ORIGINAL ARTICLE

Automatic Detection of Laparoscopic Videos Distortion Using Machine Learning Classification

Mohamed Belmokeddem [1*], Kamila Khemis [1], Salim Loudjedi [2]

[1] *Biomedical Engineering Department, Faculty of Technology, University Abou-Bekr Belkaid of Tlemcen, Algeria*

[2] *Surgery B Tlemcen Hospital, Department of Medicine, University Abou-Bekr Belkaid of Tlemcen, Algeria*

*Corresponding Author: Mohamed Belmokeddem
Email: mohamed.belmokeddem@univ-tlemcen.dz

Received: 07 November 2023 / Accepted: 04 March 2024

Abstract

Purpose: Ensuring excellent video quality is crucial for the success of minimally invasive surgical procedures without disrupting the surgical procedure flow. Real-time laparoscopic video frequently encounters issues such as blur and smoke, often stemming from lens contamination. The automatic detection of these distortions is imperative to assist surgeons, ultimately reducing operative time and mitigating risks for the patient.

Materials and Methods: In this paper, we leverage the Laparoscopic Video Quality (LVQ) database developed by Khan *et al*. to train and validate our model. To classify defocus blur, motion blur, and smoke in the laparoscopic video, we adopt a novel approach utilizing a cascade support vector machine (SVM) classifier, which combines decisions from three binary classifiers. The first classifier categorizes videos into two classes: good and distorted. The second classifier focuses on detecting smoke and blur, while the third is dedicated to distinguishing between defocus blur and motion blur.

Results: In this study, we calculate performance metrics, including accuracy rate, precision, recall, F1 score, and execution time, which are crucial indicators for evaluating quality detection results. The machine-learning classification demonstrates notable performance, with an accuracy rate of 96.55% for the first classifier, 100% for the second, and 99.67% for the third classifier. Additionally, the classification achieves a high inference speed of 37 frames per second (fps).

Conclusion: The experimental results showcased in this paper underscore the efficacy of the proposed approach in automatically detecting distortions in a laparoscopic video. The method exhibits high performance, excelling in both accuracy and processing speed. Notably, the method's advantage lies in its simplicity and the fact that it does not necessitate high-performance computer hardware.

Keywords: Automatic Detection of Laparoscopic Video Distortion; Smoke and Blur Detection; Machine-Learning Classification.

DOI:

1. Introduction

Outstanding video quality is essential for minimally invasive surgical procedures, ensuring the flow of operative procedures [1]. Laparoscopic videos are susceptible to various distortions, including noise, smoke, uneven illumination, defocus blur, and motion blur. These are considered the five categories of laparoscopic video distortion [1]. Reduced video quality not only diminishes a surgeon's visibility but also compromises the outcomes of computational tasks in robot-assisted surgery and image-guided navigation systems [2, 3]. Examples of such tasks are segmentation [4-7], instrument tracking [8-11], and augmented reality [12, 13]. A noticeable drop in accuracy was observed when the laparoscopic image quality was degraded (blur, noise, or resolution drop) [14].

Numerous solutions have been proposed to address these issues, each grounded in the detection of one or more causes of distortion. The first methods focused on the detection and classification of smoke, using Support Vector Machines (SVM) [15] and Convolutional Neural Networks (CNN) [16-19].

Researchers have also proposed alternative approaches, including using the histogram of the saturation channel (S) in the HSV color space. These methods reveal that the color saturation decreases when smoke obscures the Field Of View (FOV) [20].

To classify the other distortions, the authors constructed a Laparoscopic Video Quality database (LVQ) containing a dataset of 200 videos. The database comprises five categories of distortions and four levels of intensity [1]. A specific method was used to classify each type of distortion: a fast noise variance estimator with a threshold for noise distortion [21], a saturation analysis (SAN) classifier for smoke distortion [16], statistics of the luminance component of an image for uneven illumination distortion [1], and a Perceptual Blur Index (PBI) with a threshold classifier for motion and defocus blur distortions [22]. The authors have also utilized various methods with different datasets to detect image sharpness or blurriness. Some of these methods include:

Haar wavelet transform (HWT), Fast Fourier Transform (FFT), Laplacian operator, and Sobel operator [23, 24]. However, these techniques are constrained to specific types of distortion. Furthermore, they are only applicable for single distortion classification [25]. When confronted with multiple distortions during processing, the complexity can significantly increase, leading to intensified masking effects between video frames [26].

