International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I7P226

Volume 70 Issue 7, 260-268, July 2022
© 2022 Seventh Sense Research Group®

*Original Article*

# Experimental Evaluation of Resampling Algorithms on the Imbalance Violence Video Detection

Moch Arief Soeleman[1], Catur Supriyanto[2], Dwi Puji Prabowo[3], Pulung Nurtantio Andono[4]

*Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia.*

[1]arief22208@gmail.com

Received: 12 May 2022 Revised: 03 July 2022 Accepted: 11 July 2022 Published: 25 July 2022

***Abstract** - Violence detection is part of the video surveillance research area and has played an important role in the last decade. Convolution Neural Network (CNN) has become a very successful classifier for violence video detection. The learned features of CNN give a superior result over the handcrafted features of traditional machine learning. Long Short-Term Memory (LSTM) layer process the learned features to capture the temporal dependencies. Violence video detection is a binary classification that categorizes the instance video into violence or non-violence. However, the number of video clips in each class is not balanced, which makes it hard to collect the positive class. In this direction, this work presents the empirical results of resampling techniques to enhance the performance of video violence detection. This work compares four resampling techniques Random Under Sampling (RUS, Synthetic Minority Oversampling Technique (SMOTE), Random Over Sampling (ROS), and the combination of SMOTE and RUS. The experiments are conducted on two popular benchmark datasets, Hockey and Crowd Datasets. The number of positive classes of these datasets is reduced to create an imbalance of datasets for experimental purposes. The experiment results demonstrated that RUS produced superior performance compared to the other resampling techniques in terms of G-means and AUC.*

***Keywords** - Convolution Neural Network (CNN), Imbalance dataset, Resampling algorithm, Long Short-Term Memory (LSTM), Violence video detection.*

## 1. Introduction

Nowadays, the amount of violence in the public area is increasing. It can create discomfort in public places. Violence also occurs because of the limited number of security personnel. Therefore, automatic violence video detection can help to monitor and prevent violence in the public area. Pujol et al. [1] developed a system for detecting violence in video. The video input is obtained from an installed camera, smartphone, or YouTube video. Real-time violence detection also has been developed for football stadiums [2]. The frames are categorized into the violence model, negative model, and human part model. A histogram of Oriented Gradients (HOG) was used to extract the features.

Recently proposed methods for video violence detection can be categorized into handcrafted and learned features. The handcrafted features can be determined before feature extraction. For example, the generated features from Gray Level Co-occurrence Matrix (GLCM) are entropy, contrast, and homogeneity. Meanwhile, the learned features cannot be determined before feature extraction.

In violence video detection, some studies capture the handcrafted features using some spatiotemporal descriptor, such as MoSIFT [3], Motion Weber Local Descriptor (MoWLD) [4], OViF [5], LHOG+LHOF [6], and DiMOLIF [7], Spatio Temporal Autocorrelation of Gradients (STACOG) [8]. Baba et al. [9] proposed another work's cascade approach to separate temporal and spatial features. In violence video detection, some researchers [10][11][12][13] show that the features of convolution neural network (CNN) become the potential features compared to the handcrafted features. Although the learned features of CNN are superior to the handcrafted features, CNN has high computational complexity and needs a large number of data to train the network [14]. In image classification, Wang et al. [15] show that CNN is not better than a support vector machine (SVM) in terms of accuracy when using a small sample dataset. CNN extracts the spatial and motion features of the object through a long short-term memory (LSTM) network. CNN does not need additional feature extraction since the feature extraction and classification have been included in the structure of CNN.

