Nigerian Journal of Science and Environment 2024 Volume 22 (2) 177 – 189 ISSN: 3043 – 4440
https://doi.org/10.61448/njse2222412

A HYBRID ASSAULT DETECTION SYSTEM USING RANDOM FOREST ENABLED XGBOOST-LIGHTGBM TECHNIQUE

***[1]Ohwosoro, I.D., [2]Edje, A.E., [3]Atonuje, E. and [4]Ogeh, C.O.**

Department of Computer Science, Delta State University, Delta State PMB 01, Nigeria.

ABSTRACT

This article presents the development of an assault identification system using face recognition in a closed location, by employing a machine learning-based computer vision approach. The proposed model combines algorithms such as Random forest, XGBoost and LightGBM techniques. The objective is to accurately identify and classify instances of assault in real-time based on facial recognition. The proposed approach utilizes machine learning algorithms to analyze facial features and patterns associated with assault activities. By leveraging on a hybrid model, the system can be integrated into closed locations such as schools, workplaces, or public venues to enhance security measures and promptly respond to potential threats. The findings of this research contribute to the field of computer vision-based assault identification systems, in addressing security challenges. Further advancements of the proposed hybrid model can lead higher performance levels in various real-world scenarios and enhancing public safety and security. The system's performance was evaluated using various metrics, including precision, recall, F1 score, accuracy, and ROC score. The results shows that the proposed system outperformed the existing system with its identified weakness and limitations of: Limited Robustness in Handling Complex Variations, Inability to Handle High-Dimensional Data, Limited Scalability, e.tc. The hybrid model achieved impressive results, with a precision of 98%, recall of 98%, F1 score of 97.7%, accuracy of 97.5%, and ROC score of 97.4%. The above findings demonstrated the effectiveness and robustness of the developed system in accurately detecting and recognizing assault instances within a closed location.

Keywords: Machine Learning, Computer Vision, Face Recognition, Closed Location and Face Detection.

INTRODUCTION

Assault is an illegal act of causing physical harm or unanticipated physical contact to another person. This is more dangerous in a closed location. Assault can lead to bodily harm, reduction of self-esteemed, low productivity or death in a place of work. The existing system relied on traditional facial recognition methods that often struggled to handle complex variations in facial features, lighting conditions, and with changes in posture, e.t.c. The requirement for more investigation to ascertain the most effective machine learning algorithms or combinations of algorithms for identifying assaulting individuals and large number of cameras and computational resources to enable real-time identification and monitoring of assaulting individuals in a closed location, an improvement to this problem was a major reason for this proposed system.

Sukkar *et al.* (2012) created a computer vision-based surveillance system that uses motion trajectory analysis to identify violent incidents. However, facial recognition was not a part of this system. Parkhi et al. (2015), Taigman *et al.* (2014), Schroff *et al.* (2015), Sun *et al.* (2014), and Huang *et al.* (2017) reported that different deep learning-based methods improved face recognition tasks greatly and offered reliable solutions under different settings.

According to Farfade *et al* (2015), he proposed a method of face detection based on deep learning, which called Deep Dense Face Detector (DDFD).

The method does not require pose/landmark annotation and is able to detect faces in a wide range of orientation using a single model.

Similarly, *Filali et al.* (2018) provided a comparative study between four methods (Haar–AdaBoost, LBP–AdaBoost, GF-SVM, GFNN) for face detection. These techniques vary according to the way in which they extract the data and the adopted learning algorithms.

Ren *et al.* (2017) have presented a method for real time detection and tracking of the human face. The method combines the Convolution Neural Network detection and the Kalman filter tracking. Convolution Neural Network is used to detect the face in the video, which is more accurate than traditional detection method.

The study of developing an assault identification system through face recognition in a closed location using a machine learning-based computer vision approach is rooted in the increasing need for advanced security measures.

Traditional security measures, such as CCTV cameras and security personnel, have often been found lacking in their ability to prevent assaults as they occur. These systems are typically reactive, responding to incidents after they have already taken place. This reactive nature of traditional security measures has led to the exploration of more proactive solutions, such as the use of machine learning and computer vision for assault identification.

