the basic block and the identity mapping. Assume that the input vector of ResNet in the $l - th$ module is x_l and the output vector of $l - th$ module is x_{l+1}. the output vector of the ResNet module is formulated as (2) and (3) [22].

$$f(x_l, \{W_l^i\}) = ((W_l^2)^T \sigma(W_l^1)^T x_l) \tag{2}$$

$$x_{l+1} = \sigma\left(f(x_l, \{W_l^i\})\right) + x_l \tag{3}$$

3.1. BiLSTM

Long short-term memory, or LSTM, is frequently thought of as a development of recurrent neural networks (RNN). The capacity of RNN to function as "short-term memory" made it possible to apply prior knowledge—but only up to a certain point—for the current task. LSTM architecture, which is an extension of RNN, offers "long-term memory," or the ability to store all of the prior data for a given neural node, rather than just a specific point in time [23].

LSTM architecture consists of three gates Input, output, and forget. The network's present long-term memory, which houses the history of data, is referred to as the "cell state." Short-term memory can be thought of as the output at a previous point in time, which is referred to as the previous Hidden State. The input value at the current time step is contained in the input data. BiLSTM adopts consideration of both the past and future properties of the network. The BiLSTM hidden layer consists of two components: the forward and backward LSTM cell states. To participate in the forward and reverse calculations, network traffic information first enters the hidden layer through the input layer, and the data computed by the hidden layer is subsequently communicated to the output layer; ultimately, the output layer combines the forward and reverse LSTM outputs based on a predetermined weight to produce the desired output. Figure 3 shows the structure of the BiLSTM network.
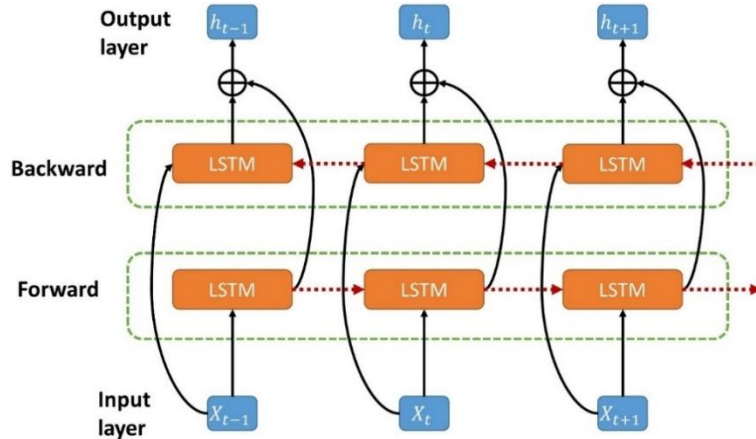
Figure 3. Functional architecture of BiLSTM used for anomaly detection [23]

Bi-direction LSTM network allows the model to learn inter-dependent features from input video. Initially, feature matrices are processed with a forward pass of LSTM and consequently calculated in backward pass also. The proposed study has used different interconnected layers of BiLSTM. However, good accuracy is found in the four layers architecture of BiLSTM. The hidden state of the proposed architecture ht can be calculated as (1). Whare C_t, as learning of cell state formulated as (2) [24].

$$h_t = o_t \times tanh(C_t) \tag{1}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C_t} \tag{2}$$

Forward and backward passes consist of three gates as the vanilla architecture of LSTM. The input gate of the proposed architecture helps to identify relevant information from anomaly frames (video). The output of input gate is set as enabling abnormal features for the rest of the video data its set to 0 as disabled. Similarly, output gate enables and disables based on feature information, whether it should be forwarded to the next layer or not. Formulation of input and output gate as (3) and (4) [25].

$$it = \sigma(Wi \cdot [ht - 1, xt] + bi) \tag{3}$$

$$ot = \sigma(Wo \cdot [ht - 1, xt] + bo) \tag{4}$$

Output and input gates are interconnected in forward and backward propagation anomaly features. However, the behavior of input and output gate control by memory cell, which is updated by forget gate (formulated as equation 5) of the proposed architecture. The output of the BiLSTM layer is the concatenation of the forward and backward hidden states at each time step as equation 6. Where is the forward hidden state at time step t and $h_t^{(b)}$ is the backward hidden state at time step t.

$$ft = \sigma(Wf \cdot [ht - 1, xt] + bf) \tag{5}$$

$$ht = [ht(f), ht(b)] \tag{6}$$

The proposed architecture improves the recognition of abnormal activity from video sequences [25]. The combinational study of convolution and recurrent networks helps to extract advanced features from complex activity videos. The deeper architecture of ResNet50 helps to reduce vanishing issues and generate a feature matrix which is further analyzed with Bi-direction LSTM architecture for improvises in the detection of abnormality in common areas.

