



Contents lists available at [www.journal.unipdu.ac.id](http://www.journal.unipdu.ac.id)



Journal Page is available to [www.journal.unipdu.ac.id/index.php/register](http://www.journal.unipdu.ac.id/index.php/register)



Research article

# Detecting violent scenes in movies using Gated Recurrent Units and Discrete Wavelet Transform

Elly Matul Imah <sup>a,\*</sup>, Ivan Kurnia Laksono <sup>b</sup>, Karisma Karisma <sup>c</sup>, Atik Wintarti <sup>d</sup>

<sup>a,d</sup> Department of Mathematics, Universitas Negeri Surabaya, Surabaya, Indonesia

<sup>b,c</sup> Unit of Artificial Intelligence and Scientific Publication, Universitas Negeri Surabaya, Surabaya, Indonesia

email: <sup>a,\*</sup> [ellymatul@unesa.ac.id](mailto:ellymatul@unesa.ac.id), <sup>b</sup> [ivankurnialaksono@gmail.com](mailto:ivankurnialaksono@gmail.com), <sup>c</sup> [karisma.17030214035@mhs.unesa.ac.id](mailto:karisma.17030214035@mhs.unesa.ac.id), <sup>d</sup> [atikwintarti@unesa.ac.id](mailto:atikwintarti@unesa.ac.id)

\* Correspondence

## ARTICLE INFO

### Article history:

Received 7 September 2021

Revised 21 October 2021

Accepted 31 December 2021

Available online 07 April 2022

### Keywords:

deep learning

gate recurrent unit

violence detection

video processing

wavelet

Please cite this article in IEEE

style as:

E. M. Imah, I. K. Laksono, K.

Karisma and A. Wintarti,

"Detecting violent scenes in

movies using Gated Recurrent

Units and Discrete Wavelet

Transform," *Register: Jurnal*

*Ilmiah Teknologi Sistem Informasi*,

vol. 8, no. 2, pp. 94-103, 2022.

## ABSTRACT

The easiness of accessing video on various platforms can negatively impact if not done wisely, especially for children. Parental supervision is needed so that movies platforms avoid inappropriate displays such as violence. Violent scenes in movies can trigger children to commit acts of violence, which is not desired. Unfortunately, it is not easy to supervise them fully. This study proposed a method for automatic detection of violent scenes in movies. Automatic violence detection assists the parents and censorship institutions in detecting violence easily. This study uses Gated Recurrent Units (GRU) algorithm and wavelet as feature extraction to detect violent scenes. This paper shows comparative studies on the variation of the mother wavelet. The experimental results show that GRU is robust and deliver the best performance accuracy of 0.96 while combining with mother wavelet Symlet and Coiflets8. The combination of GRU with wavelet Coiflets8 shows better results than the predecessor.

Register with CC BY NC SA license. Copyright © 2022, the author(s)

## 1. Introduction

Violence against children is one of the social events that need to be handled immediately. Violence against children can be in the form of physical, sexual, emotional abuse, neglect, and exploitation. Based on a study conducted by Hills and colleagues, the severe violence of incidence against children in Asia is around 64% [1]. Based on data from the WHO report, the incidence of violence against children during the COVID-19 pandemic is increasing [2, 3, 4]. Violence experienced by children has a tremendous impact, ranging from learning difficulties, experiencing excessive anxiety, experiencing depression, disability, and experiencing mental disorders to death [5].

It is suspected that accessing various videos through various platforms has an essential role in increasing violence cases because viewing or watching violent videos can affect a person's behavior [6]. Especially now that anyone can easily access multiple shows that are not necessarily worth watching and negatively impact psychology, especially for children. Parents play an important role in supervising and controlling their children, but they also cannot have complete control throughout the day. Supervision from an older person is necessary so that streaming platforms avoid serving inappropriate content such as violence. Violent scenes in movies can induce children to carry out violent acts. However, supervising children while watching movies is a problematic task [7, 8]. Violence detection scenes in movies will help parents by alerting them to violent content so they can skip violent scenes or address their children to watch other suitable programs.

Violence detection based on visual video is a challenging topic in computer vision. Several methods used for detection are a combination of audio and visual information [9, 10, 11]. However,

visual information is dominant in determining the results, so we focus on methods that use visual information in this study—some of the studies that have been carried out. Authors in [12] researched video violence detection using CNN and LSTM. The data used consists of various types and video formats. Mahmoodi and Salajeghe's study detects violence using the Histogram of Optical flow Magnitude and Orientation based on the optical flow method [13]. This study developed an artificial intelligence-based device to detect violent acts on videos automatically. This system uses the Gated Recurrent Unit (GRU) deep learning algorithm to detect the violence contained in a film. This automation will make it easier for censorship agencies or parents to control every film that children watch for speeding up work without having to totally watch all the movies [14].

Samuel et al. researched real-time violence detection using the LSTM method. They used data in movie scenes consisting of 2314 videos with 1077 fight ones and 1237 no-fight ones [15]. Violence detection using VGG 16 and SVM shows promising results for the hockey dataset [16]. Although the results obtained are quite adequate, obtaining an accuracy of 94.5%, the time required is quite long. Deepak et al. conducted a study on autocorrelation of gradients-based violence detection [17]. Wang et al. researched Violence detection and face recognition based on deep learning. The model they use is CNN for the Crowd and Hockey dataset [18]. In 2021, Sen and Deb categorized actions in soccer videos using Gated Recurrent Unit (GRU) and VGG as extraction methods. Sen and Deb revealed that this architecture could overcome the difficulties of detecting actions accurately [19]. Unsupervised recurrent deep learning scheme also used for process monitoring [20, 21, 22].