Face recognition technology, which identifies or verifies a person's identity using their facial features, plays a pivotal role in this system. It has been used in various applications, from smartphone security to criminal investigations. In the context of an assault identification system, face recognition can be used to identify known threats and alert security personnel, allowing for timely intervention before an assault takes place. However, the development of such a system also raises important ethical and privacy concerns. The potential for misuse of face recognition technology and the implications for personal privacy must be carefully considered during the development and implementation of the system.

Face recognition output system consists of two approaches: identification and verification (authentication). Face identification is a one-to-many mapping where a face is checked against a database of known faces, whilst, face verification is a one-to-one mapping, where a face is checked with an identity in the database.

According to a study conducted by Nusir *et al.* (2016), it was suggested to use supervised machine learning algorithms as the basis for an automated assault detection method. A Support Vector Machine (SVM) algorithm was trained by the authors using pre-defined features such as skin color, motion history, and face detection after they collected video material. The algorithm's accuracy rate in identifying attacks was 92.5%, while its accuracy rate in identifying regular encounters was 97%.

In a similar vein, Zhang *et al.* (2018) investigated the use of a vision-based method for the identification of physical disputes in a nursing home. The authors state that a Convolutional Neural Network (CNN) was utilized to identify activities and facial expressions during occurrences. The findings demonstrated an accuracy of 89.8% in identifying physical disputes.

In a different study, Chen *et al.* (2019) suggested a machine learning technique called Long hort-Term Memory (LSTM) to create a context-aware assault detection system. In light of their findings, scientists employed LSTM to identify sequential patterns of facial emotions and body movements in smaller pieces of video material. The accuracy rate attained by the method was 95.5%. Deep learning was used by Rony *et al.* (2019) to identify illegal facial expressions. They found that using the CK+ dataset, the system's accuracy rate for identifying criminal facial expressions was 90.4%. In addition, a deep neural network model for identifying violent episodes in airport CCTV data was presented by Cho *et al.* (2019). Using a convolutional neural network, they achieved an accuracy rate of 92.04% on the AVA dataset, according to their research.

Using convolutional neural networks, *Kazi et al.* (2020) created a framework for identifying and categorizing hostile conduct in a crowd. On the UCF-Crime dataset, the authors' accuracy rate was 88.15%, based on their research. *Zhang et al.* (2020) presented a facial recognition system for real-time assault identification in enclosed spaces in a different study. Their research indicates that the system achieved great detection and identification rates by using deep convolutional neural networks for face detection and recognition.

Similarly, Kim and colleagues (2021) created a system based on deep learning to identify instances of fighting in a subway station. Their paper states that using the SKT-C dataset, their algorithm attained an accuracy rate of 94.45%. S. Das *et al.,* (2021) provide methods for detecting violence. Three kinds are distinguished: visual-based methods employing SVM classifier. Progressive methodology that is grounded in sound principles. It is reliant on Hidden Markov models (HMM) and Gaussian mixture models. Additionally, the hybrid method uses the k-Nearest Neighbor classifier to determine whether or not the supplied sequence is violent. Its accuracy on the KTH dataset is 88.19%.

Table 1: Differences between face detection and face recognition systems.

Face Detection	Face Recognition
1. It is a form or subclass of face recognition.	It consists of face identification and face verification.
2. The initial step to face recognition.	To be able to recognize a face.
3. Detects face and crop image to be pre-processed.	Checks the database whether face identity is present.

Table 2: Summary of identification and verification approach explained.

Description	Identification	Verification
Known as	1:N matching problem	1:1 matching problem
Narration	The unknown face is compared with the captured pictures of faces in the database	The identified queried face is compared with the captured faces in the databases.
Nature of task/result	•closed-set – a person is known to be in the database. • open-set – a person is unknown to be in the database.	•confirmed. •rejected.

Various research works have been carried out in the development of an assault identification system using face recognition in closed locations. This section presents a comprehensive review of the works carried out in this field by various Authors.