The deep learning approach is gaining popularity and is successfully employed not only in image classification but also in text classification. Various architectures, such as Recurrent Neural Networks (RNN) or bidirectional RNN [27], contribute to its versatility.

In the analysis of laparoscopic images, a diverse set of deep learning methods was applied to detect and classify the five categories of laparoscopic video distortions. Some of these methods utilized single-label classification and a single Deep Neural Network (DNN) [28]. Others, based on multi-label distortion classification [29], used the convolutional neural network ResNet50 as DNN to extract the features from laparoscopic frames. Recently, a multi-label classification method was proposed, based on a combination of ResNet and a Fully Connected Neural Network (FCNN) [30], using the LVQ database introduced in [1]. However, most of these solutions are time-consuming during the training phase [27] and rely on the high performance of computer hardware. Additionally, there are challenges in the medical domain, including limited data access due to privacy concerns and the need for high-quality labeled datasets for training deep learning models [28]. The lack of availability of large-scale data is a common problem in medical applications [29].

To benefit from the distortion classification, it is necessary to separate distortions caused by laparoscopic lens contamination from distortions caused by technical problems. During minimally invasive surgical procedures, most distortions surgeons encountered were caused by contamination such as blur and smoke [31]. Therefore, our work focuses on detecting these types of distortions.

In this paper, we propose an alternative approach using three SVM classifiers in cascade for automatic detection and classification of defocus blur, motion blur, and smoke in laparoscopic video. In the first classifier, we assess the video quality as high or low. In the second, we identify the presence of blur or

smoke. Finally, the third classifier distinguishes between defocus blur and motion blur. The first and second classifiers use the variance and maximum values of the Laplacian and Sobel operators and the SAN classifier as features of the image. The third classifier uses the variance and the maximum value of the Laplacian and the Sobel operators as features of the image.

We propose a three-staged computational framework for the automatic detection of these distortions. The real-time detection of defocus blur in laparoscopic video can be used to activate an automatic in-situ lens cleaning system.

This paper introduces several contributions:

- A novel distortions classification system used to identify distortions, specifically those caused by laparoscopic lens contamination.
- A real-time solution to identify distortions in live-captured videos.
- A simple machine-learning, classification-based method that does not demand high-performance resources. Hence, it avoids the challenge of the limited availability of large-scale medical data.
- To our knowledge, this study is the first to focus on detecting distortions caused by laparoscopic lens contamination using a cascading SVM.

The structure of this paper is as follows: Section II, Materials and Methods, provides an overview of the datasets and a detailed description of the method employed. Section III, Results, delves into the experimental setup and results. Section IV, Discussion, provides an analysis and discussion of the results. Lastly, Section V, Conclusion, summarizes the outcomes and significance of this work, offering insights into potential future improvements.

2. Materials and Methods

2.1. Dataset Overview

In this paper, we utilize the LVQ database developed by Khan *et al.* [1]. The database consists of 10 reference videos, each lasting ten seconds. These videos are extracted from the Cholec80 dataset [32], which includes ten distinct categories of scene variations: Bleeding (BL), Grasping and Burning (GB), Multiple Instruments (MI), Irrigation (IR), Clipping (CL), Stretching Away (SA), cutting (CU), stretching forward (SF), Organ Extraction (OE), and Burning (BU).

Each reference video was distorted by five different categories of distortions with four levels of severity, as shown in Table 1 and Table 2. This results in a total of 200 videos annotated with subjective scores performed by expert and non-expert observers. Khan *et al.* [1] used these severity ranks to investigate how experts and non-experts perceive the themes of each distortion. The distortions include smoke, noise, uneven illumination, defocus blur, and motion blur, as shown in Figure 1.