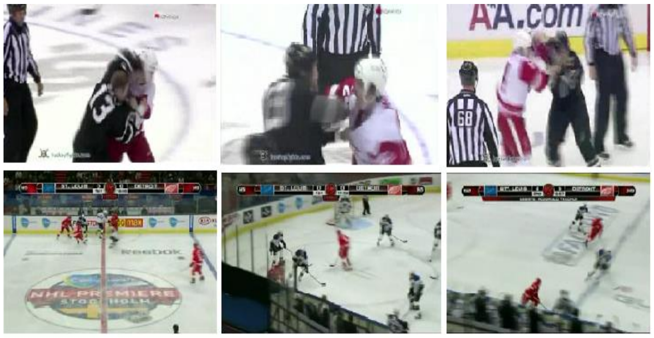

**Fig. 1 Hockey dataset, first row for violence and second row for non-violent.**

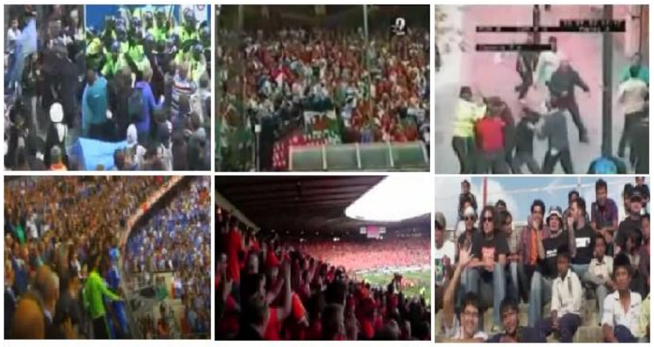

**Fig. 2 Crowd dataset, first row for violence and second row for non-violent.**

In CNN, some pretrained models exist for classification or detection purposes, such as Resnet50, Xception, and InceptionV3. Bintang and Kusuma show that the performance of the pretrained model is more accurate than the simple CNN [16]. Xiao et al. [17] compared these pretrained models in ultrasonic breast mass discrimination. The results show that InceptionV3 produces better performance in terms of accuracy. In our previous work, InceptionV3 also performs better violence video detection [18]. These pretrained models can be utilized to classify the new datasets. Although the number of datasets is small, the classification can use transfer learning instead of building a CNN model from scratch.

In video violence detection, data in the positive class is hard to be collected. Therefore, the dataset becomes imbalanced, and the CNN performs poorly on the minority classes. The imbalance of datasets also becomes a serious problem in other research areas, such as the automatic detection of arrhythmia [19], credit scoring analysis [20], and software defect prediction [21]. Normally, the classifier needs a balanced dataset in training to classify the testing dataset accurately. Therefore, this paper aims to evaluate the resampling techniques for video violence detection, such as ROS, RUS, SMOTE, and SMOTE+RUS.

The resampling technique is a pre-processing that balance the label or class proportions on an imbalanced dataset. There are two main procedures in the resampling techniques, increase the size of minority examples (oversampling) or decrease the size of majority examples (undersampling). The evaluation of this study contains both over-sampling, under-sampling, and hybrid strategies. Random Under Sampling (RUS) [22] is a resampling technique that removes the majority of class examples. RUS can balance the examples by using the imbalance ratio (IR). Random Over Sampling (ROS) [22]

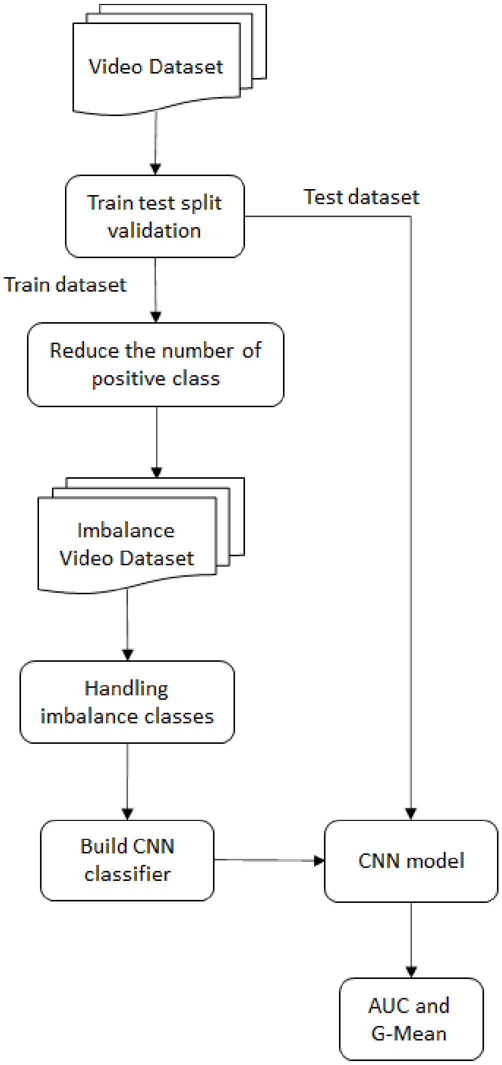

**Fig. 3 The proposed model**

| | | Actual | |
|---|---|---|---|
| | | Violence | Non-violence |
| Predicted | Violence | TP | FN |
| | Non-violence | FP | TN |

**Fig. 4 Confusion Matrix**

Consists of duplicating the examples in the minority class. The example duplication will reduce the number of IR in the dataset. Synthetic Minority Oversampling Technique (SMOTE) [23] is like ROS, increasing the number of minority classes by creating synthetic examples. SMOTE generates the synthetic examples between the line of two minority examples. Hybrid resampling [22] combines the two previous techniques, under-sampling, and under-sampling. Our work combines SMOTE and RUS to balance the dataset.