In the classification required feature extraction process. Several researchers have carried out the integration of feature extraction with machine learning classification methods, as in [23, 24]. Wavelet is a feature extraction method that shows good video and computer vision [25, 26]. In their research, Chatterjee and Halder show that wavelets have good accuracy and resistance combining deep learning for violence detection [27]. Wavelets also show good results in face anti-spoofing [25, 28]. The author in [29] combines DWT with CNN-BiLSTM for violence detection. The proposed method is computationally lighter because the authors used only part of the structure instead of passing a sequence of complete frames to the neural network. Based on these references, we use the GRU algorithm and wavelet as an extraction method to detect violent scenes in the movie. Based on these references, we use the GRU algorithm and wavelet as an extraction method to detect violent scenes in the movie. The use of this approach is because the GRU's parameters are fewer than LSTM [30], so it can achieve a time-efficient with a satisfactory amount of accuracy.

The rest of this paper is organized as follows. The second section describes the material and method that explain wavelet feature extraction and GRU classification: the experiment set up, evaluation measurement, experiment results, and discussion in the third section. The last section is the conclusion of this study.

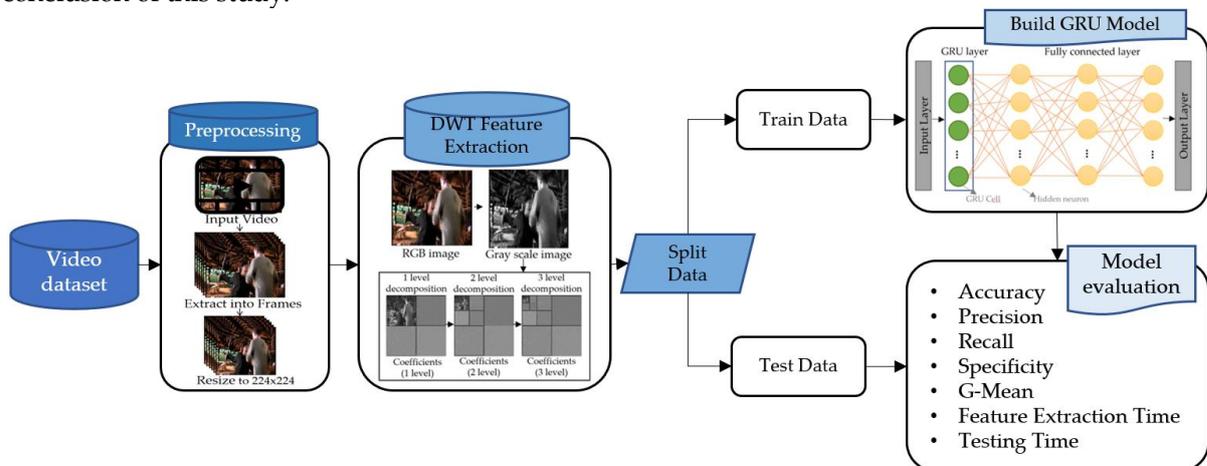


Fig. 1. Violence detection system

## 2. Materials and Methods

The violence detection process can be seen in Fig. 1. The process is divided into six parts: input video dataset, splits data, pre-processing, feature extraction, classification, and evaluation. The first process is input video data for pre-processing. After that, split the dataset into train and test data using k-fold

cross-validation. We captured several images from each video in the pre-processing and then resized the collection images to  $224 \times 224$  pixels. The next step is to convert the image batch into grayscale and extract the features using DWT with several mother wavelets. The approximation coefficient of each video obtained from the DWT process is reshaped into 1-Dimensional vectors, then used as input to the GRU. These vectors will pass through the GRU cell consisting of 2 gates: the reset gate and the update gate. We use the SoftMax function to generate violence or non-violence labels in the output layer. The label is then evaluated. In addition, we save the CPU time by evaluating the time required during the testing process.

### 2.1. Dataset

The dataset used in this research is secondary data downloaded from the Movie dataset [31]—This dataset contains 200 videos divided into two classes, namely violent and non-violent. The violent class consists of 100 videos in the .avi format, and the non-violent class has the same number of videos but in .mpg format. The duration of each video ranges from 1 to 2 seconds. The sample images of this dataset are shown in Fig. 2.

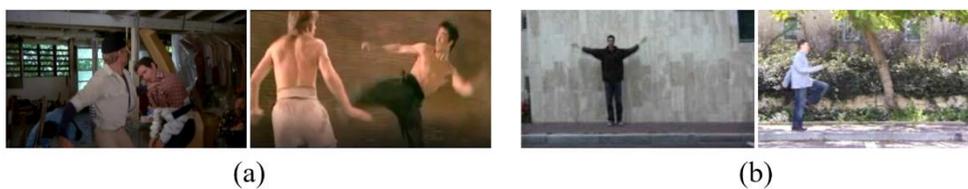


Fig. 2. (a) Sample of movie dataset violent scene and (b) Non-violent scene images

### 2.2. Movie Data Pre-processing

The first process carried out in this research is pre-processing. The pre-processing stage is done by extracting the video into a collection of images. This collection of images is obtained by capturing video. From one video, as many as 20 images are collected. This collection of images is later referred to as a batch of images. After obtaining a batch of images, the next step is to change the image size to  $224 \times 224$  pixels. It is used to adjust the size of the input on the input layer when performing feature extraction. This process produces a data matrix with dimensions of  $n \times 20 \times 224 \times 224 \times 3$ , where  $n$  is the number of videos. The illustration of the pre-processing stage is shown in Fig. 3.