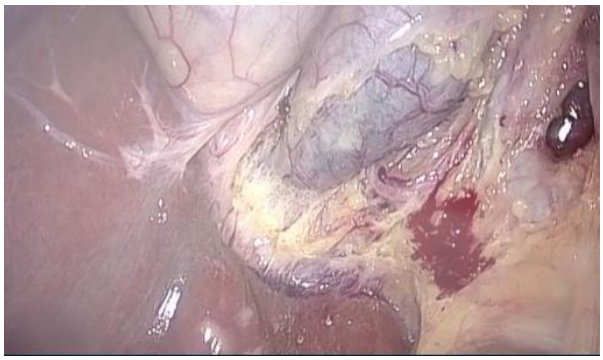
To validate our method, we utilized an extended version of the LVQ database (ICIP LVQ Challenge dataset) which was proposed in the International Conference on Image Processing Challenge (ICIP2020) [32]. The ICIP_LVQ_test has 200 laparoscopic videos, which include both single and multiple distortions. Specifically, there are 60 videos with noise, 50 videos with defocus blur, 45 videos with motion blur, 95 videos with smoke, and 95 videos with uneven illumination.

Table 1. Dataset description [1]

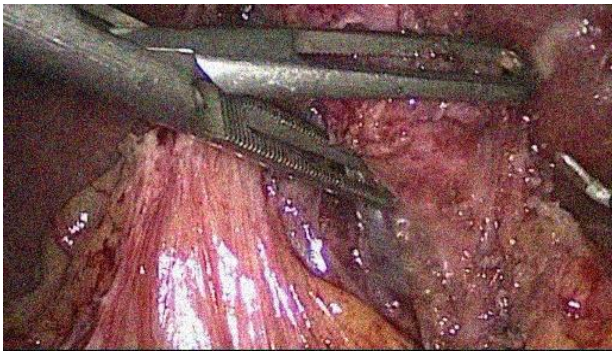
Resolution of each video	Frame rate	Number of distortions	Number of levels
512 X 288	25 fps	5	4

Table 2. Dataset summary [1]

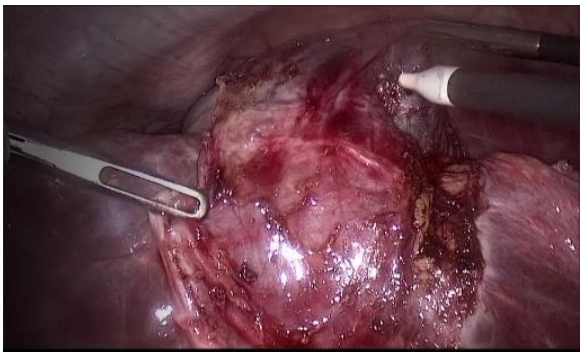
Distortions categories	Number of distorted videos	Number of videos by level	Number of levels of severity
Smoke	40	10	4
Noise	40	10	4
Uneven illumination	40	10	4
Defocus blur	40	10	4
Motion blur	40	10	4
Total	200	50	/



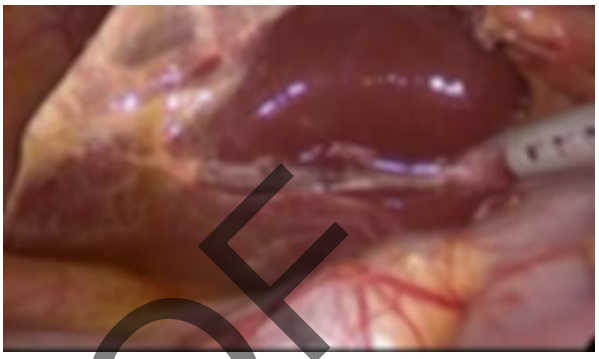
a: Smoke



b: Noise



c: Uneven illumination



d: Defocus blur

e: Motion blur

Figure 1. Various distorted images extracted from the LVQ database [1]

2.2. The Proposed Method

In this study, we propose a new approach using three SVM classifiers in cascade for the automatic detection and classification of defocus blur, motion blur, and smoke in laparoscopic video. The first classifier assesses the video quality as either high or low, using two classes: 'Degraded Image' and 'Good Image.' The second classifier identifies the presence of blur or smoke. Lastly, the third classifier differentiates between defocus blur and motion blur.