**Table 1. Information on Imbalance Hockey Dataset**

| Dataset | # Violence | # Non-Violence | IR |
|---|---|---|---|
| Low imbalance $IR \leq 9$ | | | |
| Hockey 1 | 350 | 400 | 1.14 |
| Hockey 2 | 300 | 400 | 1.33 |
| Hockey 3 | 250 | 400 | 1.6 |
| Hockey 4 | 200 | 400 | 2 |
| Hockey 5 | 150 | 400 | 2.67 |
| Hockey 6 | 100 | 400 | 4 |
| Hockey 7 | 75 | 400 | 5.33 |
| Hockey 8 | 50 | 400 | 8 |
| Hockey 9 | 45 | 400 | 8.89 |
| Medium imbalance $9 < IR \leq 20$ | | | |
| Hockey 10 | 40 | 400 | 10 |
| Hockey 11 | 35 | 400 | 11.43 |
| Hockey 12 | 30 | 400 | 13.33 |
| Hockey 13 | 25 | 400 | 16 |
| Hockey 14 | 20 | 400 | 20 |
| High imbalance $IR > 20$ | | | |
| Hockey 15 | 15 | 400 | 26.67 |
| Hockey 16 | 10 | 400 | 40 |

**Table 2. Information on Imbalance Crowd Dataset**

| Dataset | # Violence | # Non-Violence | IR |
|---|---|---|---|
| Low imbalance $IR \leq 9$ | | | |
| Crowd 1 | 80 | 100 | 1.25 |
| Crowd 2 | 60 | 100 | 1.67 |
| Crowd 3 | 40 | 100 | 2.5 |
| Crowd 4 | 20 | 100 | 5 |
| Medium imbalance $9 < IR \leq 20$ | | | |
| Crowd 5 | 10 | 100 | 10 |
| Crowd 6 | 5 | 100 | 20 |
| High imbalance $IR > 20$ | | | |
| Crowd 7 | 3 | 100 | 33.33 |

## 2. Methods

### *2.1. Experimental Data*

This study uses two popular datasets in violence video detection to demonstrate the potency of resampling techniques in CNN LSTM. The datasets consist of violence and non-violence classes. The first dataset is Hockey Fight containing 1000 video clips. Each video has a resolution of 360×288 pixels. There are 500 videos in each class. The violent crowd as the second dataset consists of 246 videos. In the violent crowd, there are 123 videos in each class. Figures 1 and 2 show the samples clip of Hockey and crowd video violence datasets.

### *2.2. Experimental Design and Parameters Setting*

Figure 3 presents the process of the experiment. To evaluate the performance of resampling techniques, this study follows the split validation as referring to previous work [11]. For that reason, our study divided the dataset into training and testing with a ratio of 80:20. To evaluate the proposed model for an imbalanced dataset, this study creates the imbalanced datasets from the training dataset with several imbalance ratios (IR). This study followed Fan et al. [24] to decide the IR of the dataset.

Fan et al. [24] classified the IR into three levels of imbalance, low, medium, and high. Based on the level of imbalance, this research reduced the number of violent videos. Table 1 and Table 2 show the imbalance proportion of Hockey and Crowd. Then, the generated imbalance datasets were implemented in our model. This study uses InceptionV3 as the pretrained model for our CNN. To evaluate the proposed model, this study repeated 10 experiments to collect the average performances. The experiment was executed with Keras library on python programming. The number of epochs for the neural network is set to 200.

As discussed in the Introduction, the experiment adopted InceptionV3 as the pretrained network. The structured network of InceptionV3 extracts the features of each image. Our pretrained network has been trained on the ImageNet dataset. ImageNet contains more than 14 million images belonging to more than 20.000 classes. The experiment resizes each frame to 299x299 to feed the InceptionV3 model.

### *2.3. Metric Evaluation*

This section presents the two standard metrics to evaluate the classification performance when the dataset is unbalanced. G-mean and Area Under Curve (AUC) scores are the most frequent metrics for this purpose. High AUC and G-mean mean that the detection has high performance. The evaluation adopted these metrics based on several works of literature [25]. The two metrics are calculated from the confusion matrix, which presents the number of correct and incorrect classification tasks, as shown in Figure 4.

Geo metric-mean (G-mean) is a suitable metric for the imbalance problem and is computed by Equation 1.

$$G - Mean = \sqrt{\frac{TP}{TP+FN} \times \frac{TN}{TN+FP}} \qquad (1)$$

Where $TP, FP, FN, TN$ represent the number of true positive, false positive, false negative, and true negative.

**Table 3. Reported G-mean of Hockey Dataset. The color value is the best performance in each row. The value of $\tau_F$ is $4.45$, and the critical value of $F(4,60) = 2.04$.**