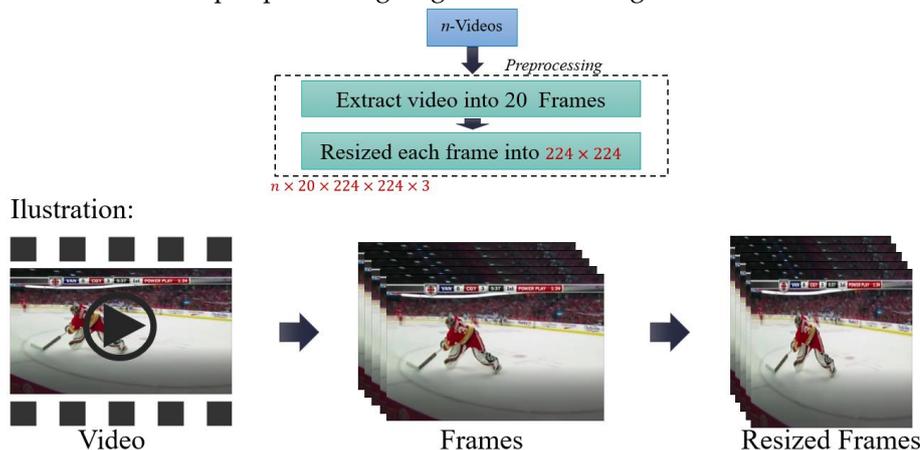


Fig. 3. Movie data pre-processing

### 2.3. Discrete Wavelet Feature Extraction

Wavelet is a method that can be used for signal processing, image analysis, and compression [32]. Wavelets were first introduced by Alfred Haar in 1909 and developed by Jean Morlet and Alex Grossman in the 1980s. In general, wavelets are divided into two types of transformations: Continuous Wave Transformation (CWT) and Discrete Wave Transform (DWT) [33]. The difference between the two is in the value of the scale coefficient ( $a$ ) and the translational coefficient ( $b$ ), namely by limiting  $a$  and  $b$  only to discrete values with  $a = a_0^m$ ,  $b = nb_0 a_0$  for  $m, n \in Z$ ,  $a_0 > 1$ ,  $b_0 > 0$ . In this study, the focus is on applying DWT because it is relatively easy to use. DWT follows the Eq. 1.

$$C_f(a, b) = a_0^{-\frac{m}{2}} \int_{-\infty}^{\infty} f(t) \psi(a_0^{-m} t - nb_0) dt \quad (1)$$

Wavelet has many variants, for DWT including Haar, Daubechies, Symlets, Coiflets, Biorthogonal, Reverse biorthogonal, and Discrete Meyer. The Haar wavelet is the oldest and simplest wavelet. The Haar wavelet is the same as the Db1 Wavelet (Daubechies order 1). The length of the Haar Wavelet sieve is 2 [26]. The Daubechies wavelet was discovered by a mathematician named Ingrid Daubechies. In the Daubechies Wavelet filtration process, there are low-pass and high-pass filtering. A low-pass filter produces a low-frequency subfield coefficient, and a high-pass filter makes a high-frequency subfield. The Daubechies wavelet has an order where the order describes the number of filter coefficients. For the Daubechies wavelet with order  $N$  ( $db-N$ ), the Daubechies wavelet has a filter coefficient size of  $2N$  [34]. Daubechies wavelet has an asymmetrical shape; to improve it, Ingrid Daubechies expands and forms Coiflet. The Coiflet filters are not much different from Daubechies.

The Symlet or Symmetric Wavelet is a development of Daubechies. Although the resulting shape is not entirely symmetrical, the Symlet filter has a balanced design character [34]. Biorthogonal wavelet is an extension of orthogonal wavelet. The term biorthogonal refers to two primary functions that are orthogonal to each other, but each does not form an orthogonal set. The degree of freedom of the biorthogonal wavelet is higher than that of the orthogonal wavelet. In the discrete wavelet, the orthogonality condition satisfies the Eq. 2,

$$\sum_{m \in \mathbb{Z}} h(k)h^*(k) + 2m + 2m = 2\delta m_0 \tag{2}$$

with  $h(k)$  is a function coefficient used on the first scale wavelet transformation and  $h^*(k)$  is the coefficient of the second scale function used in the inverse transformation. Biorthogonal wavelets are widely used in signal and image processing because of their symmetry of perfect reconstruction. Although Biorthogonal and Reverse Biorthogonal filters have different orthogonality relationships than orthogonal wavelets, they still have ideal reconstruction properties. In addition to reducing the size of the Biorthogonal wavelet, it is also used for image enhancement to obtain detailed coefficients.