According to a study conducted by Kim *et al*. (2021) created a system based on deep learning to identify instances of fighting in a subway station. Their paper states that using the SKT-C dataset, their algorithm attained an accuracy rate of 94.45%. With a high and enhanced predictive and detective performance record, the research works of Vijeikis *et al*. (2022) with an accuracy of 82% and Marius Baba *et a.,* (2019) with accuracy of 86.93% were further examined and updated in this proposed system

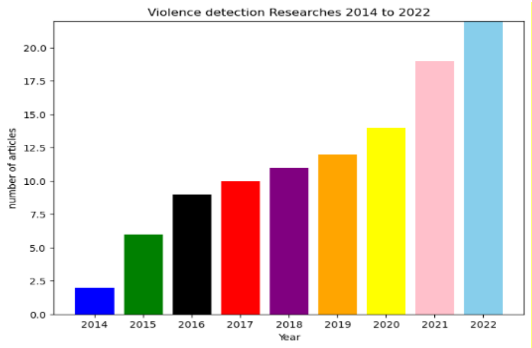

Figure 1: Violence detection researches

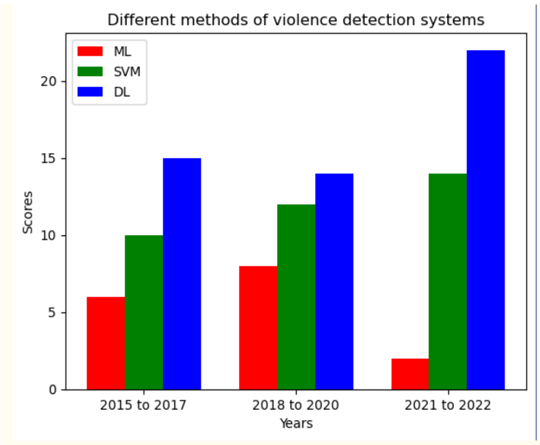

Figure 2: Various methods of violence detection systems

MATERIALS AND METHODS

In this chapter, the methodology for developing an Assault recognition System of a random forest-enabled Xboost-lightgbm through Face Detection in a Closed Location was discussed. The methodology adopted a hybrid model that combined both The Random Forest and boosting-based algorithms to enhance the system's performance and accuracy. Specifically, the hybrid model consisted of traditional algorithms such as (Random Forest, along with boosting algorithms like XGBoost and LightGBM: To begin, the analysis of the existing algorithms were utilized as a foundational component of the hybrid model.

The Random Forest, an ensemble learning method, was incorporated into the hybrid model. This algorithm constructs multiple decision trees and combines their predictions to make a final decision. By training each decision tree on random subsets of features and data samples, Random Forest reduces the risk of overfitting and improves generalization. Its robustness and capability to handle high-dimensional data made it an important component of the hybrid model.

Random Forest RF is a technology that represents a set of an ensemble learning methods for random classification that functioning by making decisions using a multitude of trees votes and predicting the features of data as follow: classification results from each tree are collected for the input image, after that, the majority voting is gathered to give the resulting class label. Random forest (RF) is artificial intelligence technique and the strong modern method to a classification of data and modeling. RF has been applied to compare extracted template (vectors features) from both training and testing stage to match the corresponding person. Suitable features vector that able to characterize, as much

Nigerian Journal of Science and Environment 2024 Volume 22 (2) 177 – 189 ISSN: 3043 – 4440
https://doi.org/10.61448/njse2222412

as possible is highly recommended for the classification process.

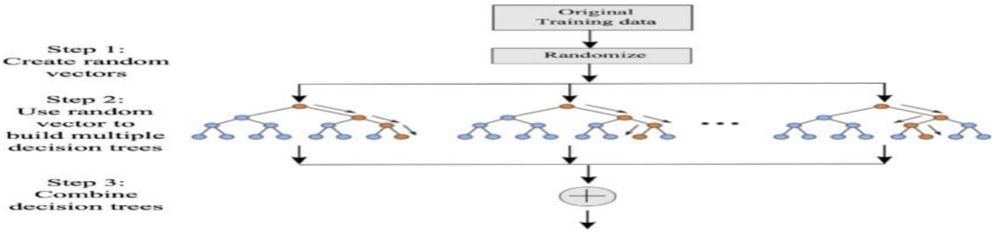

Figure 3: Shows Random Forest algorithm block diagram

In addition to the Random Forest algorithm, boosting algorithms were integrated into the hybrid model to further enhance its performance XGBoost, an eXtreme Gradient Boosting algorithm, was employed for its powerful learning ability and the capacity to capture complex relationships within the data by sequentially combining weak learners, XGBoost improved the system's predictive performance. and its capability to capture non-linear relationships and intricate data patterns.

