KOMTEKINFO JOURNAL
LPPM Universitas Putra Indonesia YPTK Padang
Jl. Raya Lubuk Begalung, Padang, West Sumatera, Indonesia, Zip Code: 25221
Volume: 12, Issue: 4, Page: 220 - 225, Desember 30th, 2025, e-ISSN: 2502-8758
Available online at website: https://jkomtekinfo.org/ojs/index.php/komtekinfo

Convolutional Neural Network Method in Detecting Digital Image Based Physical Violence

Hasibuan Elpina[1]✉, Yuhandri[2], Sumijan[3]
[1,2,3]Master Program, Informatics Engineering, Faculty of Computer Science, Universitas Putra Indonesia YPTK Padang, Padang, 25221, Indonesia

elpina28pina@gmail.com

Abstract

Physical violence in the educational environment has a serious impact on mental health, safety, and student achievement, in addition to causing physical injury, violence can cause psychological trauma that interferes with the learning process, due to the limited supervision system, lack of officers, and the absence of automatic detection technology. This research aims to design and develop an automatic detection system of physical violence using digital image processing technology. This study uses the Convolutional Neural Network (CNN) method with the stages of digital image collection and labeling, preprocessing, model training, and evaluation using accuracy, precision, recall, and F1-score metrics. The CNN architecture was chosen because it is efficient and accurate, and it supports data augmentation to improve generalization. The dataset was taken from kaggle and primary data at the al-falah huraba Islamic boarding school which consisted of 2000 images which included: 800 images of violence on CCTV of the dormitory room, 500 images of violence simulation of training videos and 500 non-violent images. The results showed that the developed CNN model was able to detect physical violence with an accuracy of above 88%, making it feasible to apply in surveillance camera-based school surveillance systems (CCTV). The system is able to classify images in real-time into two categories: safe and hard. This research contributes to the use of artificial intelligence to support efficient and affordable technology-based education security.

Keywords: physical violence, violence detection, digital imagery, CNN, U-Net

1. Introduction

The implementation of AI (Artificial Intelligence) has grown rapidly over the past few years [1]. In general, artificial intelligence is currently divided into two main branches, namely Deep Learning and Machine Learning [2]. One of the most widely used Deep Learning techniques today is the Convolutional Neural Network (CNN) [3]. U-Net is a neural network architecture to upsample and produce high-resolution segmentation [4].

The U-Net architecture is used to classify images, where this architecture can be used as a solution to physical violence cases to increase surveillance [5]. Violence, whether physical, mental, or emotional, is a serious problem [6]. Surveillance systems can be implemented effectively to detect and respond to acts of violence and bullying quickly and accurately. Installing CCTV is a form of security or supervision of students [7]. This is one of the efforts to help schools [8].

Therefore, an automated system can automatically detect digital images or videos [9] of violence detection is needed to increase effectiveness without the need for continuous human supervision [10]. System development by utilizing cutting-edge machine learning techniques to detect violence [11]. Violence detection apps can be useful for monitoring bullying and reducing the number of bullying cases that occur in schools [12].

One type of neural network that is commonly used in image data is Convolutional Neural [13]. Artificial intelligence used in education is Deep learning [14].Convolutional Neural Network (CNN) is one of the methods in Deep Learning [15]. CNN is a development of Multi Layer Perceptron (MLP) and is one of the algorithms of Deep Learning. The CNN algorithm plays a role in the process of extracting traits from the input image data [16]. The CNN method has the most significant results in image recognition [17].

The development of a traffic violation detection system, using the YOLO3 method combined with CNN and LSTM has a higher accuracy of 89%. Meanwhile, in the CNN base model, the resulting accuracy is 85% [16]. The combination of Convolutional Neural Network and Face-API for Effective and Efficient Online Attendance

Submitted: 15-10-2025 | Revised: 15-11-2025 | Accepted: 14-12-2025 | Published: 30-12-2025

Tracking methods in the development of face detection applications for online attendance is designed to help teachers track student attendance. CNN is designed to process two-dimensional data because it is the development of Multilayer [18][19].