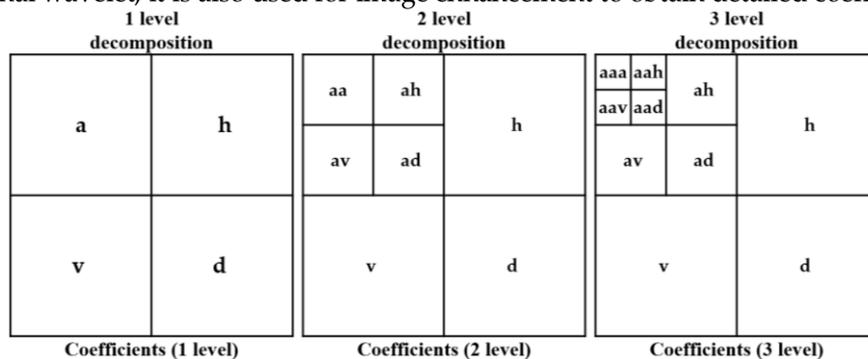


Fig. 4. Feature Extraction Using DWT

Meyer's wavelet construction is fundamentally a solvent method for solving the two-scale equation. Given a basis  $\phi$  for the approximation space  $V_0$  Meyer employed Fourier techniques to derive the DTFT of the two-scale equation coefficients,  $g_0[n]$ , from  $\Phi(\omega)$  in Eq. 3 [35].

$$G_0(e^{f\omega}) = \sqrt{2} \sum_k \phi(2\omega + 2k\pi) \tag{3}$$

The multilevel discrete wavelet transform involves further decomposition of the sub-bands at each level, but only the approximation coefficients (sub-band low-pass), which are further decomposed at the next level. In 2D, the discrete wavelet transform produces four sets of coefficients corresponding to the four possible combinations of the wavelet decomposition filter. The detail of the multilevel wavelet transform can be seen in Fig. 4 [33].

#### 2.4. Gated Recurrent Unit (GRU)

Gated Recurrent Unit (GRU) was first introduced by Kyunghyung et al. in 2014. The idea of the GRU is to create a network that can select the essential information needed according to a particular purpose adaptively in each recurrent unit [19]. The GRU is like a long short-term memory (LSTM) with a forget gate but has fewer parameters than the LSTM, as it lacks an output gate [36]. In Fig. 5,  $r$  represents reset gates, and  $z$  represents update gates.

In the GRU, the component that regulates the flow of information is called a gate, while the reset gate is a component that is used to determine how to combine new input with past data, and the update gate is a component that serves to determine how much past information should be kept. The equations for update gate  $Z_t$  (Eq. 4) and reset gate  $r_t$  (Eq. 5) are as follow:

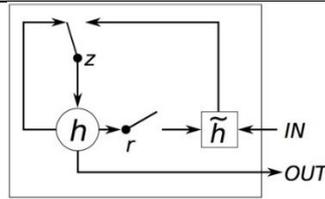


Fig. 5. Gating Structure GRU

$$Z_t = \sigma(W^z x_{t-1} + V^z h_{t-1} + b_z) \tag{4}$$

$$r_t = \sigma(W^r x_{t-1} + V^r h_{t-1} + b_r) \tag{5}$$

hidden state  $h_t$  is related to the current linear interpolation ( $t$ ) with the previous activation function ( $t-1$ ).  $h_t$  is formulated as Eq. 6 and Eq. 7,

$$h_t = (1 - Z_t) \otimes h_{t-1} + Z_t \otimes \hat{h}_t \tag{6}$$

$$\hat{h}_t = \tanh(W^c x_t + V^c (r_t \otimes h_{t-1})) \tag{7}$$

The matrices  $W^z$ ,  $W^r$ , and  $W^c$  are the weighted input matrices, and  $V^z$ ,  $V^r$ , and  $V^c$  are the recurrent weight matrices. The vectors  $b_z$  and  $b_r$  are the bias vectors. The function  $\sigma$  is the activation function.

### 3. Results and Discussion

Table 1. The Classification of GRU and Wavelet Level-2

Feature Extraction Methods	Accuracy	Precision	Recall	Specificity	$G_{mean}$	Feature Extraction Time (s)
db4	0.92	1	0.84	1	0.9165	0.0022
db8	0.92	1	0.84	1	0.9165	0.0036
sym4	0.92	1	0.84	1	0.9165	0.0022
sym8	0.92	1	0.84	1	0.9165	0.0035
haar	0.66	1	0.32	1	0.5657	0.0016
coif4	0.86	1	0.72	1	0.8485	0.0066
coif8	0.92	1	0.84	1	0.9165	0.0179
bior4.4	0.64	1	0.28	1	0.5292	0.0026
bior6.8	0.92	1	0.84	1	0.9165	0.0066
rbio 4.4	0.7	1	0.4	1	0.6325	0.8574
rbio 6.8	0.92	1	0.84	1	0.9165	0.0037
dmey	0.86	1	0.72	1	0.8485	0.1908

Table 2. The Classification of GRU and Wavelet Level-3

Feature Extraction Methods	Accuracy	Precision	Recall	Specificity	$G_{mean}$	Feature Extraction Time (s)
db4	0.92	1	0.84	1	0.9165	0.0023
db8	0.92	1	0.84	1	0.9165	0.0037
sym4	0.92	1	0.84	1	0.9165	0.0023
sym8	0.92	1	0.84	1	0.9165	0.0035
haar	0.92	1	0.84	1	0.9165	0.0016
coif4	0.92	1	0.84	1	0.9165	0.0055
coif8	0.96	1	0.92	1	0.9592	0.0147
bior4.4	0.92	1	0.84	1	0.9165	0.0026
bior6.8	0.92	1	0.84	1	0.9165	0.0039
rbio 4.4	0.92	1	0.84	1	0.9165	0.0025
rbio 6.8	0.92	1	0.84	1	0.9165	0.0071
dmey	0.96	1	0.92	1	0.9592	0.0231

This research uses the dataset of 200 movie videos consisting of 100 scenes of violence and 100 non-violence movies [11, 37]; the following is the training process flow using the GRU algorithm. Each video is captured and resized as many as 20 images measuring 224×224 pixels. After that, the captured image was extracted using various types of multilevel Discrete Wavelet Transform, including Haar, Daubechies, Symlets, Coiflets, Biorthogonal, Reverse biorthogonal, and Discrete Meyer. The levels used in this study are levels 2, 3, 4, and 5. After that, split the data into 160 training data and 40 test data. This data is processed in the GRU network. The following are the results of the classification with GRU and DWT at level 2 as feature extraction.