To generate degraded frames from the dataset, we exclude frames that exhibit the first level of defocus blur, motion blur, and smoke distortions. This exclusion is determined by the good scores obtained in expert evaluation, which fall between 0.8 and 1, as mentioned in [1]. Hence, the second, third, and fourth levels of these distorted videos were taken to create a training dataset encompassing three classes: 'Defocus Blurred Image,' 'Motion Blurred Image,' and 'Smoked Image.'

Consequently, 90 videos are chosen and converted into frames with a rate of 5 fps, as depicted in Table 3. These three classes are combined into a single class labeled 'Degraded Image.' It is important to note that the classes 'Noise' and 'Uneven Illumination' are

excluded from consideration, as this work focuses on detecting blur and smoke.

From the reference video in the LVQ database, all frames from 10 videos were extracted, with a frame rate of 25 fps. This class is labeled ‘Good Image’.

Among the various imbalanced learning approaches, the data-level approach was selected [33] to address the issue of imbalanced class distribution in the dataset. As a result, we resampled the training dataset by oversampling the 'Good Image' class, which is the minority class, achieved by vertically flipping the extracted frames, as illustrated in Table 3.

Table 3. Training data summary for the first classifier

Dataset	Number of frames
Number of defocus blurred frames	1548
Number of motion-blurred frames	1548
Number of smoked frames	1548
Number total of degraded frames	4644
Number of reference frames	2561
Number of reference frames after data augmentation	4644
Number of total frames	9288

For the second classifier, the two classes, 'Defocus Blurred Image' and 'Motion-Blurred Image,' are combined into a single class labeled 'Blurred Image.' To achieve balance with the 'Smoked Image' class, data augmentation is performed by vertically flipping the extracted frames, as shown in Table 4.

Table 4. Training data summary for the second classifier

Dataset	Number of frames
Number of blurred frames (defocus and motion)	3096
Number of smoked frames	1548
Number of smoked frames after data augmentation	3096
Number of total frames	6192

For the third classifier, we utilize two classes: 'Defocus Blurred Image' and 'Motion Blurred Image' as shown in Table 5.

Video frames are resized to 224x224 pixels before being applied to the SVM classifiers.

Table 5. Training data summary for the third classifier

Dataset	Number of frames
Number of defocus blurred frames	1548
Number of motion blurred frames	1548
Number of total frames	3096

2.3. Features Extraction

Blur features: To identify blur in laparoscopic videos, we utilize two pixel-based edge detection algorithms: the Laplacian, a second derivative operator, and the Sobel operator, a first derivative operator. This selection is made due to the observation that non-blurred images exhibit sharper edges in comparison to blurred images [24].

The Laplacian accentuates regions with rapid changes in intensity values. By computing the variance of the Laplacian of an image, we can discern between a blurred image and a sharp image. High variance indicates a widespread response, signifying an edge or sharp image, while low variance indicates a minimal spread of response, indicating blur [34]. Additionally, we employ the Sobel operator for edge detection due to its insensitivity to noise and its use of a relatively small mask in images [35].

In this study, we utilize the variance and the maximum value derived from each operator as image features for the SVM classifier, facilitating the discrimination between blurred and non-blurred frames [24, 35]. For instance, for blurred images, the variance and the maximum values of the Laplacian and Sobel operators are below a threshold automatically determined by the SVM classifier.

-Smoke features: In smoke detection, we utilize the SAN classifier, which relies on the histogram of the saturation channel from the HSV color space. Smoke-containing images typically exhibit reduced color, and as a result, the saturation channel in the HSV color space demonstrates a notable correlation with the

presence of smoke [16]. Leibetseder *et al.* employed saturation peak analysis (SPA) to assess the saturation of a frame by converting it into the HSV color space. This method isolates the corresponding S-channel and generates an intensity histogram. It is observed that images lacking color or containing smoke exhibit numerous low-saturation pixels. Consequently, their histograms display higher values in the lower bins, creating an inverse scenario for the upper bins [36]. If the majority of bin values in the histogram fall below the suggested threshold of tc = 0.35, as proposed in [16], the frame is classified as having smoke. The probability of an image having smoke is defined as (Equation 1):

$$Preds(H) = \frac{1}{|H|} \sum_{\substack{i=0 \\ b \in H \\ i \leq tc}} b_i \qquad (1)$$

Where: bi is the i-th bin value of histogram H, with |H|=256, and t_c the threshold (t_c = 0.35) [16].