| Data ID | IR | Without Sampling | ROS | SMOTE | RUS | SMOTE+RUS |
|---|---|---|---|---|---|---|
| Hockey 1 | 1.14 | 0.9066 (1) | 0.9012 (2) | 0.8991 (3) | 0.8376 (5) | 0.8893 (4) |
| Hockey 2 | 1.33 | 0.905 (3) | 0.9057 (2) | 0.9117 (1) | 0.8411 (5) | 0.8959 (4) |
| Hockey 3 | 1.60 | 0.8991 (2) | 0.9032 (1) | 0.8925 (3) | 0.8382 (5) | 0.8889 (4) |
| Hockey 4 | 2.00 | 0.8807 (4) | 0.9037 (1) | 0.9024 (2) | 0.8564 (5) | 0.8865 (3) |
| Hockey 5 | 2.67 | 0.8714 (4) | 0.8857 (2) | 0.8766 (3) | 0.8542 (5) | 0.8888 (1) |
| Hockey 6 | 4.00 | 0.8926 (1) | 0.8879 (3) | 0.8888 (2) | 0.8532 (5) | 0.8874 (4) |
| Hockey 7 | 5.33 | 0.8369 (5) | 0.8514 (4) | 0.8741 (2) | 0.8826 (1) | 0.862 (3) |
| Hockey 8 | 8.00 | 0.6777 (5) | 0.8076 (2) | 0.7854 (4) | 0.8875 (1) | 0.7958 (3) |
| Hockey 9 | 8.89 | 0.618 (5) | 0.7811 (2) | 0.7245 (4) | 0.8943 (1) | 0.7385 (3) |
| Hockey 10 | 10.00 | 0.4845 (5) | 0.7415 (3) | 0.7481 (2) | 0.8849 (1) | 0.5854 (4) |
| Hockey 11 | 11.43 | 0.2967 (5) | 0.6966 (2) | 0.6827 (3) | 0.8806 (1) | 0.6699 (4) |
| Hockey 12 | 13.33 | 0.4762 (5) | 0.653 (3) | 0.6666 (2) | 0.8896 (1) | 0.5024 (4) |
| Hockey 13 | 16.00 | 0.1764 (4) | 0.2063 (2) | 0.1886 (3) | 0.7518 (1) | 0.1436 (5) |
| Hockey 14 | 20.00 | 0.0141 (5) | 0.1414 (2) | 0.1097 (4) | 0.5923 (1) | 0.1228 (3) |
| Hockey 15 | 26.67 | 0.0313 (5) | 0.1555 (2) | 0.1262 (3) | 0.5096 (1) | 0.0765 (4) |
| Hockey 16 | 40.00 | 0 (5) | 0.1148 (2) | 0.0906 (3) | 0.5783 (1) | 0.0807 (4) |
| Avg. G-mean | | 0.5605 | 0.6585 | 0.6480 | 0.8020 | 0.6197 |
| *AR* | | 4.00 | 2.19 | 2.75 | 2.50 | 3.56 |

**Table 4. Reported G-mean of Crowd Dataset. The color value is the best performance in each row. The value of $\tau_F$ is $3.55$, and the critical value of $F(4,24) = 2.19$.**

| Data ID | IR | Without Sampling | ROS | SMOTE | RUS | SMOTE+RUS |
|---|---|---|---|---|---|---|
| Crowd 1 | 1.25 | 0.9121 (1) | 0.8988 (2) | 0.8856 (3) | 0.8226 (4) | 0.7850 (5) |
| Crowd 2 | 1.67 | 0.8603 (3) | 0.8751 (1) | 0.8738 (2) | 0.8579 (4) | 0.7736 (5) |
| Crowd 3 | 2.50 | 0.8254 (4) | 0.8292 (3) | 0.8306 (2) | 0.8661 (1) | 0.7569 (5) |
| Crowd 4 | 5.00 | 0.5943 (5) | 0.6954 (3) | 0.6999 (2) | 0.7889 (1) | 0.6783 (4) |
| Crowd 5 | 10.00 | 0.5526 (5) | 0.5894 (3) | 0.5827 (4) | 0.8099 (1) | 0.6032 (2) |
| Crowd 6 | 20.00 | 0.2631 (5) | 0.3446 (3) | 0.3642 (2) | 0.7193 (1) | 0.3155 (4) |
| Crowd 7 | 33.33 | 0.0204 (5) | 0.2039 (4) | 0.2213 (2) | 0.7014 (1) | 0.2128 (3) |
| Avg. G-mean | | 0.5755 | 0.6338 | 0.6369 | 0.7952 | 0.5893 |
| *AR* | | 4.00 | 2.71 | 2.43 | 1.86 | 4.00 |

This study used the Friedman test to statistically analyze the performance of the resampling techniques. At a significant level of 0.05, the Bonferroni-Dunn post-hoc test was implemented to evaluate which techniques differ significantly. The Friedman test of this study follows Demsar [25]. The Friedman test is calculated as follows:

$$\tau_{\chi^2} = \frac{12N}{M(M+1)}\left(\sum_{j=1}^{M} AR_j^2 - \frac{M(M+1)^2}{4}\right) \qquad (2)$$

$$AR_j = \frac{1}{N}\sum_{i=1}^{N} R_i^j \qquad (3)$$

Where $N$ indicates the number of imbalance datasets, $M$ indicates the number of detections with and without resampling techniques; AR denotes the average rank of the comparison. After the Friedman test is calculated, its varian $\tau_F$ can be used to conduct the statistical test, which is computed as Equation 4.