Based on Table 1, the classification using GRU and wavelet level 2 gets the best accuracy of 0.92

with an average of 0.8466. For the  $G_{mean}$  value, the best score is 0.9165, with an average  $G_{mean}$  score of 0.82. The best accuracy and  $G_{mean}$  values were obtained using the wavelets db4, db8, sym4, sym8, bior6.8, and rbio6.8. The following are the results of the classification with GRU and DWT level 3 as extraction features. In Table 2, the classification using GRU and wavelet level 3 gets the best accuracy of 0.96 with an average accuracy of 0.9267. For the  $G_{mean}$  value, the best score is 0.9592, with an average  $G_{mean}$  score of 0.9236. The best accuracy and  $G_{mean}$  values were obtained using the bior4.4 and dmey wavelets.

Table 3. The Classification of GRU and Wavelet Level-4

Feature Extraction Methods	Accuracy	Precision	Recall	Specificity	$G_{mean}$	Feature Extraction Time (s)
db4	0.92	1	0.84	1	0.9165	0.0025
db8	0.92	1	0.84	1	0.9165	0.0038
sym4	0.92	1	0.84	1	0.9165	0.0025
<b>sym8</b>	<b>0.96</b>	<b>1</b>	<b>0.92</b>	<b>1</b>	<b>0.9592</b>	<b>0.0075</b>
haar	0.92	1	0.84	1	0.9165	0.0016
coif4	0.92	1	0.84	1	0.9165	0.0098
<b>coif8</b>	<b>0.96</b>	<b>1</b>	<b>0.92</b>	<b>1</b>	<b>0.9592</b>	<b>0.0211</b>
bior4.4	0.92	1	0.84	1	0.9165	0.003
bior6.8	0.92	1	0.84	1	0.9165	0.0043
rbio 4.4	0.92	1	0.84	1	0.9165	0.0029
rbio 6.8	0.92	1	0.84	1	0.9165	0.007
<b>dmey</b>	<b>0.96</b>	<b>1</b>	<b>0.92</b>	<b>1</b>	<b>0.9592</b>	<b>0.0292</b>

Classification results with GRU and DWT level 4 as extraction features can be seen in Table 3. Based on Table 3, the classification using GRU and wavelet level 4 gets the best accuracy of 0.96 with an average accuracy of 0.93. For the  $G_{mean}$  value, the best score is 0.9592, with an average  $G_{mean}$  score of 0.9272. The best accuracy and  $G_{mean}$  values were obtained using the sym8, coif8, and dmey wavelets. The classification results with GRU and wavelet transform level 5 as feature extraction can be seen in Table 4. Table 4 shows the best accuracy and  $G_{mean}$  values when using the wavelets coif8 and dmey. The coif8 shows the best result because Coiflet efficiently distinguishes local features and better approximates signals.

Table 4. The Classification of GRU and Wavelet Level-5

Feature Extraction Methods	Accuracy	Precision	Recall	Specificity	$G_{mean}$	Feature Extraction Time (s)
db4	0.92	1	0.84	1	0.9165	0.0026
db8	0.92	1	0.84	1	0.9165	0.0041
sym4	0.92	1	0.84	1	0.9165	0.0025
sym8	0.92	1	0.84	1	0.9165	0.0063
haar	0.92	1	0.84	1	0.9165	0.0018
coif4	0.92	1	0.84	1	0.9165	0.0061
<b>coif8</b>	<b>0.96</b>	<b>1</b>	<b>0.92</b>	<b>1</b>	<b>0.9592</b>	<b>0.0238</b>
bior4.4	0.92	1	0.84	1	0.9165	0.0029
bior6.8	0.92	1	0.84	1	0.9165	0.0044
rbio 4.4	0.92	1	0.84	1	0.9165	0.003
rbio 6.8	0.92	1	0.84	1	0.9165	0.0046
<b>dmey</b>	<b>0.96</b>	<b>1</b>	<b>0.92</b>	<b>1</b>	<b>0.9592</b>	<b>0.0302</b>

Training time can be seen in Fig. 6 and testing time in Fig. 7. The fastest training time for wavelet level 2 is obtained using bior4.4, which is 113.36 seconds with an average training time of 176.651. The fastest level 3 wavelet training time is obtained using 'sym4', which is 30.516 seconds with an average training time of 57.622. In wavelet level 4, the fastest training time is obtained when using 'sym4', which is 20.564 seconds with an average training time of 38.314. And the quickest training time for wavelet level-5 is obtained using bior4.4, which is 22.444 seconds with an average training time of 43,505.

A comparison of the results of the accuracy of the proposed model with previous research is shown in Table 5. Fig. 7 shows that the fastest testing time for wavelet level 2 is obtained when using 'haar,' 0.749 seconds with an average test time of 1.015. For wavelet level-3 the fastest testing time is obtained when using 'db4', which is 0.583 seconds with an average test time of 0.709. For Wavelet level 4, the shortest testing time is obtained when using 'Rbio6.8', which is 0.138 seconds with an average test

time of 0.589. The fastest testing time for wavelet level-5 is obtained when using 'bior4.4', which is 0.0098 seconds with an average test time of 0.483.

