Suspicious Activity Classification in Classrooms using Deep Learning

Neha Gupta
Computer Science & Engineering Department, IFTM University, India | Computer Science & Engineering Department, Moradabad Institute of Technology, India
discoverneha@gmail.com (corresponding author)

Bharat Bhushan Agarwal
Computer Science & Engineering Department, School of Computer Science and Applications, IFTM University, India
bharat_agarwal@iftmuniversity.ac.in

Received: 26 July 2023 | Revised: 10 August 2023 and 8 October 2023| Accepted: 19 August 2023

 DOI: https://doi.org/10.48084/etasr.6228

ABSTRACT

Video processing is attracting the attention of both research and industry. The existence of intelligent surveillance cameras with high processing power has paved the way for designing intelligent visual surveillance systems. Along with analyzing video for information recovery, it is nowadays used to analyze live surveillance video to detect activities. These systems are implemented in real time. The proposed work's goal is to create a method that can examine and discover suspicious behaviors in the lecture room environment. Video analytics offers the most efficient answer because it enables pointing an occasion and retrieves applicable statistics from the video recorded. The method aims to identify suspicious activities like fighting, sleeping, looking elsewhere, eating, etc. that the students might be doing. The proposed method involves breaking a video input into frames and converting it into image data because the model has been trained on images collected from the internet. Several models were tested and experimented with, including efficientnet_b2, spnasnet_100, efficientnet_b3, and mobilenetv3_large_100. Parameters such as the learning rate were optimized to find out the best method and create a system with the best results.

Keywords-classroom surveillance;suspicious activity;video processing;anomalous activity detection

I. INTRODUCTION

Surveillance is the observation of changing behavior, activities, or other information, usually of individuals, in order to influence, control, direct, or protect them. Visual surveillance plays a pivotal role in today's world. Governments use it to gather information, prohibit crime, protect a process, person, group, or object, or solve crime cases [1]. A useful video surveillance system depends on detecting suspicious activities [2]. The detection of anomalies in human behavior using such systems can provide clues and prevent security breaches.

The solutions proposed for suspicious classroom activity detection are not in abundance, however more work has been done on the related topic of detecting cheating and malpractices in an examination hall [3]. Automated video surveillance provides an optimal solution by monitoring students and identifying suspicious events. Monitoring an exam room is a very challenging task in terms of manpower. Therefore, a system needs to be developed that automatically detects suspicious activity and minimizes such bad practices in the exam room with as small error probability as possible. The suspicious activities in a classroom can include the use of mobile phones, playing games, sleeping, fighting, passing hand signs, or eating. A video surveillance system that could monitor the students in the classroom and classify their behavior as suspicious or not based on the activities stated above is the aim of the current work.

The recent trend of automation has an impact on the field of video analysis [4]. Video analytics can be used for a variety of applications such as motion detection, human activity prediction, person identification, abnormal activity detection, vehicle counting, etc. In this area, the two factors used for personal identification are face and gait recognition. Between these two techniques, face recognition is more versatile for automatically identifying people through surveillance videos. Face recognition can be used to predict the orientation of a person's head, which in turn helps predict a person's behavior. Motion detection along with face recognition are very useful in many applications such as verification of an individual, identification of an individual, and detection of the presence or absence of an individual at a specific location and time. Advanced technologies such as artificial intelligence, machine learning, and deep learning are now used into such systems. Human behavior is unpredictable and it is very difficult to find whether one's behavior is suspicious or normal. Monitoring is often performed through consecutive frames which are extracted from the video. These frames are fed into the machine-learning model to classify the activity as suspicious or not.

We focused on detecting suspicious activities during an offline academic examination in an exam hall because suspicious activities during normal classroom lectures are more subjective [5]. Therefore, the proposed method will detect malpractices happening in an exam hall which might include cheating through various ways like whispering, passing gestures, talking, looking back, etc. This system will serve as a useful surveillance system for educational institutions. This paper aims to analyze real-time video from examination halls and identify whether a particular person's activity is suspicious or not. The system can monitor and analyze the activities of students in an examination hall and can quickly alert the educational institution on the occurrence of any such malpractices.