**Table 5. Reported AUC of Hockey Dataset. The color value is the best performance in each row. The value of $\tau_F$ is $3.88$, and the critical value of $F(4,60) = 2.04$.**

| Data ID | IR | Without Sampling | ROS | SMOTE | RUS | SMOTE+RUS |
|---|---|---|---|---|---|---|
| Hockey 1 | 1.14 | 0.907 (1) | 0.902 (2) | 0.9 (3) | 0.8455 (5) | 0.89 (4) |
| Hockey 2 | 1.33 | 0.906 (3) | 0.9065 (2) | 0.912 (1) | 0.8485 (5) | 0.8965 (4) |
| Hockey 3 | 1.60 | 0.9 (2) | 0.9035 (1) | 0.894 (3) | 0.846 (5) | 0.89 (4) |
| Hockey 4 | 2.00 | 0.883 (4) | 0.904 (1) | 0.903 (2) | 0.862 (5) | 0.8885 (3) |
| Hockey 5 | 2.67 | 0.8735 (4) | 0.8885 (2) | 0.881 (3) | 0.86 (5) | 0.891 (1) |
| Hockey 6 | 4.00 | 0.894 (1) | 0.8895 (4) | 0.89 (2.5) | 0.859 (5) | 0.89 (2.5) |
| Hockey 7 | 5.33 | 0.845 (5) | 0.8565 (4) | 0.8765 (2) | 0.8855 (1) | 0.8665 (3) |
| Hockey 8 | 8.00 | 0.754 (5) | 0.824 (2) | 0.807 (4) | 0.89 (1) | 0.812 (3) |
| Hockey 9 | 8.89 | 0.735 (5) | 0.803 (2) | 0.7615 (4) | 0.896 (1) | 0.7695 (3) |
| Hockey 10 | 10.00 | 0.6645 (5) | 0.774 (3) | 0.776 (2) | 0.8855 (1) | 0.67 (4) |
| Hockey 11 | 11.43 | 0.599 (5) | 0.7405 (2) | 0.7305 (3) | 0.8825 (1) | 0.7215 (4) |
| Hockey 12 | 13.33 | 0.661 (4) | 0.713 (3) | 0.7215 (2) | 0.8905 (1) | 0.624 (5) |
| Hockey 13 | 16.00 | 0.049 (5) | 0.5215 (2) | 0.519 (3) | 0.7685 (1) | 0.5105 (4) |
| Hockey 14 | 20.00 | 0.5005 (5) | 0.5095 (2) | 0.508 (4) | 0.6585 (1) | 0.5085 (3) |
| Hockey 15 | 26.67 | 0.5015 (5) | 0.5105 (2) | 0.508 (3) | 0.6235 (1) | 0.5045 (4) |
| Hockey 16 | 40.00 | 0.5 (5) | 0.507 (2) | 0.5055 (3.5) | 0.6575 (1) | 0.5055 (3.5) |
| Avg. AUC | | 0.6983 | 0.7596 | 0.7558 | 0.8224 | 0.7400 |
| $AR$ | | 4.00 | 2.25 | 2.81 | 2.50 | 3.44 |

**Table 6. Reported AUC of Crowd Dataset. The color value is the best performance in each row. The value of $\tau_F$ is $4.71$, and the critical value of $F(4,24) = 2.19$.**

| CROWD | IR | Without Sampling | ROS | SMOTE | RUS | SMOTE+RUS |
|---|---|---|---|---|---|---|
| Crowd 1 | 1.25 | 0.913 (1) | 0.9 (2) | 0.887 (3) | 0.837 (4) | 0.8 (5) |
| Crowd 2 | 1.67 | 0.8652 (4) | 0.8783 (1) | 0.8761 (2) | 0.8674 (3) | 0.7891 (5) |
| Crowd 3 | 2.50 | 0.8348 (4) | 0.837 (3) | 0.8391 (2) | 0.8696 (1) | 0.7783 (5) |
| Crowd 4 | 5.00 | 0.6848 (5) | 0.7239 (3) | 0.7261 (2) | 0.7891 (1) | 0.7152 (4) |
| Crowd 5 | 10.00 | 0.6326 (4) | 0.6478 (2) | 0.6435 (3) | 0.6236 (5) | 0.6652 (1) |
| Crowd 6 | 20.00 | 0.5283 (5) | 0.5457 (3) | 0.55 (2) | 0.7413 (1) | 0.537 (4) |
| Crowd 7 | 33.33 | 0.5 (4.5) | 0.5 (4.5) | 0.5065 (2) | 0.7239 (1) | 0.5043 (3) |
| Avg. AUC | | 0.7084 | 0.7190 | 0.7183 | 0.7788 | 0.6842 |
| $AR$ | | 3.93 | 2.64 | 2.29 | 2.29 | 3.86 |

$$\tau_F = \frac{(N-1)\tau_{\chi^2}}{N(M-1)-\tau_{\chi^2}} \qquad (4)$$

The value of $\tau_F$ is compared against the critical value in the F distribution to reject or accept the null hypothesis. If the results reject the null hypothesis, the performance of the methods is significantly different, and then the Bonferroni-Dunn post-hoc test can be conducted.