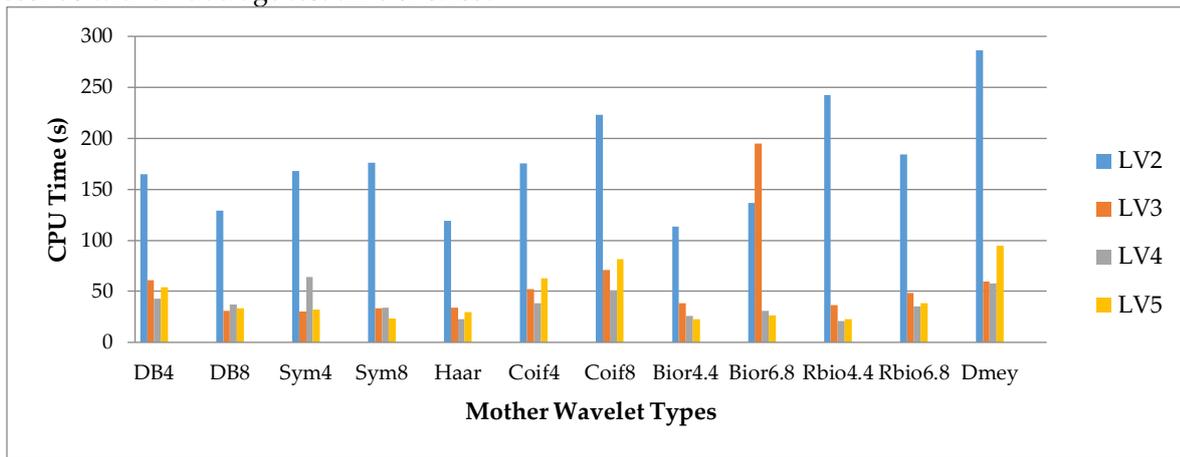


Fig. 6. CPU Time Training Phase GRU

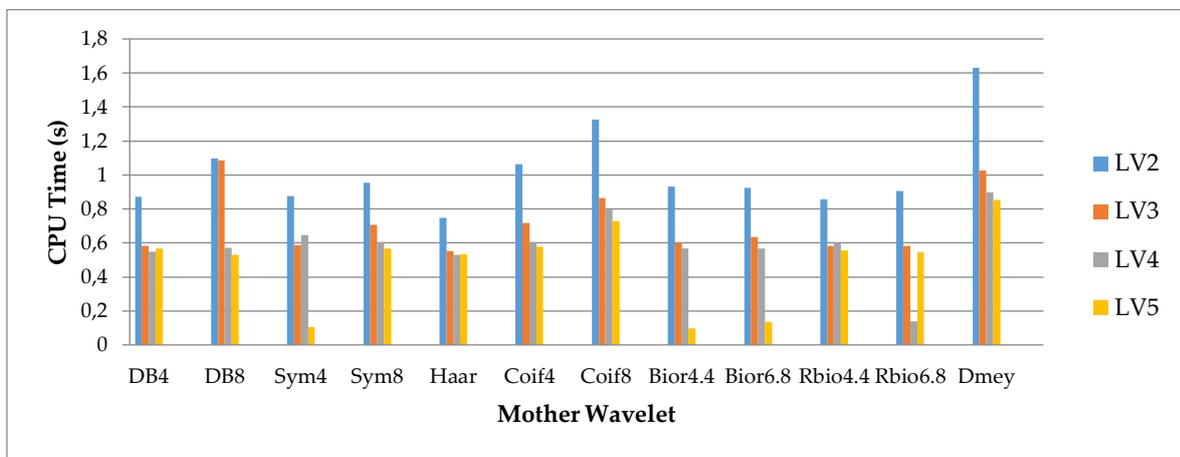


Fig. 7 CPU Time Testing Phase GRU

Table 5. Accuracy between related works and proposed method

Method	Accuracy	Dataset
ResNet50 +LSTM [21]	0.8874	Movie
VGG19 +LSTM [21]	0.8486	Movie
Xception +LSTM [21]	0.8357	Movie
Proposed: DWT lv-5 'coif8' +GRU	0.96	Movie

Based on Table 5, our proposed model has outstanding accuracy compared to the model proposed in [20]. The several advantages of Coiflets are more symmetrical; therefore, it can give a better approximation in the study of symmetrical signals. The presence of zero moments of the scaling function in Coiflets leads to better compressibility [38]. Besides that, Coiflets can efficiently distinguish local features of images since local features aim to detect the interest points in an image and describe them by a set of vectors [39]. With this capability, a collection of images from videos can be appropriately classified.

The study in [20] uses several deep CNN architectures, namely VGG, ResNet, and Xception combined with LSTM as a classifier. It makes Coiflets more precise in extracting features and provides better performance among mother wavelets. The best accuracy is obtained when using a combination of ResNet50 and LSTM, which is 0.8874. This value is still lacking compared to the best accuracy we got, which is 0.96. The difference in accuracy obtained is quite significant, namely 0.073-0.125.