Nowadays, having human invigilators in exam rooms is a task that is sometimes inconvenient. The proposed system has the ability to automatically monitor an exam room. It offers a high accuracy rate and minimizes computing time.

II. LITERATURE OVERVIEW

An extensive amount of research has been done on human activity detection and video surveillance. The general human activity detection framework includes several stages such as motion detection, background and foreground modeling and segmentation, object classification and tracking, and finally behavior and activity detection and person identification. Features such as eyes, lip and hand movements are recognized using facial recognition techniques [9]. Manifold learning maps high-dimensional to low-dimensional feature spaces. Although manifold learning is easy to implement for classification procedures, the deep learning approaches show superior performance due to the automatic feature learning. In some papers, the proposed methodologies include dividing the system into modules like facial gesture recognition or hand sign recognition and handling the modules with separate types of algorithms [11]. Thus, keeping those modules independent as features, the mainstay methods have turned out to be something like filter approaches involving feature scoring which includes PCA. But PCA is sensitive to scale and hence it should be applied only to data that have approximately the same size. It may lose important information for discrimination between different classes.

III. METHODOLOGY

In this section, dataset capturing, preprocessing, model building, and parameter tuning [15] are discussed along with the novelty of the current work.

Dataset generation was conducted by web scraping. Web scraping is an automatic method to obtain large amounts of data from websites. Most of these data are unstructured data in HTML format which are converted into structured data in a spreadsheet or a database. There are many ways to perform web scraping to obtain data from websites. These include using online services, APIs, or web scraping from scratch. Many large websites, like Google, Twitter, Facebook, and Stack Overflow, have APIs that allow data access in a structured format. Web scraping requires two parts, namely the crawler and the scraper. The crawler is an artificial intelligence algorithm that browses the web to search for the particular data required by following links across the internet. The scraper, on the other hand, is a specific tool created to extract data from a website. The extracted data are assigned classes (suspicious and non-suspicious). Data labeling is the process of identifying raw data (images, text files, videos, etc.) and adding one or more meaningful and informative labels to provide context so that a machine learning model can learn from them. Data labeling is an important part of data preprocessing for ML, particularly for supervised learning, in which both input and output data are labeled for classification to provide a learning basis for future data processing.

A total of 88 images were acquired and considered and from them, 58 images were further distributed in training, validation, and testing datasets for model building. Their division in training, testing and validation subsets along with their corresponding labels can be seen in Table I.

TABLE I. DATASET

	Number of samples		
Dataset	**Training**	**Testing**	**Validation**
Suspicious	19	14	3
Non-Suspicious	13	7	2

A. Models

The model was pre-trained on the Imagenet dataset. This model typically classifies images into a thousand classes. The Imagenet dataset is a very large series of snapshots pre-annotated and designed with the help of scientists to develop computer vision algorithms. It scales all depth/width/resolution dimensions uniformly using a simple but highly efficient composite coefficient. EfficientNets [14] are a family of models that use neural architecture search and achieve much better accuracy and efficiency than previous ConvNets [13]. The state-of-the-art models can achieve up to 84.3% accuracy on ImageNet and are 8.4 times smaller and 6.1 times faster than the best existing ConvNets. EfficientNets also achieve state-of-the-art accuracy with an order of magnitude fewer parameters [8]. MobileNet [6] is tuned to cellular CPUs through a combination of Hardware Network Architecture Search (NAS), augmented by the NetAdapt algorithm and then enhanced by new architectural advances. MobileNetV3-Large is more accurate than ImageNet with less latency than MobileNetV2 [10]. In [16-27] one can review some machine learning models that show its importance and improved results.