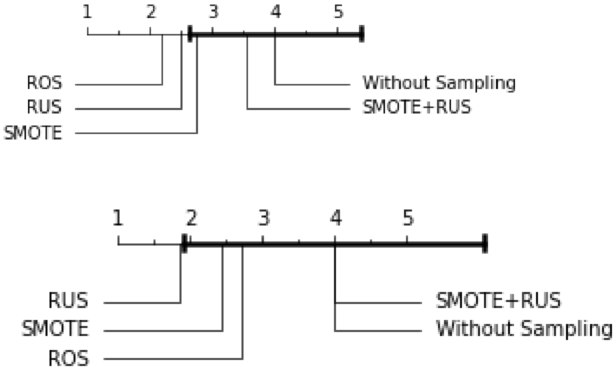

**Fig. 5 Comparison of resampling techniques with the Friedman and Bonferroni-Dunn tests in G-mean. First row for Hockey imbalanced datasets ($CD = 1.37$) and the second row for Crowd imbalanced datasets ($CD = 2.08$).**

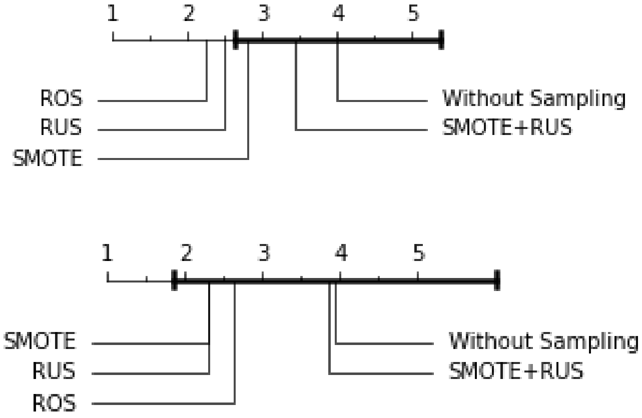

**Fig. 6 Comparison of resampling techniques with the Friedman and Bonferroni-Dunn tests in terms of AUC. First row for Hockey imbalanced datasets ($CD = 1.37$) and the second row for Crowd imbalanced datasets ($CD = 2.08$).**

The Bonferroni-Dunn test reports significant differences between the resampling technique and the control algorithm. The test is applied when the results reject the null hypothesis, which means there is a significant difference between the methods. In this study, the control algorithm refers to detecting without resampling. The critical difference is evaluated as follows:

$$CD = q_\alpha \sqrt{\frac{M(M+1)}{12N}} \qquad (5)$$

If the difference in AR between the resampling technique and the control is larger than the value of $CD$, their performance difference is significant.

## 3. Results and Discussion

This research developed video violence detection with created imbalanced datasets using CNN. There are 23 imbalanced datasets. Tables 3 and 4 show the performance of CNN on Hockey and Crowd datasets based on G-mean. The large number of IR reduces the performance of imbalanced datasets with and without resampling techniques. These experiments noticed that CNN without resampling techniques produced the best performance on 3 low imbalanced datasets (Hockey 1, Hockey 6, and Crowd 1). The two datasets have a small number of IR. Resampling techniques such as ROS, SMOTE, and SMOTE+RUS improve the G-mean only on one or two low imbalanced datasets. RUS results in the best G-mean on 16 imbalance datasets. The compared resampling techniques are significantly different on the high imbalance datasets. Vast improvement happens after applying RUS on high imbalance datasets. In the experiments, SMOTE is the worst technique. In Crowd imbalanced datasets, as shown in Table 4, SMOTE improves the performance of video violence detection, but the performance is not the best result.

The results also show the AUC of the Hockey and crowd imbalance video violence datasets with and without resampling. In Tables 5 and 6, AUC reported the Performance results like G-mean.

**Table 7. AUC comparison with the other approaches**

| Dataset | Related Works | AUC |
|---|---|---|
| Hockey | ViF [27] | 0.8801 |
| | **Ours** | **0.9120** |
| | OViF [5] | 0.9193 |
| | DiMOLIF [7] | 0.9323 |
| | HOMO [28] | 0.9518 |
| | MoWLD [4] | 0.9758 |
| | LHOG+LHOF [6] | 0.9798 |
| Crowd | HOMO [28] | 0.8284 |
| | ViF [27] | 0.8804 |
| | DiMOLIF [7] | 0.8925 |
| | **Ours** | **0.9130** |
| | OViF [5] | 0.9182 |
| | MoWLD [4] | 0.9408 |
| | LHOG+LHOF [6] | 0.9703 |

The AUC of RUS gives the best performance on the most imbalanced datasets. Although the IR is high, the performances of RUS are reduced insignificantly.

Based on Table 3, the Friedman test can be computed as follows:

$$\tau_{\chi^2} = \frac{12 \times 16}{5(5+1)}\left((4^2 + 2.19^2 + 2.75^2 + 2.5^2 + 3.56^2) - \frac{5(5+1)^2}{4}\right) = 14.65$$

$$\tau_F = \frac{(16-1)14.65}{16(5-1) - 14.65} = 4.45$$

With 5 methods and 16 datasets, the degree of freedom is $5 - 1 = 4$ and $(5-1) \times (16-1) = 60$. The critical value of $F(4{,}60)$ for $\alpha = 0.1$ (90%) is 2.04 (See F distribution table). The value of $\tau_F$ is greater than the critical value, so there is a performance difference between the resampling techniques. It also happens in Tables 4, 5, and 6; the value of $\tau_F$ is higher than the critical value. Therefore, the Bonferroni-Dunn test was performed.