4.   RESULTS AND DISCUSSION

In this section, it is explained the results of research and at the same time is given the comprehensive discussion. Results can be presented in figures, graphs, tables and others that make the reader understand easily [14], [15]. Simulation of proposed combinational architecture implemented on two

diverse anomaly detection datasets as University of Central Florida (UCF) Crime [26] and ShanghaiTech Campus dataset [27]. Properties of benchmark anomaly detection datasets are demonstrated in Table 2.

Table 2. Dataset information

Dataset	Classes	Avg. Frames	Length
UCF Crime	14	7247	128 hr
ShanghaiTech	13	24415	130 hr

The proposed approach is simulated on a PC with 32 GB RAM, GeForce RTX 3080, and Core i7 processor. The model was executed using Python's Keras package and TensorFlow. Equations (7), (8), and (9) are used to build evaluation matrices such as precision, recall, F1-Score, and accuracy [28]. The simulation of the proposed study has achieved remarkable performance as 97.55% accuracy for the UCF Crime dataset and 91.94% accuracy for the ShanghaiTech dataset. The accuracy and loss graph of the proposed ResNet50 with BiLSTM architecture is illustrated in Figure 4.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{7}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{8}$$

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Positive} \tag{9}$$
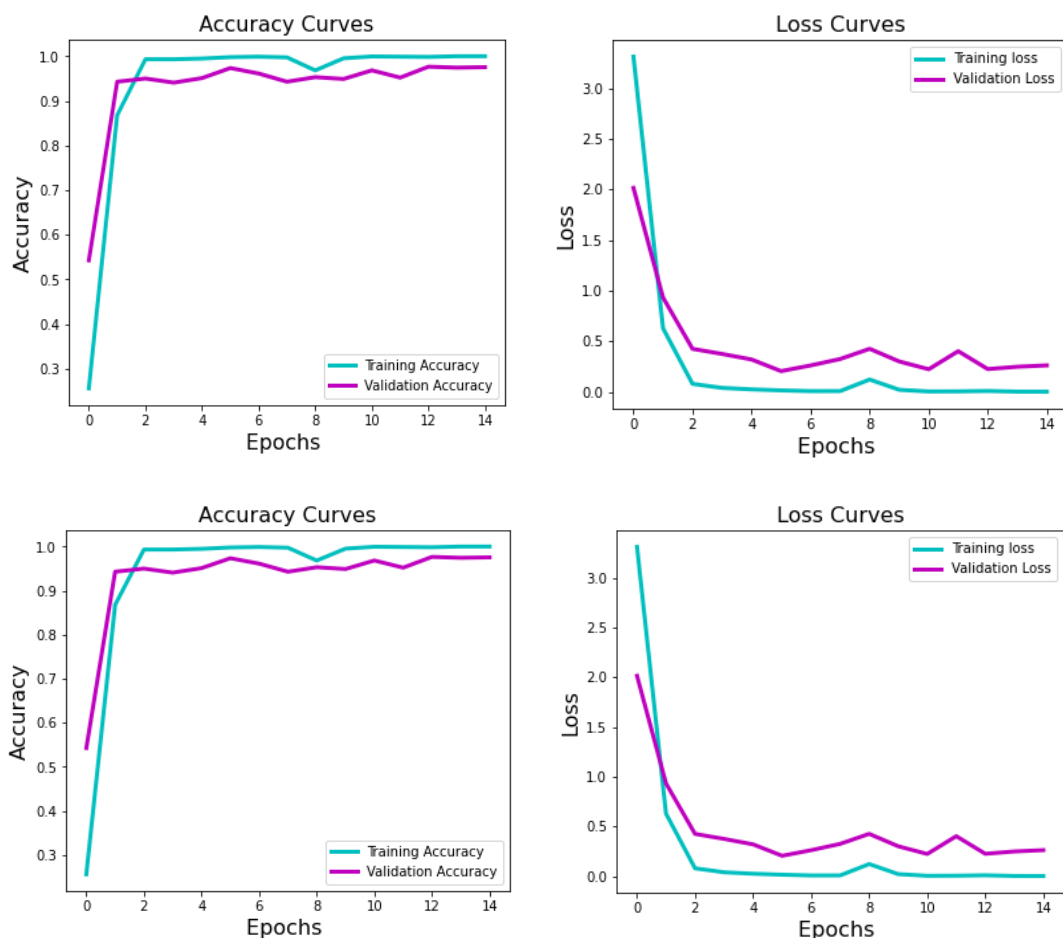