B. The Proposed Methodology

The proposed solution includes three parts. Firstly, the dataset was synthesized. We decided to train the model on the image dataset and consequently, the input video data were converted to a suitable format, i.e. images, before feeding the model. In order to synthesize the image dataset, we scraped the web on the Google Search engine to download images for different types of relevant queries that could provide the desired result. Figure 1 shows the proposed approach.

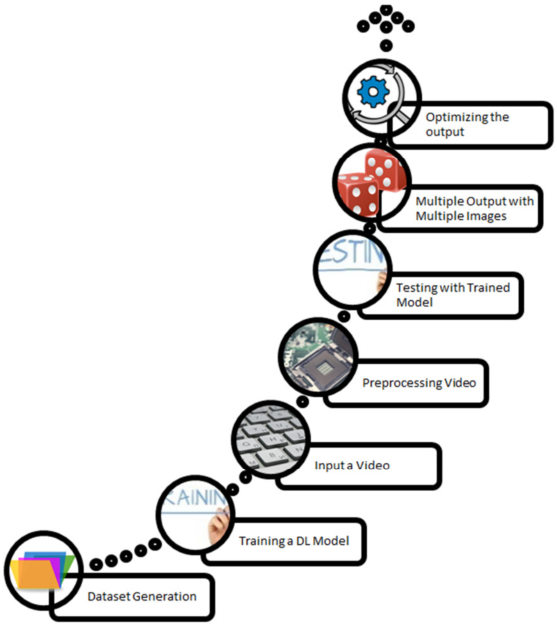

Fig. 1. Workflow of proposed approach.

There are numerous software applications for automating data scraping [12]. Bing Image Search API v7 helps users scour the web for images. Results include thumbnails, full image URLs, publishing website info, image metadata, and more. Image metadata include machine-generated captions, visually similar images, shopping and recipe sources, related image searches, etc. The second part was to train the model. In order to obtain good accuracy, we decided to use a deep learning model as a starting point of the training phase (transfer learning approach). Therefore, we used a deep learning model for images and fine-tuned it according to the dataset and the requirements. The model was trained on the data that had been synthesized and was then tested. The third part is to put the model to use. To achieve this, the video input will be converted into image data that will be fed to the model. A certain number of frames will be extracted from the videos, which are saved in the form of images. This will be done by using OpenCV library. OpenCV is a library containing predefined functions for real-time computer vision tasks like object detection, processing of images captured, etc. This will then be used for prediction by our model. The model will output the probabilities or percentages, to which category a particular belongs. The initial experiments on all the types of models have been performed by fixating on a specific initial learning rate for all the models. Thus, the initial experiments have a random learning rate [7]. After recording these values, the learning rate parameter was optimized for all models. After the optimum value was found, the training was conducted again, the experiments were re-performed and the readings and figures were again plotted and covered.

IV. EXPERIMENTS AND RESULTS

Tables II-VI show the training loss, validation accuracy and testing accuracy for the used models (Efficientnet b2, Efficientnet b3, Mobilenetv3 large 100 and Spnasnet 100) over different number of epochs and different learning rates.

Figure 2 shows the optimization of the learning rates of the different models with the help of Pytorch. Accuracy plots and a comparative study between the various considered models are presented in Figures 3 and 4. Figure 3 shows how the different losses and accuracy were distributed throughout the training process along different learning rates. Figure 4 shows training loss, validation loss, and validation accuracy for the different used models.

TABLE II. EFFICIENTNET B2, LR = 0.144

Epochs	Training loss	Validation accuracy	Testing accuracy
60	1.892	0.583	46.67
80	11.202	0.583	46.67

TABLE III. EFFICIENTNET B3, LR = 0.005

Epochs	Training loss	Validation accuracy	Testing accuracy
60	0.113	0.75	80
80	0.103	0.756	80
100	0.099	0.583	80

TABLE IV. MOBILENETV3 LARGE 100, LR = 0.238

Epochs	Training loss	Validation accuracy	Testing accuracy
60	2.881	0.833	63.33
80	2.29	0.667	63.33
100	6.739	0.583	63.33