The findings in this study suggest that deep transfer learning and image augmentation can improve detection accuracy [20]. In the study, the detection of the level of hate speech with the LSTM model achieved an overall accuracy of 93.14% [21]. The selection of the right classification method is indispensable according to the factors that support the level of processing to solve the problem [22]. In measuring the performance of the training process, it will be compared with some basic detection models used, such as CNN, VGG16, ResNetNet50, MobileNetV2, YOLO3, and YOLO3+LSTM. [23].

However, most previous research has focused on violence in public spaces such as stadiums, stations, and highways. Few specifically explore the context of schools, where the characteristics of the environment are more homogeneous and the types of violence tend to occur spontaneously and in short duration. Therefore, this study aims to develop a visual detection system of physical violence based on digital imagery designed specifically for educational environments. This system is expected to be able to distinguish between safe and violent activities accurately, quickly, and can be integrated with existing surveillance cameras (CCTV) in schools.

2. Methods

This study uses an experimental quantitative approach with a deep learning-based model development method. The framework of this study describes systematic steps to develop and test a physical violence detection system using the CNN method, which is focused on visual identification in the environment of the Al-Falah Huraba Islamic Boarding School. The research stages are designed to produce accurate and adaptive detection models, as shown in Figure 1.

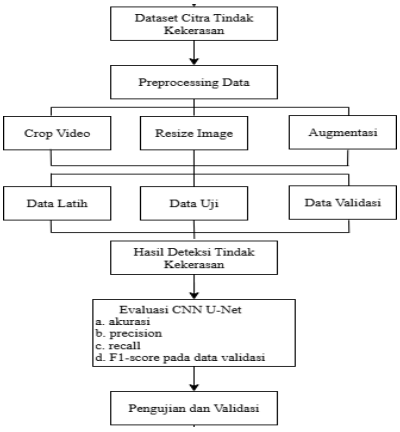

Figure 1. Research Framework

Figure 1 shows the process of detecting violence with CNN U-Net, starting from the collection and preprocessing of datasets (video crop, image resize, augmentation), the division of data into training, testing, and validation, followed by detection, evaluation using accuracy, precision, recall, and F1-score, and ending with testing and validation of the model.

2.1 Violence Imagery Dataset

Datasets are an important component in machine learning-based research. At this stage, a dataset of images or video footage containing violent and non-violent scenes is collected. This dataset is the basis for training and testing of the violence detection model that will be developed.

To improve the accuracy of the model and avoid overfitting, data augmentation was carried out in the form of random rotation, exposure changes, zoom in/out, and horizontal flipping. This technique has been shown to improve the model's resistance to variations in lighting conditions and viewing angles. This data is taken directly from surveillance cameras (CCTV). To complement the limited number of real data, simulated data on physical violence was also made in a controlled manner.

2.2 Preprocessing Data

Data preprocessing aims to standardize the format and reduce noise so that the CNN U-Net model can recognize the data optimally. This process includes cropping to crop an area of the image or video so that only important parts that contain an indication of hardness, resize to equalize the dimensions of the image according to the model's input, as well as augmentation such as rotation, flipping, or lighting adjustment to increase data variety and increase model durability. After that, the dataset is divided into training data for the learning process, test data to measure initial performance, and validation data to evaluate the model's generalization ability against new data.

2.3 Results of Violence Detection

CNN and U-Net models that have gone through the training process are then used on test data to detect acts of violence. This process results in a prediction of whether an image or video footage contains elements of violence or not, which is then used as a reference in evaluating the model's performance.

2.4 Evaluation of the CNN Model

Model performance evaluation was conducted to assess the ability of CNN and U-Net to detect harshness, using accuracy, precision, recall, and F1-score parameters. The calculation of each parameter is described in the following formula.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} X\ 100\% \quad (1)$$

$$Precision = \frac{TP}{TP+FP}\ X\ 100\% \quad (2)$$

$$Recall = \frac{TP}{TP+FN} X\ 100\% \quad (3)$$

$$F1 - Score\ = 2\ X \frac{Precision\ X\ Recall}{Precision + Recall}\ X\ 100\% \quad (4)$$

Table 1 shows the model's evaluation results, with an accuracy of 91.47%, precision of 93.05%, recall of 85.90%, and an F1 score of 89.0%. The average inference time is 117 ms per frame, indicating efficient performance.