Figure 4. Accuracy and loss graphs of the proposed model with both benchmark datasets

Initially, experiments were conducted on the core architecture of convolution and recurrent networks to identify the potential of anomaly detection for both datasets, as demonstrated in Table 3. Implementation of anomaly detection using core and state-of-the-art models finds good accuracy, but not promising to make the model more generalized and robust. The simulation of the study uses core (4-layer) recurrent network models such as recurrent neural network (RNN), long short-term memory (LSTM), and gated recurrent unit (GRU) [29]. All this recurrent addition of the BiLSTM network improves the recognition rate of abnormal activities in both datasets.

Table 3. Relative analysis of individual deep learning methods for anomaly detection

Model	Dataset	Accuracy
CNN	UCF	90.61
	ShanghaiTech	86.06
RNN	UCF	85.82
	ShanghaiTech	74.37
LSTM	UCF	94.51
	ShanghaiTech	89.76
GRU	UCF	94.78
	ShanghaiTech	88.45
BiLSTM	**UCF**	**95.22**
	ShanghaiTech	**90.01**

Simulation of the proposed study also compared with different combinations over convolution model such as vanilla CNN, DenseNet, ResNet50/34, and VGG16 [30], [31]. Table 4 demonstrates a comparative analysis of the SOTA convolution model with the proposed Deep BiLSTM model. Simulation of the proposed study also extends to analyze the impact of different hyperparameters such as batch size, split ratio, and learning rate. Figure 5 illustrates the impact of the learning rate and train-test split ratio for both datasets. The simulation was also analyzed with different numbers of epochs and learning layers in the proposed BiLSTM architecture.

Table 4. Comparative analysis of different combinational architectures for anomaly detection

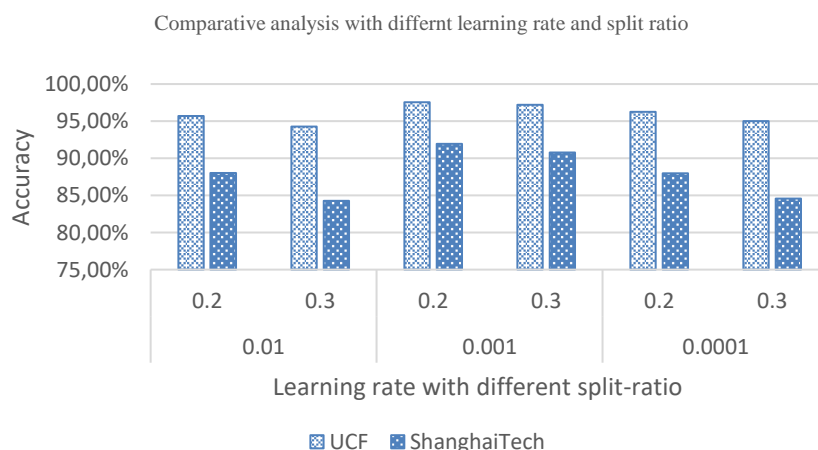| Model	Params	Accuracy	Precision	Recall
CNN+BiLSTM	14989	92.67	0.92	0.92
VGG16 + BiLSTM	15208	93.09	0.93	0.92
DenseNet + BiLSTM	17594	95.91	0.95	0.96
ResNet34 + BiLSTM	16248	94.28	0.94	0.94
ResNet50 + BiLSTM	**18464**	**97.55**	**0.97**	**0.97**

Figure 5. Comparative analysis of proposed architecture with different learning parameters