TABLE V. MOBILENETV3 LARGE 100, LR = 3E-14

Epochs	Training loss	Validation accuracy	Testing accuracy
60	5.604	0.750	63.33
80	2.290	0.667	63.33
100	6.739	0.583	63.33

TABLE VI. SPNASNET 100, LR = 1E-8

Epochs	Training loss	Validation accuracy	Testing accuracy
60	0.620	0.667	60
80	0.147	0.667	60
100	0.140	0.667	60

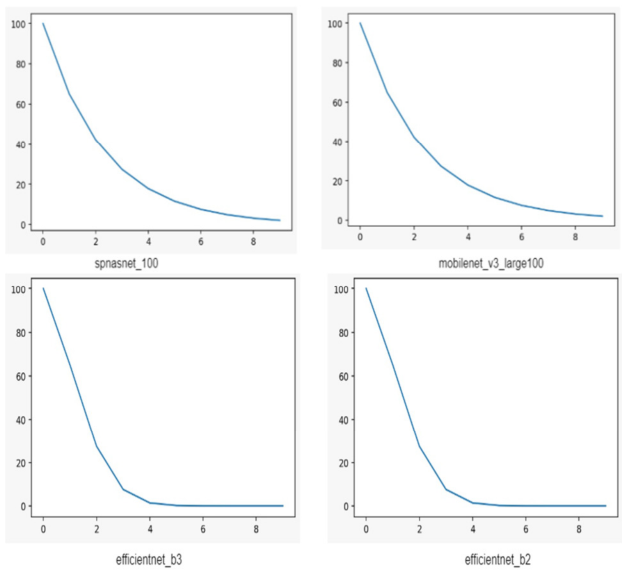

Fig. 2. Learning rate plots.

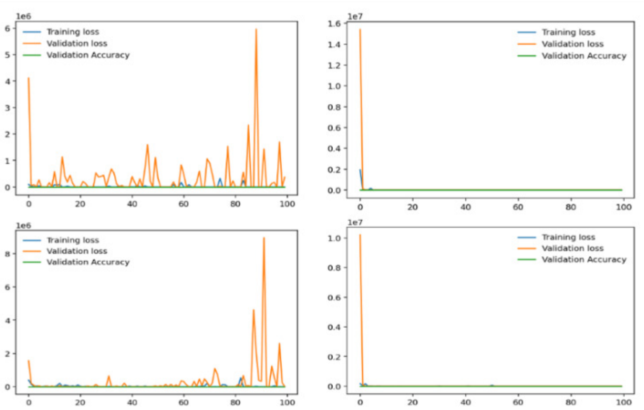

Fig. 3. Accuracy plots.

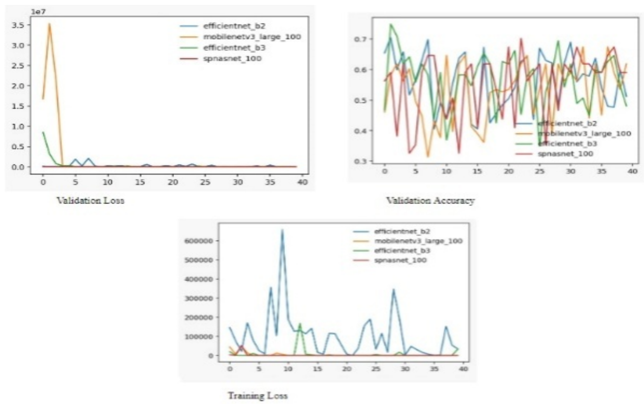

Fig. 4. Comparative study between efficientnet, mobilenetv3large and spnasnet models.

After training, Spnasnet 100 was considered for further evaluation and experimentation. Another set of 30 images were given to the model as input. Table VII shows the results and Table VIII the confusion matrix for the test dataset and the classification into suspicious and non-suspicious activity. The F1 score of 0.81 of the proposed model is sufficient for the proposed approach in limited datasets.