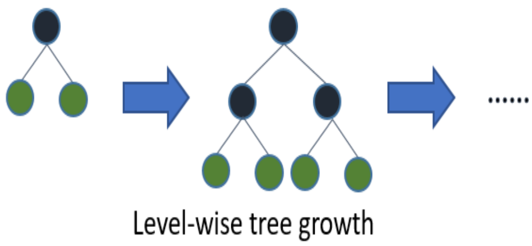

Figure 4: Shows Xgboost algorithm block diagra

Another boosting algorithm, is LightGBM, was also included in the hybrid model. LightGBM is known for its efficiency and ability to handle large-scale datasets. By adopting a histogram-based approach for feature discretization, LightGBM reduced computational complexity while maintaining high accuracy. Its efficiency and scalability made it a valuable addition to the hybrid model. By combining algorithms of (Random Forest) with boosting algorithms (XGBoost and LightGBM), the hybrid model leveraged the strengths of both approaches. The Random Forest provides a solid foundation for identifying patterns and capturing complex relationships, while the boosting algorithms enhanced the model's performance and predictive capabilities.

LightGBM carries out leaf-wise (vertical) growth that results in more loss reduction and in turn higher accuracy while being faster. But this may also result in overfitting on the training data which could be handled using the max-depth parameter that specifies where the splitting would occur. Hence, XGBoost is capable of building more robust models than LightGBM.

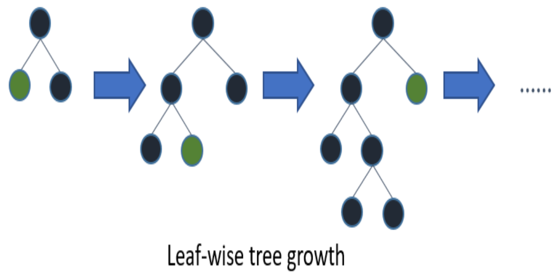

Figure 5: Shows Lightgbm algorithm block diagram.

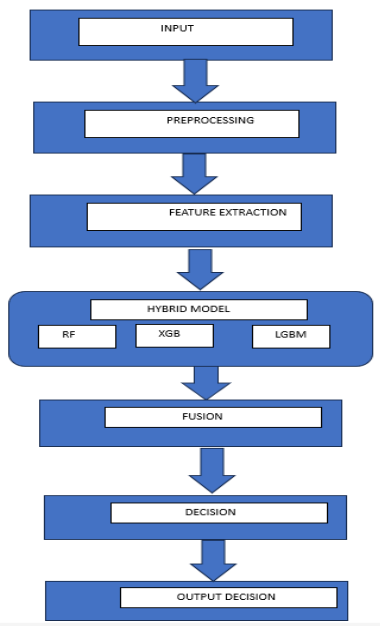

Figure 6: High Level Model of the Proposed System architecture.

Method of Data Collection.

Data Preprocessing: This is crucial steps taken in preparing the dataset for training the hybrid model. The following steps were undertaken:

i. **Data Collection**: The dataset used in this study was acquired from the esteemed Kaggle community, a platform well-known among data scientists and machine learning enthusiasts. This dataset encompassed a diverse array of images that were diligently classified into two specific categories: "assaulted/violence" and "non-assaulted/non-violence" images.

Table 3: Summary of Dataset collection.