Table 1. Evaluation Results

Evaluation Metrics	Result
Accuracy	91.47%
Precision	93.05%
Recall	85.90%
F1 Score	89.0%
Average inference	117 ms/frame

2.5 Testing and Validation

After the evaluation process, the model is further tested using validation data to ensure its performance stability. This stage aims to verify that the model not only delivers good results on training data and test data, but is also able to maintain performance when dealing with new data that has never been processed before.

2.6 CNN Model Architecture

The CNN model is built with a layered architecture consisting of:

a. Convolutional layers to extract spatial features from images,
b. Pooling layer to reduce dimensions and prevent overfitting,
c. Fully connected layer for final classification.

As a variation, additional experiments were conducted with the CNN-LSTM hybrid model to capture the temporal sequence of the video frame. They showed that the integration of CNN and LSTM was able to improve recall and precision in detecting physical violence in short videos.

The model was trained using the Adam optimizer, learning rate 0.0001, batch size 32, and epoch 50 times. To speed up the training process and improve accuracy, transfer learning with the pretrained ResNet50 model is also used, CNN Model Architecture shown in Figure 2.

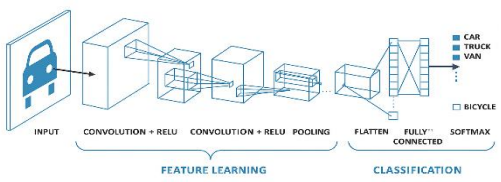

Figure 2. CNN Model Architecture

This image illustrates the basic architecture of the Convolutional Neural Network (CNN) used in the study to classify images from school environments into two classes: "Safe" and "Violent". The flowchart shows the process of transforming data from input to output, namely:

a. Input Image
It is an image or frame taken from school surveillance video (CCTV) recording.
b. Convolutional Layer
Some convolutional layers function to extract spatial features from the image (e.g., body movements, hand direction, object position). Each layer uses filters to detect edges, angles, and visual patterns.
c. Relu Layer (Rectified Linear Unit)
It is used to add non-linearity to the model and speed up the convergence process during training.
d. Fully Connected Layer
A final layer that combines all the features and provides predictions of the target class.
e. Output Layer
Provides a final classification, i.e. "Safe" or "Hard", based on the results of the feature analysis.

2.7 Devices and Infrastructure

The model was developed using the Python programming language with the TensorFlow and Keras frameworks. The experiment was conducted on laptops with Intel Core i7 processor specifications, 16 GB of RAM, and an NVIDIA RTX 3060 GPU. For real-time speed testing, the system was also implemented on the Jetson Nano as a simulation deployment in schools with limited infrastructure [6]. Hardness Detection System Flowchart shown in Figure 3.

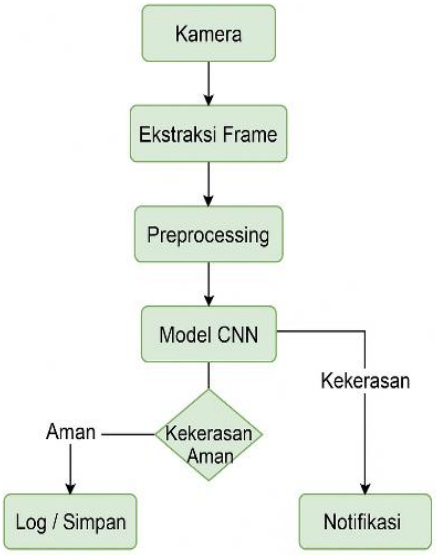

Figure 3. Hardness Detection System Flowchart

This image is a representation of the process flow of the digital image-based violence detection system designed in the research. This diagram shows the logical sequence of data processing stages from the camera input to the detection output, namely:

a. Camera
   Surveillance cameras (CCTV) continuously record activities in the school environment.
b. Frame Extraction
   The video footage is broken down into individual pieces of images (frames) for independent analysis.
c. Preprocessing
   Stages of image size normalization, color conversion (e.g. to grayscale or RGB), and augmentation (flip, rotation, brightness).
   The goal is to improve the quality and consistency of the data that goes into the model.
d. Model CNN
   The preprocessed images were fed into the Convolutional Neural Network (CNN) model for classification.
e. Classification Process
   CNN provides output in the form of two labels:
   Safe: activities that show no signs of violence.
   Violence: suspicious activity or showing aggressive behavior.
f. Output-Based Actions
   If the classification is "Secure", then the data is logged or stored as a reference.
   If the classification is "Violence", the system automatically sends a notification to the relevant parties (e.g. school security officers or homeroom teachers).

3. Results and Discussions

The analysis and design stages are a crucial foundation in the process of developing artificial intelligence-based systems, especially those that involve the process of automatically detecting physical violence images.

3.1 Stages of Analysis and Planning

The analysis and design of the system is arranged through six main stages which will be described in detail in the following discussion. To make it easier to understand the process flow, the stages are visualized in the form of a chart shown in Figure 4.

Figure 4. Planning Flow Chart

Figure 4 shows the initial flow of the development of a violence detection system, starting from data collection, analysis of needs and methods, to designing system structures and architectures to support effective detection functions.

3.2 Data on Violence

The data sources in this study come from two main types that complement each other, namely real-time data from CCTV and simulation data of violent acts. Labeling is done manually by two annotators to distinguish between violent (aggressive movement) and non-violent (normal activity) images. Quality is maintained by cross-validation and consensus review according to the principle of inter-rater reliability. The data limitations in the CNN model are addressed with image augmentation, such as rotation, flipping, lighting adjustments, noise addition, and zoom and crop, to improve generalization and reduce overfitting. Table 2 summarizes the dataset details, including the number of samples and categories.

Table 2. Dataset

Yes	Data Source	Label	Number of Images	Format
1	CCTV Dormitory Room	Violence	800	.jpg/.png
2	Classroom CCTV	Non-Violence	700	.jpg/.png
3	Simulation Training Video	Violence	500	.jpg/.png
4	Additional Datasets	Non-Violence	500	.jpg/.png
	Total	—	2.500	—

The table summarizes 2,500 images from various sources, including CCTV footage of dormitories and classrooms, simulation of training videos, and additional datasets. This data labeled *violent* and *non-violent*, in *.jpg* and *.png* formats, is used to train the system to be able to accurately distinguish between violent and non-violent activities. The following are the results of the detection experiment on the detection of physical violence. Violent Detection Result shown in Figure 5.

Figure 5. Violent Detection Results

The image shows the results of the detection of physical violence from CCTV footage, where the *computer vision* system marked the area of the incident with a red box and labeled it "Violence Detected" as an indication of an act of violence.

3.3 System Analysis

The analysis of this system includes the design of a CNN-based and Computer Vision-based violence detector at the Al-Falah Huraba Islamic Boarding School, including user needs, data flows, algorithms, module integration, and risk mitigation. The modular system consists of CCTV video acquisition, frame preprocessing, CNN classification, U-Net segmentation, result storage, and real-time monitoring dashboard, designed based on software engineering principles and the latest AI image processing practices.

3.4 System Design

The output design displays the results of the violence detection in real-time. The live feed shows detected objects with a red box, the detection log records details of the time, camera location, and probability, while notifications provide quick alerts when violence occurs. The camera and date filter feature makes it easy to search for detection data, which can be seen in Figure 6 below.

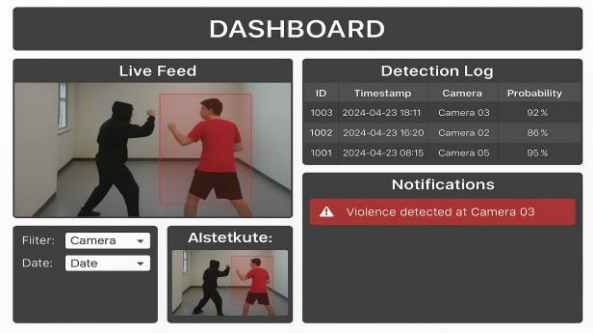

Figure 6. System Design

4. Conclusions

In conclusion, there is a research gap in the development of violence detection systems that are appropriate for school contexts, as most previous studies have focused more on public spaces. With the characteristics of a homogeneous school environment and spontaneous and brief incidents of violence, a solution is needed that is able to quickly and accurately recognize the difference between safe and violent activities. This research answers this need by designing a digital image-based visual detection system that is integrated with school CCTV, so that it can improve security and response to violence in the educational environment.