5. DISCUSSION

The proposed study analyzes the detection of abnormal activity from benchmark datasets like UCF crime and ShanghaiTech using enhanced deep learning models such as ResNet50 with BiLSTM. Simulation

of proposed architecture finds remarkable performance with 93.81 and 87.91% accuracy for benchmark datasets. The simulation section also analyzes the comparison with different SOTA models for anomaly detection and identifies the impact of standalone convolution and recurrent networks. Simulation of the proposed study improves recognition accuracy for abnormal activity by combining convolution and recurrent architecture. The performance of the proposed architecture for unseen prediction data is illustrated in Figure 6 as a confusion matrix for UCF crime datasets.
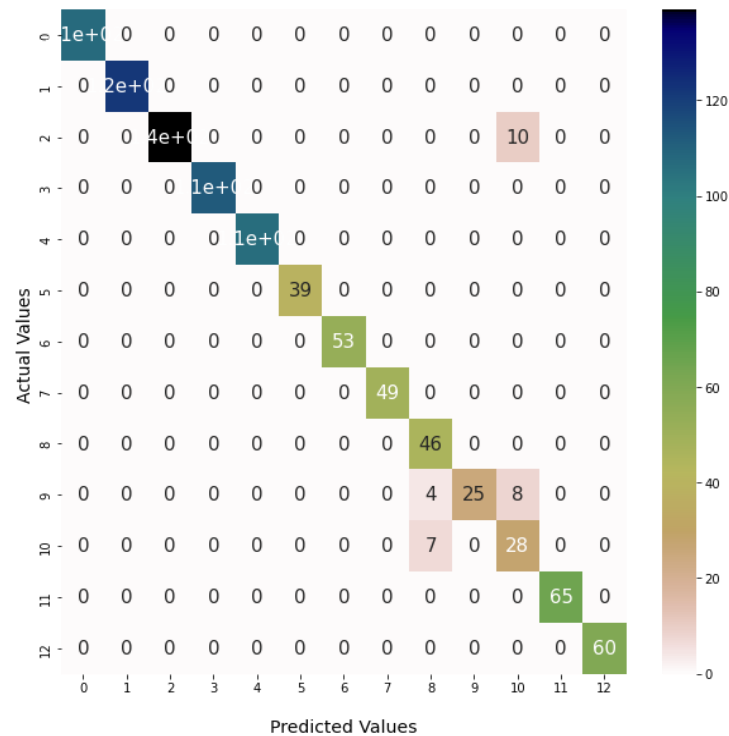
Figure 6. Confusion matrix of proposed architecture on unseen UCF dataset

6. CONCLUSION

Video surveillance become essential to prevent unexpected accidents in public places. Automatic detection of abnormal activity helps to reduce crime and improve critical care in the medical sector. The proposed study helps to improve the detection of abnormal activity from video streams. The proposed study uses convolution and sequential architecture to enhance the accuracy of anomaly detection. Simulation of the proposed has proven impact of the proposed architecture over two benchmark datasets. Simulation has also explored the compassion of the proposed architecture with state-of-the-art deep learning models for anomaly detection. The proposed combinational architecture can achieve remarkable performance with 97.55% and 91.94% accuracy for UCF and ShanghaiTech datasets respectively. Simulation has also explored the performance of different combinations of sequential models with convolution models, but BiLSTM finds promising results compared to other sequential models like GRU and LSTM. The proposed study can be extended to detect abnormal activity from live video feeds. An extension of this work can be designed with robust and advanced architecture for anomaly detection.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Dipak Ramoliya	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Amit Ganatra		✓				✓		✓		✓	✓			

C	:	**C**onceptualization	I	:	**I**nvestigation
M	:	**M**ethodology	R	:	**R**esources
So	:	**So**ftware	D	:	**D**ata Curation
Va	:	**Va**lidation	O	:	Writing - **O**riginal Draft
Fo	:	**Fo**rmal analysis	E	:	Writing - Review & **E**diting

Vi	:	**Vi**sualization
Su	:	**Su**pervision
P	:	**P**roject administration
Fu	:	**Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

ETHICAL APPROVAL

The conducted research is not related to either human or animal use.

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author, DR, on request.