Items	No of Violence Images	No of Non-violence Images	Tot No of trained datasets
No of Images	5,832	5,531	11,063
(% of Images)	52.7	47.3	100

ii. Image Resizing and Standardization: The photographs were downsized to a standard size to guarantee uniformity and enable effective processing.

iii. Feature Extraction: Images must be transformed into numerical feature vectors to enable machine learning algorithms to process them. Using methods like flattening the pixel values or applying more sophisticated feature extraction techniques. With the use of these methods, pertinent features that contained discriminative information could be extracted from the pictures.

iv. Label Encoding: The dataset featured labels or annotations designating whether a picture depicted an attack or a non-assault situation. To make training and evaluating the hybrid model easier, these labels were encoded as numerical values, such as 0 for non-assault and 1 for assault.

v. Train-Test Split: In order to assess the hybrid model's performance, the dataset was divided into sections for testing and training. The testing set was used to evaluate the model's performance on untested data, while the training set was used to train the model. Typical split ratios of the ratio of 80:20 (i.e., 80% and 20% for dataset training and testing respectively) was utilized. By following these data preprocessing and transformation steps, the dataset of assault and non-assault images was prepared for training the hybrid model. The resized and standardized images were converted into feature vectors, and the corresponding labels were encoded for supervised learning.

Development Process

Designing a project for violence detection in a closed location involves using technologies like computer vision, machine learning, and possibly sensor data to identify and respond to violent behavior. Here is an overview of how the proposed system's objective: which was to develop a system that can detect violent behavior in a closed location such as a school, workplace, or public space, and alert the authorities or security personnel for action was implemented:

i. **Data Collection**: A dataset of videos or images were collected from a closed location where violence may occur. This dataset should contain examples of both violent and non-violent behaviours.
ii. **Data Pre-processing**: Pre-process the data by extracting relevant features, resizing images, and converting videos into frames for analysis.
iii. **Violence Detection Model:** Train a machine learning model using techniques like object detection, action recognition, or anomaly detection to identify violent behavior in the videos or images.
iv. **Real-time Detection**: Implement the model to perform real-time violence detection in a closed location using live video feeds or surveillance cameras.
v. **Alert System**: An alert system was developed to triggers notifications to security personnel or authorities when violent behaviour is detected.
vi. **Integration with Security Systems**: The violence detection system will be iintegratedd with existing security systems to automate responses like locking down certain areas, sounding alarms, or contacting emergency services, e.tc.
vii. **Testing and Evaluation**: Test the system with different scenarios and evaluate its performance in terms of detection accuracy, false positives and false negatives.
viii. **Deployment and Monitoring:** Deploy the system in the closed location and monitor its performance over time. The proposed model need to be ccontinuously updated with new data to improve its accuracy and its performance.
ix. **Technologies to Consider:** Computer Vision: OpenCV, TensorFlow, PyTorch;

Machine Learning/Deep Learning models (CNNs, RNNs), Scikit-learn and Alert System: Email notifications, SMS alerts.

Programming Language Use.

Python was used for the creation of this model because to its adaptability and superior usefulness when handling mathematical, statistical, and scientific processes. The Jupyter Notebook IDE was used to write our Python source code. The notebook itself is accessed using a web browser and can be hosted on either a remote server or your own computer.

Figure 3.9: An overview of Jupyter Notebook

Performance Evaluation Metrics

The evaluation of the Assault Identification System through Face Recognition in a Closed Location using a Hybrid Model (XGBoost, LightGBM and Random Forest) involves the use of various metrics to assess its performance and effectiveness. The following evaluation metrics were employed:

Precision: A metric that measures the accuracy of positive predictions made by the system. It is calculated as the ratio of true positives (correctly identified assault incidents) to the sum of true positives and false positives (instances incorrectly identified as assault incidents). A higher precision value indicates a lower rate of false positives, signifying the system's ability to accurately identify assault incidents without falsely classifying non-assault scenarios.

$$Precision = \frac{TP}{TP + FP}$$

Accuracy: Rate of data instances correctly classified by the model

$$Accuracy = \frac{TPR + TNR}{TPR + TNR + FPR + FNR}$$

Recall (Sensitivity or True Positive Rate): Recall measures the ability of the system to identify positive instances correctly. It is calculated as the ratio of true positives to the sum of true positives and false negatives. A higher recall value indicates a lower rate of false negatives, indicating the system's capability to accurately identify assault incidents and minimize instances of missed detections.