Apart from the accuracy, our proposed method also answers the previous state-of-the-art regarding the running time in detecting violence on video. The use of deep learning architecture provides good accuracy in detecting violence on video. However, these advantages are accompanied by the running time required, as in [16], which uses one of the deep learning architectures, VGG16

combined with SVM. The best accuracy obtained is 0.945, but it takes quite a lot of time. It takes 7852.86 seconds to train a dataset consisting of 1000 videos, while our proposed model only requires an average of 38.314 seconds to train a dataset with 200 videos using level 3 wavelets. The time required to test using level 5 of coif8 wavelet is only 0.728 seconds per data. The results are very far compared to the time required in [16], which is 10.21 seconds per data. These advantages provide an opportunity to apply this model to future works, namely real-time detection in live streaming videos. The study shows the same result as Imran et al. research, which says the GRU algorithm has a strong performance [30]. GRU shows the best results in the problem of detecting violent scenes with eight levels five wavelet coif feature extraction, both in terms of accuracy and time. In general, the GRU shows promising results for a wide variety of mother wavelets.

#### 4. Conclusion

Detection of violent scenes in movies can be done automatically and quickly with the help of artificial intelligence. The combination of the GRU algorithm with variations of the mother wavelet for feature extraction shows that the GRU is quite robust. The GRU model is more straightforward than other RNN models because it has two gates with fewer parameters; therefore, it is computationally more efficient and can speed up the process. Even though it only has two gates, GRU still controls the flow of information to provide excellent performance.

The automatic detection of violence with mother wavelets Haar, Daubechies, Symlets, Coiflets, Biorthogonal, Reverse biorthogonal, and Discrete Meyer showed the best accuracy obtained with mother wavelets Coiflets and Discrete Meyer with an accuracy of 0.96. In terms of time, between Coiflets and Discrete Meyer, Coiflets showed the best performance since it is more symmetrical, efficiently distinguishes local features, and provides a better approximation and compressibility. Because of these characteristics, the Coiflets become more precise and offer the best performance for the mother wavelet. This research does deliver good performance, but there are still many challenges that can be continued for further study, namely detecting live streaming videos because there are many live broadcasts on online channels.

#### Author Contributions

E. M. Imah: Conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, and writing - review & editing. I. K. Laksono: Data curation, investigation, resources, software, validation, visualization, and writing - original draft. K. Karisma: Investigation, visualization, and writing - review & editing. A. Wintarti: Investigation, project administration, and writing - review & editing.

#### Acknowledgment

This research was supported by the Grant of Higher Education Excellence Research on Applied Research with approval no. B/12126/UN38.9/LK.04.00/2021 and No: 193/SP2H/AMD/LT/DRPM/2021, funded by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia. Implementation of research in the second year of DRPM funding in 2022.

#### Declaration of Competing Interest

We declare that we have no conflict of interest.

#### References

- [1] WHO, "Global status report on preventing violence against children 2020," World Health Organization, 2020.
- [2] M. Lawson, M. H. Piel and M. Simon, "Child Maltreatment during the COVID-19 Pandemic: Consequences of Parental Job Loss on Psychological and Physical Abuse Towards Children," *Child Abuse & Neglect*, vol. 110, 2020.
- [3] H. H. Fore, "Violence against children in the time of COVID-19: What we have learned, what remains unknown and the opportunities that lie ahead," *Child Abuse & Neglect*, vol. 116, 2021.
- [4] C. Øverlien, "The COVID-19 Pandemic and Its Impact on Children in Domestic Violence Refuges," *Child Abuse Review*, vol. 29, no. 4, pp. 379-386, 2020.

- [5] R. S. McGinnis, E. W. McGinnis, J. Hruschak, N. L. LopezDuran, K. Fitzgerald, K. L. Rosenblum and M. Muzik, "Rapid detection of internalizing diagnosis in young children enabled by wearable sensors and machine learning," *PLoS ONE*, vol. 14, no. 1, p. e0210267, 2019.
- [6] R. Durham and P. Wilkinson, "Joker: how 'entertaining' films may affect public attitudes towards mental illness – psychiatry in movies," *The British Journal of Psychiatry*, vol. 216, no. 6, 2020.
- [7] T. Han, J. Zhang, Z. Zhang, G. Sun, L. Ye, H. Ferdinando, E. Alasaarela, T. Seppänen, X. Yu and S. Yang, "Emotion recognition and school violence detection from children speech," *Journal on Wireless Communications and Networking*, vol. 235, no. 2018, 2018.
- [8] H. Yao and X. Hu, "A survey of video violence detection," *Cyber-Physical Systems*, 2021.
- [9] I. P. Febin, K. Jayasree and P. T. Joy, "Violence detection in videos for an intelligent surveillance system using MoBSIFT and movement filtering algorithm," *Pattern Analysis and Applications*, vol. 23, p. 611–623, 2020.
- [10] W. Song, D. Zhang, X. Zhao, J. Yu, R. Zheng and A. Wang, "A Novel Violent Video Detection Scheme Based on Modified 3D Convolutional Neural Networks," *IEEE Access*, vol. 7, pp. 39172-39179, 2019.
- [11] A. Jain and D. K. V. A. J. a. D. K. Vishwakarma, "Deep NeuralNet For Violence Detection Using Motion Features From Dynamic Images," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India, 2020.
- [12] A.-M. R. Abdali and R. F. Al-Tuma, "Robust Real-Time Violence Detection in Video Using CNN And LSTM," in *2019 2nd Scientific Conference of Computer Sciences (SCCS)*, Baghdad, Iraq, 2019.
- [13] J. Mahmoodi and A. Salajeghe, "A classification method based on optical flow for violence detection," *Expert Systems with Applications*, vol. 127, pp. 121-127, 2019.
- [14] K. Jani, M. Chaudhuri, H. Patel and M. Shah, "Machine learning in films: an approach towards automation in film censoring," *J. of Data, Inf. and Manag.*, vol. 2, p. 55–64, 2020.
- [15] D. J. Samuel, E. Fenil, G. Manogaran, G. Vivekananda, T. Thanjaivadivel, S. Jeeva and A. Ahilan, "Real time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM," *Computer Networks*, vol. 151, pp. 191-200, 2019.
- [16] K. Karisma, E. M. Imah and A. Wintarti, "Violence Classification Using Support Vector Machine and Deep Transfer Learning Feature Extraction," in *2021 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, Surabaya, Indonesia, 2021.
- [17] K. Deepak, L. Vignesh and S. Chandrakala, "Autocorrelation of gradients based violence detection in surveillance videos," *ICT Express*, vol. 6, no. 3, pp. 155-159, 2020.
- [18] P. Wang, P. Wang and E. Fan, "Violence detection and face recognition based on deep learning," *Pattern Recognition Letters*, vol. 142, pp. 20-24, 2021.
- [19] A. Sen and K. Deb, "Categorization of actions in soccer videos using a combination of transfer learning and Gated Recurrent Unit," *ICT Express*, 2021.
- [20] F. Harrou, Y. Sun, A. S. Hering, M. Madakyaru and A. Dairi, "Chapter 7 - Unsupervised recurrent deep learning scheme for process monitoring," in *Statistical Process Monitoring Using Advanced Data-Driven and Deep Learning Approaches*, Elsevier, 2021, pp. 225-253.
- [21] E. Ditsanthia, L. Pipanmaekaporn and S. Kamonsantiroj, "Video Representation Learning for CCTV-Based Violence Detection," in *2018 3rd Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*, Bangkok, Thailand, 2018.
- [22] A. Nurhopipah and A. Harjoko, "Motion Detection and Face Recognition For CCTV Surveillance System," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 12, no. 2, pp. 107-118, 2018.
- [23] E. M. Imah, W. Jatmiko and T. Basaruddin, "Adaptive Multilayer Generalized Learning Vector Quantization (AMGLVQ) as new algorithm with integrating feature extraction and classification for Arrhythmia heartbeats classification," in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Seoul, Korea, 2012.