Based on the average G-mean and AUC in Table 3-6, RUS performs better than other resampling techniques. However, Figure 5 and 6 shows that ROS has better *AR* on the Hockey imbalanced datasets in term of G-mean and AUC. Moreover, the *AR* of ROS seems to significantly differ from the detection without resampling technique (control method). The *AR* of RUS is higher than ROS in the Hockey imbalanced datasets; it may be caused by the fact that the RUS ranks are the worst resampling technique at the small IR (Hockey 1-6 datasets).

Based on the experiment results, according to Tables 3, 4, 5, and 6, RUS shows the best results in most datasets, followed by ROS, SMOTE, and SMOTE+RUS. The results are consistent with [26], in which the performance of RUS is better than that of ROS and SMOTE. Figures 5 and 6 on Crowd imbalanced datasets show that RUS is the only resampling technique with significant differences from the control algorithm (without resampling technique). The reason may be that the small number of Crowd imbalance datasets causes the difference AR between resampling techniques and the control algorithm to be less than the value $CD$. AUC comparison with the other approaches is reported in Table 7. This study compares the results with ViF, OViF, DiMOLIF, HOMO, MoWLD, and LHOG+LHOF. In the performance view, our approach can achieve better performance than ViF [27] on the Hockey dataset and DiMOLIF [7], ViF [27], and HOMO [28] on the Crowd dataset. Our AUC differs slightly from OViF [5] on Hockey and Crowd. Although our approach CNN+LSTM has a hard computational issue, this study experimented on NVIDIA DGX A100 with no computing problem.

## 4. Conclusion

This paper evaluated four resampling techniques, ROS, SMOTE, RUS, and SMOTE+RUS, on various imbalances. The methods are evaluated on two datasets containing violence and non-violence video clips, namely Hockey and Crowd datasets. RUS achieves the best performance in terms of AUC and G-mean, which is better than the other resampling techniques. Therefore, RUS is suitable for resampling techniques in video violence detection, which imbalance could be the issue in the dataset. In the future, the research plans to propose a resampling technique based on statistic evaluation and evaluate more video violence datasets. The research also can evaluate real-time deep learning applications, such as research in [29] which proposed deep learning to detect the handgun from live video and real-time. The use of cross-validation can be proposed to evaluate the performances based on each fold. The selection of representative samples for training in cross-validation is important to avoid underfitting or overfitting.

## Funding Statement

DRPM-DIKTI Number fully supported this research: 312/E4.1/AK.04.PT/2021. The scheme of applied research for two years, 2021-2022.

## References

[1] F. A. Pujol, H. Mora, and M. L. Pertegal, “A soft computing approach to violence detection in social media for smart cities,” *Soft Comput.*, vol. 24, no. 15, pp. 11007–11017, 2020, doi: 10.1007/s00500-019-04310-x.

[2] D. J. Samuel R. *et al.*, “Real time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM,” *Comput. Netw.*, vol. 151, pp. 191–200, 2019, doi: 10.1016/j.comnet.2019.01.028.

[3] L. Xu, C. Gong, J. Yang, Q. Wu, and L. Yao, “Violent video detection based on MoSIFT feature and sparse coding,” *IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP*, 2014.

[4] T. Zhang, W. Jia, B. Yang, J. Yang, X. He, and Z. Zheng, “MoWLD: a robust motion image descriptor for violence detection,” *Multimed. Tools Appl.*, vol. 76, no. 1, pp. 1419–1438, 2017, doi: 10.1007/s11042-015-3133-0.

[5] Y. Gao, H. Liu, X. Sun, C. Wang, and Y. Liu, “Violence detection using Oriented Violent Flows,” *Image Vis. Comput.*, vol. 48–49, pp. 37–41, Apr. 2016, doi: 10.1016/j.imavis.2016.01.006.

[6] P. Zhou, Q. Ding, H. Luo, and X. Hou, “Violence detection in surveillance video using low-level features,” *PLOS ONE*, vol. 13, no. 10, p. e0203668, Oct. 2018, doi: 10.1371/journal.pone.0203668.

[7] A. Ben Mabrouk and E. Zagrouba, “Abnormal behavior recognition for intelligent video surveillance systems: A review,” *Expert Syst. Appl.*, vol. 91, pp. 480–491, Jan. 2018, doi: 10.1016/j.eswa.2017.09.029.

[8] Deepak K., Vignesh L.K.P., and Chandrakala S., “Autocorrelation of gradients based violence detection in surveillance videos,” *ICT Express*, vol. 6, no. 3, pp. 155–159, Sep. 2020, doi: 10.1016/j.icte.2020.04.014.