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: The F1-score is a metric that combines precision and recall into a single value, providing a balanced assessment of the system's performance. It is calculated as the harmonic mean of precision and recall. The F1-score considers both false positives and false negatives, providing a comprehensive evaluation of the system's accuracy and completeness in identifying assault incidents.

$$F1 - Score(F) = \frac{2PR}{P + R}$$

Nigerian Journal of Science and Environment 2024 Volume 22 (2) 177 – 189 ISSN: 3043 – 4440
https://doi.org/10.61448/njse2222412

Confusion Matrix:

The confusion matrix is a table that presents the system's performance by summarizing the predictions against the actual labels. It consists of four elements: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The confusion matrix provides an overall view of the system's performance and allows for the calculation of various metrics, including Precision, Recall, and F1-Score

Table 4: Confusion Matrix table

Description	True Positive	True Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN

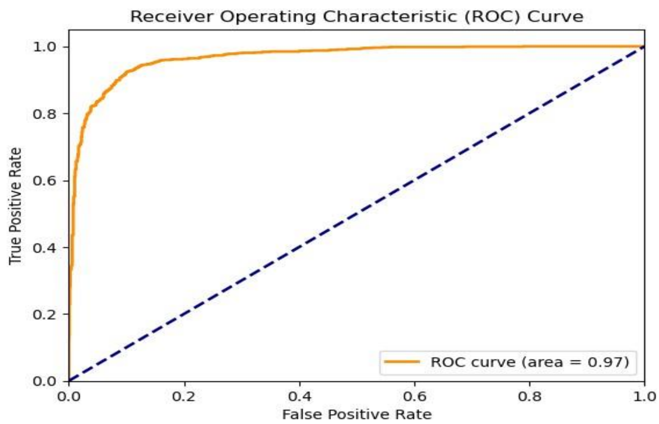

Figure 8.: ROC curve for the proposed system

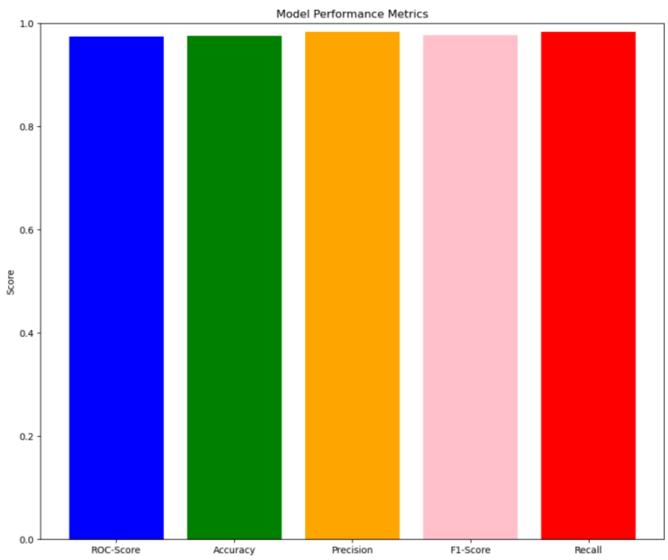

Figure 9.: Performance Metrics of the model

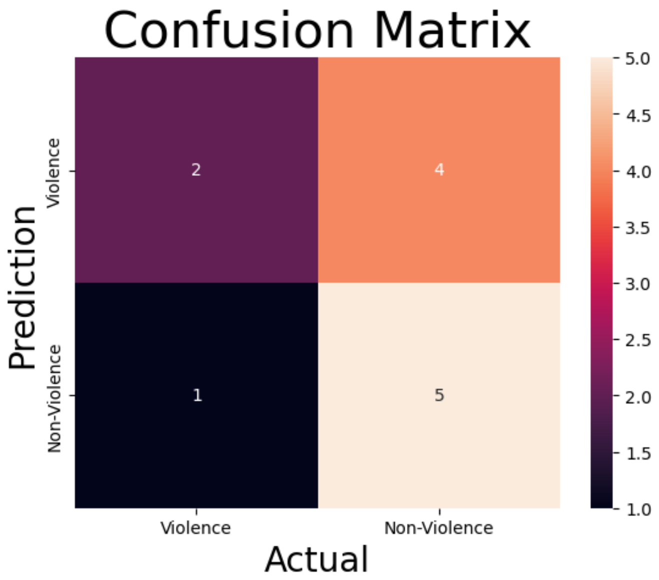

Figure 10: An overview of Jupyter Notebook

RESULT AND DISCUSSION

The findings from the proposed system has been effectively summarized and organized in the table presented below:

Table 5: Results of the proposed system.