Detecting smoke using saturation histograms essentially involves identifying a suitable concentration point for bin values in non-smoke samples called the classification threshold, or t_c. This threshold serves as a reference point for smoke samples, as they typically display a lower concentration [16].

2.4. The Proposed Work

The rationale behind the selection of the SVMs classifiers is that they are supervised learning classifiers of linear and non-linear data; they have different kernel functions (e.g., linear kernel, RBF kernel, polynomial kernel) [37]. They have relatively high predictive performance and are less affected by the overfitting problem [38]. The inputs into the SVM are feature vectors extracted from frames of the dataset cited in the preceding section. The goal is to separate the feature vectors to maximize the margins from both vectors.

Dimensionality reduction techniques can significantly decrease the time complexity during the training phase of machine learning algorithms, thus alleviating the computational burden. Two widely used dimensionality reduction methods are Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) [39]. However, we did not employ any of these methods in this work. Instead, our focus was solely on identifying the optimal SVM kernel. Our objective was to apply a transformation that yields the best separation of feature vectors, consequently achieving the highest accuracy. In this work, various SVM kernels were evaluated, and the Gaussian Radial Basis Function (RBF) kernel was found to provide the highest accuracy. The optimal parameters were determined as C = 100 and Gamma = 0.01.

SVM is trained to fit the extracted features and map them to specific classes. Three binary SVMs are utilized in cascade in this paper, as shown in Figure 2.

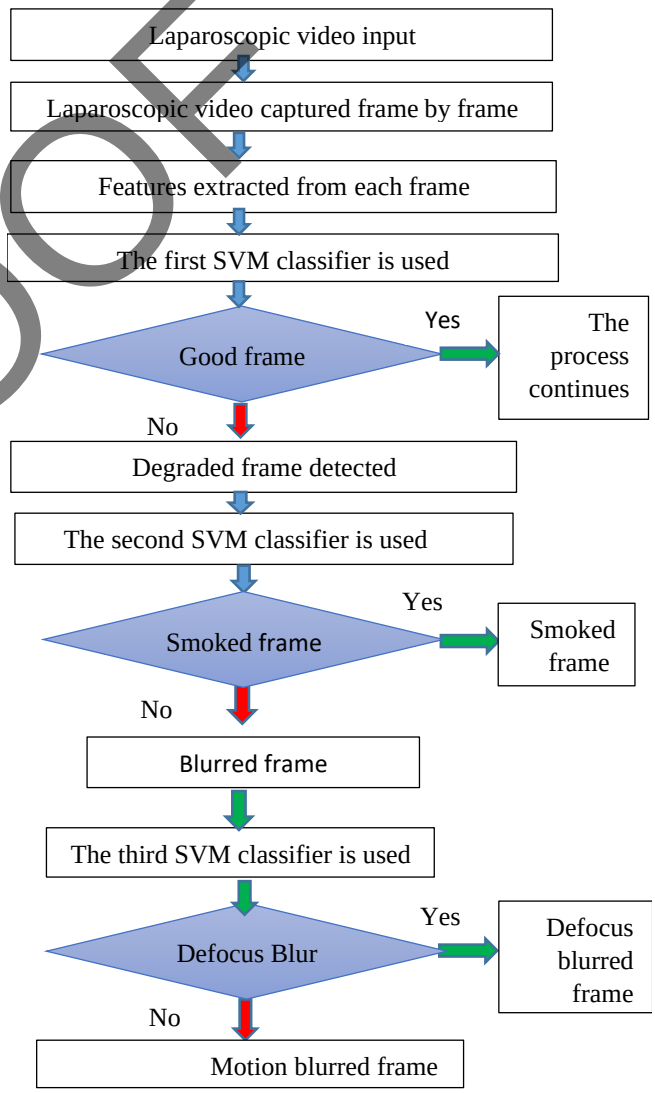


Figure 2. Flowchart of the proposed framework for laparoscopic video quality identification

2.4.1. First Step: Degraded Image/Good Image Classification

For the first classifier, training data is generated from the two classes 'Degraded Image' and 'Good Image' and stored in a data file. Subsequently, various feature combinations are explored, including:

- The variance and maximum value of the Laplacian operator and SAN classifier.