References

[1] R. D. Natasya, "Implementasi Artificial Intelligence (Ai) Dalam Teknologi Modern," *J. Komput. dan Teknol. Sains*, vol. 2, no. 1, pp. 22–24, 2023.

[2] A. Supriyadi, "Penerapan Metode Joyful Learning dalam Pembelajaran," *J. Pendidik.* , vol. 8, no. 2, pp. 123–135, 2020.

[3] J. A. Figo, N. Yudistira, and A. W. Widodo, "Deteksi Covid-19 dari Citra X-ray menggunakan Vision Transformer," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, pp. 1116–1125, 2023, [Online]. Available: https://j-ptiik.ub.ac.id

[4] I. S. D. A. N. U. Untuk, D. Dan, and P. Tingkat, "Implementasi ssd-mobilenet dan u-net untuk deteksi dan penilaian tingkat keparahan pada aplikasi pelaporan jalan berlubang," vol. 12, no. 3, pp. 4334–4344, 2024.

[5] S. A. Putri, A. Rifai, and I. Nawawi, "Physical Violence Detection System to Prevent Student Mental Health Disorders Based on Deep Learning," *J. Pilar Nusa Mandiri*, vol. 19, no. 2, pp. 103–108, 2023, doi: 10.33480/pilar.v19i2.4600.

[6] S. Nasional, T. Elektro, S. Informasi, and T. Informatika, "Sistem Deteksi Kekerasan," pp. 198–204, 2024.

[7] G. A. Sidik, "Deteksi Tindak Kekerasan dan Perundungan Pada Anaka Berbasis YOLOv8 (You Only Look Once)," vol. 3, no. 9, 2024.

[8] E. Naf'an, "Sistem Deteksi Penggunaan Masker Saat Keluar Rumah Berbasis Smartphone dan Arduino," *J. KomtekInfo*, vol. 8, pp. 232–238, 2021, doi: 10.35134/komtekinfo.v8i4.188.

[9] D. K. Doni, Yuhandri, and A. Ramadhanu, "Penerapan Algoritma Haar Cascade Clasifier dan Computer Neural Network Sebagai Presensi Karyawan," *J. KomtekInfo*, vol. 11, no. 4, pp. 398–408, 2024, doi: 10.35134/komtekinfo.v12i1.565.

[10] A. D. Fikri, P. F. Utaminingrum, and E. G. Edhi, "Sistem Pendeteksi Kekerasan di Ruang Publik Menggunakan Metode 3D Convolutional Neural Network," vol. 1, no. 1, pp. 1–6, 2023.

[11] S. A. Putri *et al.*, "PHYSICAL VIOLENCE DETECTION SYSTEM TO PREVENT STUDENT," vol. 19, no. 2, 2023, doi: 10.33480/pilar.v19i2.4600.

[12] A. Informatics, S. A. Putri, A. Rifai, I. Nawawi, and A. Info, "Aplikasi Cerdas Sistem Deteksi Tindak Kekerasan Fisik Untuk Pengawasan Perundungan Dengan Convolutional Neural Network," vol. 7, no. 2, pp. 332–340, 2024.

[13] K. Disease, "IMPLEMENTASI ALGORITMA CONVOLUTIONAL NEURAL NETWORK UNTUK MENDETEKSI PENYAKIT GINJAL IMPLEMENTATION OF CONVOLUTIONAL NEURAL NETWORK FOR DETECTING," vol. 4, no. 2, pp. 212–219, 2022.