[9] M. Baba, V. Gui, C. Cernazanu, and D. Pescaru, “A Sensor Network Approach for Violence Detection in Smart Cities Using Deep Learning,” *Sensors*, vol. 19, no. 7, p. 1676, Apr. 2019, doi: 10.3390/s19071676.

[10] F. U. M. Ullah, A. Ullah, K. Muhammad, I. U. Haq, and S. W. Baik, “Violence Detection Using Spatiotemporal Features with 3D Convolutional Neural Network,” *Sensors*, vol. 19, no. 11, p. 2472, 2019, doi: 10.3390/s19112472.

[11] M. Asad, J. Yang, J. He, P. Shamsolmoali, and X. He, “Multi-frame feature-fusion-based model for violence detection,” *Vis. Comput.*, vol. 37, no. 6, pp. 1415–1431, 2021, doi: 10.1007/s00371-020-01878-6.

[12] 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 J. Comput. Sci.*, vol. 07, no. 01, pp. 19–40, 2020, doi: 10.1142/S2196888820500013.

[13] M. Cheng, K. Cai, and M. Li, “RWF-2000: An Open Large Scale Video Database for Violence Detection,” *ArXiv191105913 Cs*, Oct. 2020, Accessed: Apr. 02, 2022. [Online]. Available: http://arxiv.org/abs/1911.05913

[14] H. Wang *et al.*, “Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features,” *J. Med. Imaging*, vol. 1, no. 3, pp. 034003, 2014, doi: 10.1117/1.JMI.1.3.034003.

[15] P. Wang, E. Fan, and P. Wang, “Comparative analysis of image classification algorithms based on traditional machine learning and deep learning,” *Pattern Recognit. Lett.*, vol. 141, pp. 61–67, 2021, doi: 10.1016/j.patrec.2020.07.042.

[16] I. Bintang and G. P. Kusuma, “Porn Detection in a Video Streaming Using Hybrid Network of CNN and LSTM,” *Int. J. Eng. Trends Technol.*, vol. 69, no. 11, pp. 248–255, 2021, doi: 10.14445/22315381/IJETT-V69I11P231.

[17] T. Xiao, L. Liu, K. Li, W. Qin, S. Yu, and Z. Li, “Comparison of Transferred Deep Neural Networks in Ultrasonic Breast Masses Discrimination,” *BioMed Res. Int.*, p. 10, 2018.

[18] M. A. Soeleman, C. Supriyanto, and D. P. Prabowo, “An Empirical Study of CNN-LSTM on Class Imbalance Datasets for Violence Video Detection,” in *The 2021 International Conference on Computer, Control, Informatics and Its Applications*, Virtual/online conference Indonesia, Oct. 2021, pp. 81–85. doi: 10.1145/3489088.3489126.

[19] S. K. Pandey and R. R. Janghel, “Automatic detection of arrhythmia from imbalanced ECG database using CNN model with SMOTE,” *Australas. Phys. Eng. Sci. Med.*, vol. 42, no. 4, pp. 1129–1139, 2019, doi: 10.1007/s13246-019-00815-9.

[20] I. Brown and C. Mues, “An experimental comparison of classification algorithms for imbalanced credit scoring data sets,” *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3446–3453, 2012, doi: 10.1016/j.eswa.2011.09.033.

[21] Z. Xu *et al.*, “Software defect prediction based on kernel PCA and weighted extreme learning machine,” *Inf. Softw. Technol.*, vol. 106, pp. 182–200, Feb. 2019, doi: 10.1016/j.infsof.2018.10.004.

[22] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, “A study of the behavior of several methods for balancing machine learning training data,” *ACM SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, 2004, doi: 10.1145/1007730.1007735.

[23] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[24] Q. Fan, Z. Wang, D. Li, D. Gao, and H. Zha, “Entropy-based Fuzzy Support Vector Machine for Imbalanced Datasets,” *Knowl.-Based Syst.*, p. 32, 2016.

[25] J. Demsar, “Statistical Comparisons of Classifiers over Multiple Data Sets,” *J. Mach. Learn. Res.*, vol. 7, p. 30, 2006.

[26] J. Xiao, Y. Wang, J. Chen, L. Xie, and J. Huang, “Impact of resampling methods and classification models on the imbalanced credit scoring problems,” *Inf. Sci.*, vol. 569, pp. 508–526, 2021, doi: 10.1016/j.ins.2021.05.029.

[27] T. Hassner, Y. Itcher, and O. Kliper-Gross, “Violent flows: Real-time detection of violent crowd behavior,” in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Providence, RI, USA, Jun. 2012, pp. 1–6. doi: 10.1109/CVPRW.2012.6239348.

[28] J. Mahmoodi and A. Salajeghe, “A classification method based on optical flow for violence detection,” *Expert Syst. Appl.*, vol. 127, pp. 121–127, 2019, doi: 10.1016/j.eswa.2019.02.032.

[29] M. Ghazal, N. Waisi, and N. Abdullah, “The detection of handguns from live-video in real-time based on deep learning,” *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 18, no. 6, p. 3026, 2020, doi: 10.12928/telkomnika.v18i6.16174.