- [24] A. H. Masruroh, E. M. Imah and E. Rahmawati, "Classification of Emotional State Based on EEG Signal using AMGLVQ," *Procedia Computer Science*, vol. 157, pp. 552-559, 2019.
- [25] W. Zhang and S. Xiang, "Face anti-spoofing detection based on DWT-LBP-DCT features," *Signal Processing: Image Communication*, vol. 89, 2020.
- [26] J. Wang and Z. Du, "A method of processing color image watermarking based on the Haar wavelet," *Journal of Visual Communication and Image Representation*, vol. 64, 2019.
- [27] R. Chatterjee and R. Halder, "Discrete Wavelet Transform for CNN-BiLSTM-Based Violence Detection," in *Advances in Systems, Control and Automations: Select Proceedings of ETAEERE 2020*, Singapore, 2021.
- [28] F. Deeba, S. Kun, F. A. Dharejo and Y. Zhou, "Wavelet-Based Enhanced Medical Image Super Resolution," *IEEE Access*, vol. 8, pp. 37035-37044, 2020.
- [29] R. Chatterjee and R. Halder, "Discrete Wavelet Transform for CNN-BiLSTM-Based Violence Detection," in *Advances in Systems, Control and Automations: Select Proceedings of ETAEERE 2020*, Singapore, 2021.
- [30] J. Imran, B. Raman and A. S. Rajput, "Robust, efficient and privacy-preserving violent activity recognition in videos," in *SAC '20: Proceedings of the 35th Annual ACM Symposium on Applied Computing*, Brno Czech Republic, 2020.
- [31] I. S. Gracia, O. D. Suarez, G. B. Garcia and T.-K. Kim, "Fast Fight Detection," *PLoS ONE*, vol. 10, no. 4, 2015.
- [32] J. Panyavaraporn and P. Horkaew, "DWT/DCT-based Invisible Digital Watermarking Scheme for Video Stream," in *2018 10th International Conference on Knowledge and Smart Technology (KST)*, Chiang Mai, Thailand, 2018.
- [33] D. Zhang, "Wavelet Transform," in *Fundamentals of Image Data Mining*, Cham, Springer, 2019, pp. 35-44.
- [34] D. Popov, A. Gapochkin and A. Nekrasov, "An Algorithm of Daubechies Wavelet Transform in the Final Field When Processing Speech Signals," *Electronics*, vol. 7, no. 7, p. 120, 2018.
- [35] D. Zhang, "Windowed Fourier Transform," in *Fundamentals of Image Data Mining*, Cham, Springer, 2019, pp. 25-34.
- [36] S. A. Sumon, R. Goni, N. B. Hashem, T. Shahria and R. M. Rahman, "Violence Detection by Pretrained Modules with Different Deep Learning Approaches," *Vietnam Journal of Computer Science*, vol. 7, no. 1, p. 19-40, 2020.
- [37] C.-H. Demarty, C. Penet, M. Soleymani and G. Gravier, "VSD, a public dataset for the detection of violent scenes in movies: design, annotation, analysis and evaluation," *Multimedia Tools and Applications*, vol. 74, p. 7379-7404, 2015.
- [38] V. Varuikhin and A. Levina, "Continuous Wavelet Transform Applications In Steganography," *Procedia Computer Science*, vol. 186, pp. 580-587, 2021.
- [39] L. Kabbai, M. Abdellaoui and A. Douik, "Image classification by combining local and global features," *The Visual Computer*, vol. 35, p. 679-693, 2019.