TABLE VII. RESULTS

True Positives	18	**Accuracy**	0.73
False Positives	8	**Recall**	1.00
True Negatives	4	**Precision**	0.69
False Negatives	0	**F1 Score**	0.81

TABLE VIII. CONFUSION MATRIX

	False	**True**
False	4	8
True	0	18

V. CONCLUSION AND FUTURE WORK

Detecting suspicious activity is an open area of research. Timely automatic detection of suspicious student activity helps the supervisors take accurate and fair action. This work includes the classification of suspicious activity using several types of models, training them on our dataset and finally using the optimized learning rate. Thus, the proposed system detects suspicious human activity carried out in an examination hall using the specified models. This method gives better results and reduces the number of false positives when compared to other approaches.

It must be noted that very good accuracy couldn't be achieved due to the insufficient amount of data. Data of this kind couldn't be found in abundance. However, this system is a good tool with fair accuracy. In the future, we aim to improve this system with the help of a better quality dataset.

REFERENCES

[1] C. Sujatha and U. Mudenagudi, "A Study on Keyframe Extraction Methods for Video Summary," in *2011 International Conference on Computational Intelligence and Communication Networks*, Jul. 2011, pp. 73–77, https://doi.org/10.1109/CICN.2011.15.

[2] T. Senthilkumar and G. Narmatha, "Suspicious Human Activity Detection in Classroom Examination," in *Computational Intelligence, Cyber Security and Computational Models*, Singapore, 2016, pp. 99–108, https://doi.org/10.1007/978-981-10-0251-9_11.

[3] G. S. Devi, G. S. Kumar, and S. Chandini, "Automated Video Surveillance System for Detection of Suspicious Activities during Academic Offline Examination," *International Journal of Computer and Information Engineering*, vol. 11, no. 12, pp. 1265–1271, Oct. 2017.

[4] A. Katiyar, A. Singh, M. Parvez, N. Rajora, L. K. Sharma, and G. Singal, "Thresholding based Smart Home Automation System using K-means," in *2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, Jan. 2023, pp. 47–52, https://doi.org/10.1109/Confluence56041.2023.10048795.

[5] M. D. Genemo, "Suspicious activity recognition for monitoring cheating in exams," *Proceedings of the Indian National Science Academy*, vol. 88, no. 1, pp. 1–10, Mar. 2022, https://doi.org/10.1007/s43538-022-00069-2.

[6] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." arXiv, Apr. 16, 2017, https://doi.org/10.48550/arXiv.1704.04861.

[7] T. Mamalis, D. Stipanovic, and P. Voulgaris, "Stochastic Learning Rate Optimization in the Stochastic Approximation and Online Learning Settings." arXiv, Mar. 21, 2022, https://doi.org/10.48550/arXiv.2110.10710.

[8] M. Shaha and M. Pawar, "Transfer Learning for Image Classification," in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, Mar. 2018, pp. 656–660, https://doi.org/10.1109/ICECA.2018.8474802.

[9] P. Deraiya, J. Pandya, and P. G. Student, "A Survey for Abnormal Activity Detection in Classroom," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 5, no. 12, pp. 20511–20516, 2007, https://doi.org/10.15680/IJIRSET.2016.0512066.

[10] G. Wang, A. El Saddik, X. Lai, G. Martinez Perez, and K.-K. R. Choo, Eds., *Smart City and Informatization: 7th International Conference, iSCI 2019, Guangzhou, China, November 12–15, 2019, Proceedings*, Springer, 2019.

[11] A. Paszke *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems*, 2019.

[12] A. Sharma *et al.*, "Heart rate and blood pressure measurement based on photoplethysmogram signal using fast Fourier transform," *Computers and Electrical Engineering*, vol. 101, Jul. 2022, Art. no. 108057, https://doi.org/10.1016/j.compeleceng.2022.108057.