REFERENCES

[1]   Y. Tan, C. Hu, K. Zhang, K. Zheng, E. A. Davis, and J. S. Park, "LSTM-based anomaly detection for non-linear dynamical system," *IEEE Access*, vol. 8, pp. 103301–103308, 2020, doi: 10.1109/access.2020.2999065.
[2]   S. Garg and S. Batra, "A novel ensembled technique for anomaly detection," *International Journal of Communication Systems*, vol. 30, no. 11, Jan. 2017, doi: 10.1002/dac.3248.
[3]   M. Said Elsayed, N.-A. Le-Khac, S. Dev, and A. D. Jurcut, "Network anomaly detection using LSTM based autoencoder," in *Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, Nov. 2020. doi: 10.1145/3416013.3426457.
[4]   H. Wei, K. Li, H. Li, Y. Lyu, and X. Hu, "Detecting video anomaly with a stacked convolutional LSTM framework," in *Computer Vision Systems*, Springer International Publishing, 2019, pp. 330–342. doi: 10.1007/978-3-030-34995-0_30.
[5]   D. Mane, "Real-time anomaly detection in video surveillance: A mathematical modeling and nonlinear analysis perspective with MobileNet and Bi-LSTM," *Communications on Applied Nonlinear Analysis*, vol. 31, no. 2s, pp. 306–319, Jun. 2024, doi: 10.52783/cana.v31.651.
[6]   E. Duman and O. A. Erdem, "Anomaly detection in videos using optical flow and convolutional autoencoder," *IEEE Access*, vol. 7, pp. 183914–183923, 2019, doi: 10.1109/access.2019.2960654.
[7]   Y. Zhu, N. M. Nayak, and A. K. 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, Feb. 2013, doi: 10.1109/jstsp.2012.2234722.
[8]   T. N. Nguyen and J. Meunier, "Anomaly detection in video sequence with appearance-motion correspondence," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 1273–1283. doi: 10.1109/iccv.2019.00136.
[9]   P. Christiansen, L. Nielsen, K. Steen, R. Jørgensen, and H. Karstoft, "DeepAnomaly: Combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field," *Sensors*, vol. 16, no. 11, p. 1904, Nov. 2016, doi: 10.3390/s16111904.
[10]   M. Sabokrou, M. Fathy, and M. Hoseini, "Video anomaly detection and localisation based on the sparsity and reconstruction error of auto-encoder," *Electronics Letters*, vol. 52, no. 13, pp. 1122–1124, Jun. 2016, doi: 10.1049/el.2016.0440.
[11]   S. W. Khan *et al.*, "Anomaly detection in traffic surveillance videos using deep learning," *Sensors*, vol. 22, no. 17, p. 6563, Aug. 2022, doi: 10.3390/s22176563.
[12]   V.-T. Le and Y.-G. Kim, "Attention-based residual autoencoder for video anomaly detection," *Applied Intelligence*, vol. 53, no. 3, pp. 3240–3254, May 2022, doi: 10.1007/s10489-022-03613-1.
[13]   D. Avola *et al.*, "A novel GAN-based anomaly detection and localization method for aerial video surveillance at low altitude," *Remote Sensing*, vol. 14, no. 16, p. 4110, Aug. 2022, doi: 10.3390/rs14164110.
[14]   M. Zaigham Zaheer, J.-H. Lee, M. Astrid, and S.-I. Lee, "Old is gold: Redefining the adversarially learned one-class classifier training paradigm," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 14171–14181. doi: 10.1109/cvpr42600.2020.01419.
[15]   Y. Tang, L. Zhao, S. Zhang, C. Gong, G. Li, and J. Yang, "Integrating prediction and reconstruction for anomaly detection," *Pattern Recognition Letters*, vol. 129, pp. 123–130, Jan. 2020, doi: 10.1016/j.patrec.2019.11.024.
[16]   J. T. Zhou, J. Du, H. Zhu, X. Peng, Y. Liu, and R. S. M. Goh, "AnomalyNet: An anomaly detection network for video surveillance," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2537–2550, Oct. 2019, doi: 10.1109/tifs.2019.2900907.
[17]   B. Wan, W. Jiang, Y. Fang, Z. Luo, and G. Ding, "Anomaly detection in video sequences: A benchmark and computational model," *IET Image Processing*, vol. 15, no. 14, pp. 3454–3465, May 2021, doi: 10.1049/ipr2.12258.