Method adopted	Precision	Recall	F1 Score	Accuracy	ROC Score
Hybrid model	98%	98%	97.7%	97.5%	97.4%

This hybrid model demonstrated excellent performance in identifying assaults using face recognition system. It achieved a precision of 98%, recall of 98%, F1 score of 97.7%, and an accuracy of 97.5%. These results indicate the effectiveness of the hybrid approach in accurately detecting assault incidents in closed locations. This hybrid model outperformed the existing system, table 6 shows the comparison of the findings with other exisitng model's findings from all the reviewed literatures.

Findings of other existing models and the proposed models are summarized in the table below

Table 6: Summary of some related studies in violent detection models and the proposed model

S/N	Reference	Detection Methods	Feature Extraction	Strength/Accuracy
1	Gao *et al.*	SVM and AdaBoost	Oriented violent flows (OViF)	Performance of the proposed OViF and LTP was able to achieve a more satisfactory accuracy of 87.50% and

				88.00% for Hockey Fights and Violent Flow.
2	Zhou *et al.*	SVM	Low-level features are the local histogram of oriented gradient (LHOG), bag-of-words (BoW), local histogram of optical flow (LHOF) descriptor	The proposed features extraction showed an effective detection model in automatic violent behaviors in comparison with the state-of-the-art algorithms.
3	Kazi *et al*	CNN	Classification of perceived aggressive behaviours.	Effective use of technology in aggressive human behaviour detection. Accuracy of 88.15%.
4	Long, B *et al*	K-NN Algorithm	An application that could autonomously detect a bullying event, without the knowledge of the bullies.	Accuracy of 84.0%
5	O. Sharma *et al*	PCA	Analysis of violence and non-violence detection.	Accuracy of 89.5%
6	Isaac, O *et al*	Xgboost-Lightgbm and Random Forest technique	The proposed approach utilizes machine learning algorithms to analyze facial features and patterns associated with assault activities by leveraging on hybrid techniques.	Accuracy of 97.5%

AKNOWLEDGEMENTS

The research leading to these results were as a result of the immense support from my Project Supervisor who doubled as the HOD of the department and the MSc Projector Coordinator for her invaluable guide and has received the attention of the 7th Delta State University Faculty of Science International conference held in November, 2023 for publication.

REFERENCES

Abbass, M. A.B and Kang, H. S. (2023). Violence Detection Enhancement by Involving Convolutional Block Attention Modules Into Various Deep Learning Architectures: Comprehensive Case Study for UBI-Fights Dataset. *IEEE Access, 11, 37096-37107.*

Anwarul, S and Dahiya, S. (2020). A Comprehensive Review on Face Recognition Methods and Factors Affecting Facial Recognition Accuracy. *In: Singh, P., Kar, A., Singh, Y., Kolekar, M., Tanwar, S. (eds) Proceedings of ICRIC 2019 . Lecture Notes in Electrical Engineering, vol 597.*

Bhowmik, MK; Saha, P and Singha, A. (2019). Face Recognition Techniques: A Comprehensive Review.

Buolamwini, J and Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 81-91.*

Chen, L., Xu, C., Liu, J., Yuan, Z and Zhang, B. (2019). Context-aware smart surveillance for real-time automated assault detection. *IEEE Journal on Selected Areas in Communications, 37(2), 415-426.*

Dehghani, M and Karami, A. (2020). Ethical considerations of artificial intelligence for face recognition: a systematic literature review. *Ethics and Information Technology, 22(1), 65-79.*

Durães, D., Marcondes, F.S., Gonçalves, F., Fonseca, J., Machado, J and Novais, P. (2021). Detection violent behaviors: a survey. In: Novais, P., Vercelli, G., Larriba-Pey, J.L., Herrera, F., Chamoso, P. *(eds.) ISAmI 2020. AISC, vol. 1239, pp. 106–116. Springer, Cham.*

Durães, D., Santos, F., Marcondes, F.S., Lange and S., Machado, J (2021). Comparison of transfer learning behaviour in violence detection with different public datasets. In: Marreiros, G., Melo, F.S., Lau, N., Lopes Cardoso, H., Reis,

on Intelligent Computing and Information Systems (ICICIS'19), Cairo, pp. 79-84.