[13] R. Ribani and M. Marengoni, "A Survey of Transfer Learning for Convolutional Neural Networks," in *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*, Rio de Janeiro, Brazil, Jul. 2019, pp. 47–57, https://doi.org/10.1109/SIBGRAPI-T.2019.00010.

[14] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." arXiv, Sep. 11, 2020, https://doi.org/10.48550/arXiv.1905.11946.

[15] K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, *International Conference on Machine Learning, 17-23 July 2022, Baltimore, Maryland, USA*, 2022.

[16] N. Kumar, A. Hashmi, M. Gupta, and A. Kundu, "Automatic Diagnosis of Covid-19 Related Pneumonia from CXR and CT-Scan Images," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 7993–7997, Feb. 2022, https://doi.org/10.48084/etasr.4613.

[17] N. Gupta and B. B. Agarwal, "Recognition of Suspicious Human Activity in Video Surveillance: A Review," *Engineering, Technology & Applied Science Research*, vol. 13, no. 2, pp. 10529–10534, Apr. 2023, https://doi.org/10.48084/etasr.5739.

[18] M. B. Ayed, S. Elkosantini, and M. Abid, "An Automated Surveillance System Based on Multi-Processor and GPU Architecture," *Engineering, Technology & Applied Science Research*, vol. 7, no. 6, pp. 2319–2323, Dec. 2017, https://doi.org/10.48084/etasr.1645.

[19] N. Kumar and D. Aggarwal, "LEARNING-based Focused WEB Crawler," *IETE Journal of Research*, vol. 69, no. 4, pp. 2037–2045, May 2023, https://doi.org/10.1080/03772063.2021.1885312.

[20] M. Kaur, V. Kumar, V. Yadav, D. Singh, N. Kumar, and N. N. Das, "Metaheuristic-based Deep COVID-19 Screening Model from Chest X-Ray Images," *Journal of Healthcare Engineering*, vol. 2021, Mar. 2021, Art. no. e8829829, https://doi.org/10.1155/2021/8829829.

[21] N. Kumar, N. Narayan Das, D. Gupta, K. Gupta, and J. Bindra, "Efficient Automated Disease Diagnosis Using Machine Learning Models," *Journal of Healthcare Engineering*, vol. 2021, May 2021, Art. no. e9983652, https://doi.org/10.1155/2021/9983652.

[22] N. Kumar, M. Gupta, D. Gupta, and S. Tiwari, "Novel deep transfer learning model for COVID-19 patient detection using X-ray chest images," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 1, pp. 469–478, Jan. 2023, https://doi.org/10.1007/s12652-021-03306-6.

[23] M. Gupta, N. Kumar, B. K. Singh, and N. Gupta, "NSGA-III-Based Deep-Learning Model for Biomedical Search Engines," *Mathematical Problems in Engineering*, vol. 2021, May 2021, Art. no. e9935862, https://doi.org/10.1155/2021/9935862.

[24] N. Kumar, M. Gupta, D. Sharma, and I. Ofori, "Technical Job Recommendation System Using APIs and Web Crawling," *Computational Intelligence and Neuroscience*, vol. 2022, Jun. 2022, Art. no. e7797548, https://doi.org/10.1155/2022/7797548.

[25] M. Gupta, N. Kumar, N. Gupta, and A. Zaguia, "Fusion of multimodality biomedical images using deep neural networks," *Soft Computing*, vol. 26, no. 16, pp. 8025–8036, Aug. 2022, https://doi.org/10.1007/s00500-022-07047-2.

[26] A. Hashmi et al., "Contrast Enhancement in Mammograms Using Convolution Neural Networks for Edge Computing Systems," *Scientific Programming*, vol. 2022, Apr. 2022, Art. no. e1882464, https://doi.org/10.1155/2022/1882464.

[27] K. R. Kodepogu *et al.*, "A Novel Deep Convolutional Neural Network for Diagnosis of Skin Disease," *Traitement du Signal*, vol. 39, no. 5, pp. 1873–1877, Nov. 2022, https://doi.org/10.18280/ts.390548.