[14] I. Riati, Yuhandri, and G. W. Nurcahyo, "Penerapan Convolutional Neural Network Untuk Mengidentifikasi Penyakit Tanaman Kelapa Sawit," *J. KomtekInfo*, vol. 11, no. 4, pp. 237–246, 2024, doi: 10.35134/komtekinfo.v11i4.554.

[15] N. Dewi and F. Ismawan, "Implementasi Deep Learning Menggunakan Convolutional Neural Network untuk Sistem Pengenalan Wajah," vol. 14, no. 1, pp. 34–43, 2021, doi: 10.30998/faktorexacta.v14i1.8989.

[16] C. N. Network, "Convolutional Neural Network and LSTM for Seat Belt Detection in," vol. 5, no. 158, pp. 4–6, 2024.

[17] A. Peryanto *et al.*, "Rancang Bangun Klasifikasi Citra Dengan Teknologi Deep Learning Berbasis Metode Convolutional Neural Network," vol. 8, pp. 138–147, 2019.

[18] R. A. Tilasefana and R. E. Putra, "Penerapan Metode Deep Learning Menggunakan Algoritma CNN Dengan Arsitektur VGG NET Untuk Pengenalan Cuaca," *J. Informatics Comput. Sci.*, vol. 05, no. 1, pp. 48–57, 2023.

[19] S. Carita and R. B. Hadiprakoso, "Double Face Masks Detection Using Region-Based Convolutional Neural Network," vol. 9, no. 4, pp. 904–911, 2023, doi: 10.26555/jiteki.v9i4.23902.

[20] Y. Setiawan, N. U. Maulidevi, and K. Surendro, "DETEKSI CYBERBULLYING DENGAN MESIN PEMBELAJARAN KLASIFIKASI (SUPERVISED

LEARNING): PELUANG DAN TANTANGAN CYBERBULLYING DETECTION USING SUPERVISED LEARNING :," vol. 9, no. 7, pp. 1577–1582, 2022, doi: 10.25126/jtiik.202296747.

[21] F. Setiawan, A. Wahyudi, and N. A. Febriyanti, "Deteksi Tingkat Keparahan Ujaran Kebencian Menggunakan Bi-LSTM pada Teks Bahasa Indonesia," vol. 12, pp. 21–28, 2024.

[22] B. Neural and C. Neural, "Komparasi Metode Backpropagation Neural Network dan Convolutional Neural Network Pada Pengenalan Pola Tulisan Tangan," vol. 6, no. 1, pp. 56–63, 2022.

[23] K. Yolo, F. Firdausillah, E. D. Udayanti, E. Kartikadarma, I. Komputer, and U. Dian, "JURNAL RESTI," no. 158, pp. 355–360, 2024.

Biographies of Authors

Elpina Sari Dewi Hasibuan is a Master of Informatics Engineering student at YPTK Universitas Putra Indonesia Padang. She was born on November 18, 1999, in Sinonoan. She completed her bachelor's degree in information systems at STMIK Indonesia Padang Campus in 2022. She is a teacher at SMK Negeri 1 Siabu. She lives in Sinonoan Village, Siabu District, Mandailing Natal Regency, North Sumatra Province, Indonesia.

Yuhandri was born in Tanjung Alam on May 15. He is an Assistant Professor in Faculty of Computer Science, Universitas Putra Indonesia YPTK. He received the Bachelor Degree in Informatics Management and Master Degree in Information Tecnology in 1992 and 2006 from Universitas Putra Indonesia YPTK. Moreover, he completed his Doctorate of Information Technology as Informatics Medical Image expertise from Gunadarma University in April 2017. He is a lecturer at the Faculty of Computer Science, Universitas Putra Indonesia YPTK. Scopus Id is 57193430920. E-mail: yuyu@upiyptk.ac.id

Sumijan was born in Nganjuk on May 7 1966. He received the Bachelor Degree in Informatics Management in 1991 from Universitas Putra Indonesia YPTK, Master of Information Technology in 1998 from University Technology Malaysia (UTM). He completed has Doctorate of Information Technology as Medical Image Expertise from Gunadarma University in December 2015. He is member of ACM (23145751). Scopus Id is 57194787076. E-mail: soe@upiyptk.org