Mann, B.P. (2021)“ Real-Time Violence Detection Using CNN-LSTM.

Muhammad Sajjad, Mansoor Nasir, Khan Muhammad, Siraj Khan, Zahoor Jan, Arun Mukherjee, S; Saha, S; Lahiri, S; Das, A; Bhunia, AK; Konwer, A and Chakraborty, A (2017) Convolutional neural network based Face detection. *In: Proceeding of 1st international conference on electronics, materials engineering and nano-technology, pp 1–5.*

Nusir, W. N., Tamimi, A. A and Jararweh, Y (2016). Automated assault detection using supervised machine learning algorithms. *Journal*

L.P. *(eds.) EPIA 2021. LNCS (LNAI), vol. 12981, pp. 290–298. Springer, Cham.*

Gujjar, J.P., Kumar, H.P and Chiplunkar, N.N (2021). Image classification and prediction using transfer learning in colab notebook. *Global Transit. Proc. 2(2), 382–385.*

Guo, W., Zhu, C., Yang, K., Cheng, S., and Li, S. (2021). A modified-pyramid of histograms of oriented gradients (M-PHOG) method for violent confrontation detection. *Journal of Ambient Intelligence and Humanized Computing, 12(9), 8343-8357.*

Hershey, S; Chaudhuri, S; Ellis, DPW; Gemmeke, JF; Jansen, A; Moore, RC; Plakal, M and Platt, D (2017). CNN architectures for large-scale audio classification. *In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 131–135. IEEE.*https://www.kaggle.com/datasets/karandeep98/real-life-violence-and-nonviolence-data

Kong, Y and Fu, Y (2022). Human action recognition and prediction: a survey. *Int. J. Comput. Vision 130(5), 1366–1401.*

Soliman, M; Kamal, M; Nashed, M; Mostafa, Y; Chawky, B and Khattab, D. (2019). “Violence Recognition from Videos using Deep Learning Techniques”, *Proc. 9th International Conference*

Kumar Sangaiah, Mohamed Elhoseny and Sung Wook Baik. (2020). Raspberry Pi assisted face recognition framework for enhanced law-enforcement services in smart cities, *Future Generation Computer Systems, Volume 108, 2020, Pages 995-100*

of Ambient Intelligence and Humanized Computing, 7(2), 149-157.

Santos, F., Durães, D., Marcondes, F. S., Hammerschmidt, N., Lange, S., Machado, J and Novais, P. (2021, November). In-car violence detection based on the audio signal. *In International conference on intelligent data engineering and automated learning (pp. 437-445). Cham: Springer International Publishing.*

Omarov, B; Narynov, S; Zhumanov, Z; Gumar, A and Khassanova, M. (2022). State-of-the-art violence detection techniques in video surveillance security systems: a systematic review.

Rahman, M.M; Gupta, D; Bhatt, S and Shokouhmand, S. (2020). Real-time anomaly detection and classificationfrom surveillance cameras using Deep Neural Network.

Santosh, K.D. (2021). An intelligent video surveillance system for crime detection using face recognition in smart cities. *Security and Communication Networks, 2021, 1-16.*

Serrano, G.I; Deniz S.O; Bueno G.G and Kim, T. K. (2015). Fast fight detection. *PloS one, 10(4), e0120448.*

APPENDIX

Figures and Tables that are not Author's work are listed below: Figures 3, 4 and 5 and Table 4.

Souto, H., Mello, R. and Furtado, A (2019). An acoustic scene classification approach involving domestic violence using machine learning. *In: Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional, pp. 705–716. SBC.*

Wu, P., Liu, J., Shi, Y., Sun, Y., Shao, F., Wu, Z and Yang, Z. (2020). Not only look, but also listen: Learning multimodal violence detection under weak supervision. *In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXX 16 (pp. 322-339). Springer International Publishing.*