[18] K. Doshi and Y. Yilmaz, "Towards interpretable video anomaly detection," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2023, pp. 2654–2663. doi: 10.1109/wacv56688.2023.00268.
[19] D. R. Kothadiya, C. M. Bhatt, T. Saba, A. Rehman, and S. A. Bahaj, "SIGNFORMER: DeepVision transformer for sign language recognition," *IEEE Access*, vol. 11, pp. 4730–4739, 2023, doi: 10.1109/access.2022.3231130.
[20] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A.-B. Gil-González, and J. M. Corchado, "Deepsign: Sign language detection and recognition using deep learning," *Electronics*, vol. 11, no. 11, p. 1780, Jun. 2022, doi: 10.3390/electronics11111780.
[21] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," *arXiv:1710.05941*, Oct. 2017, [Online]. Available: http://arxiv.org/abs/1710.05941
[22] L. Xu, X. Ding, D. Zhao, A. X. Liu, and Z. Zhang, "A three-dimensional ResNet and transformer-based approach to anomaly detection in multivariate temporal–spatial data," *Entropy*, vol. 25, no. 2, p. 180, Jan. 2023, doi: 10.3390/e25020180.
[23] K. Yousaf and T. Nawaz, "An attention mechanism-based CNN-BiLSTM classification model for detection of inappropriate content in cartoon videos," *Multimedia Tools and Applications*, vol. 83, no. 11, pp. 31317–31340, Sep. 2023, doi: 10.1007/s11042-023-16727-6.
[24] T. Chen, R. Xu, Y. He, and X. Wang, "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," *Expert Systems with Applications*, vol. 72, pp. 221–230, Apr. 2017, doi: 10.1016/j.eswa.2016.10.065.
[25] S. Chen, X. Zhong, L. Li, W. Liu, C. Gu, and L. Zhong, "Adaptively converting auxiliary attributes and textual embedding for video captioning based on BiLSTM," *Neural Processing Letters*, vol. 52, no. 3, pp. 2353–2369, Sep. 2020, doi: 10.1007/s11063-020-10352-2.
[26] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018. doi: 10.1109/cvpr.2018.00678.
[27] W. Liu, W. Luo, D. Lian, and S. Gao, "Future frame prediction for anomaly detection - A new baseline," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 6536–6545. doi: 10.1109/cvpr.2018.00684.
[28] D. Ramoliya and A. Ganatra, "Insights of deep learning-based video anomaly detection approaches," in *Intelligent Communication Technologies and Virtual Mobile Networks*, Springer Nature Singapore, 2023, pp. 663–676. doi: 10.1007/978-981-99-1767-9_48.
[29] A. Shewalkar, D. Nyavanandi, and S. A. Ludwig, "Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 9, no. 4, pp. 235–245, Aug. 2019, doi: 10.2478/jaiscr-2019-0006.
[30] C. Zhang *et al.*, "ResNet or DenseNet? Introducing dense shortcuts to ResNet," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2021. doi: 10.1109/wacv48630.2021.00359.
[31] D. Theckedath and R. R. Sedamkar, "Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks," *SN Computer Science*, vol. 1, no. 2, Mar. 2020, doi: 10.1007/s42979-020-0114-9.

BIOGRAPHIES OF AUTHORS

Dipak Ramoliya 🆔 📇 SC 🄲 received the B.Engg. and M.Engg. degrees in computer engineering from Gujarat Technological University, Gujarat, in 2013 and 2015 respectively. Currently, he is an assistant professor at the Department of Computer Science and Engineering, Charotar University of Science and Technology, Changa, Gujarat. His research interests include machine learning, deep learning, cloud computing and high-performance computing. He has expertise in database systems and database designs. He has worked on database designs of various ERPs and applications. He can be contacted at email: dipakramoliya.ce@charusat.ac.in.

Amit Ganatra 🆔 📇 SC 🄲 received a Ph.D. degree in computer engineering from Kadi Sarva Vishwa Vidyalaya (KSV), Gandhinagar, Gujarat, India. He has been serving as the Provost of Parul University, Vadodara since August 2022. Before that, he served as the Dean, Faculty of Technology-CHARUSAT, Gujarat. His areas of interest include artificial intelligence, machine learning and deep learning. He is an active member of IEEE (Sr. Member), ACM and CSI professional society chapters. He has supervised (guided) more than 150+ industry projects at the graduate level and 100+ dissertations at the graduate level. 15 Ph.D. Scholars have completed their Ph.D. under his supervision and currently, he is supervising (guiding) 08 research scholars at the Ph.D. level and 02 scholars at the Postdoc level. Email: ganatraamitp@gmail.com.