- The variance and maximum value of the Sobel operator and SAN classifier.

- The variance and maximum value of the Laplacian and Sobel operators and the SAN classifier.

Finally, Scikit-learn confusion matrices are presented, prediction parameters (accuracy, prediction, precision, recall, f1-score) are computed, the best feature combination is determined, and the prediction model is saved.

2.4.2. Second Step: Blurred Image/ Smoked Image Classification

In the second step, frames from the two classes, 'Blurred Image' and 'Smoked Image' are used to train the second classifier. The image features used include the maximum value and variance of the Laplacian and Sobel operators and the SAN classifier. Afterward, the second prediction model is saved. This classifier is used when a degraded frame is detected in the first step.

2.4.3. Third Step: Defocus Blurred Image/ Motion Blurred Image Classification

In the third step, frames from the two classes, 'Defocus Blurred Image' and 'Motion Blurred Image', are used to train the third classifier. The features used for the image include the maximum value and variance of both the Laplacian and Sobel operators. Finally, the third prediction model is saved. This model is used when the output of the second classifier indicates a blurred image.

3. Results

3.1. Experimental Setup

The experiments are performed on a desktop computer with the following specifications: CPU - Intel Core i3 - 2365M @ 1.40 GHz, GPU - Intel HD Graphics 3000, and 4 GB of RAM. The computer runs on Windows 7 64-bit operating system. Various frameworks, including OpenCV (CV2) v4.5.5, Scikit-Learn 1.1.2, and Scikit-Image 0.19.2, are utilized with Python 3.8.

3.2. Performance Metrics

To train and evaluate the proposed methodology, we randomly divided each dataset into two non-overlapping subsets: 80% for training and 20% for testing. The distribution for each classifier is as follows:

- First Classifier: 7430 frames for training and 1858 frames for testing.
- Second Classifier: 4953 frames for training and 1239 frames for testing.
- Third Classifier: 2476 frames for training and 620 frames for testing.

Ultimately, our method is evaluated using 200 videos from the testing dataset [31], and the results are compared to the ground truth of this dataset.

The performance metrics, encompassing accuracy rate, precision, recall, F1 score, and execution time, are computed. These metrics serve as crucial indicators for evaluating the outcomes of quality detection. The summary of performance metrics is as follows:

Accuracy calculates the percentage of correctly predicted frames out of all the test frames (Equation 2):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Precision refers to the accuracy of positive predictions [29] (Equation 3):

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

Recall (sensitivity) measures the fraction of positive predictions that were correctly identified [29] (Equation 4):

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

- The F1 score encapsulates recall and precision in a single quantity. The best score is 1.0 [29] (Equation 5).

$$\text{F1\_score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision}+\text{recall}} \quad (5)$$

- CPU execution time represents the time taken by the CPU to execute the framework and classify the frames.

- The frame rate, measured in frames per second (fps), refers to the speed at which video frames are scanned to identify distortions in live-captured laparoscopic videos [29].

3.3. Experimental Results

This section presents the evaluation results of the proposed method for each classifier. The focus is on the accuracy rate, precision, recall, F1 score, and confusion matrix.

3.3.1. The First Classifier

For the first classifier, we conducted a comparative analysis of different blur and smoke detection methods in images. Various feature combinations as described in the methodology section are compared. The confusion matrix for each method is shown in Figure 3.

Figure 3 displays the confusion matrices. In each matrix, the rows represent the true (actual) classes: 'Degraded Image' (class 0) and 'Good Image' (class 1).

The columns represent the predicted classes for each feature combination (a, b, and c), and each element (i, j) in the matrix corresponds to the number of frames known to be in real class i and predicted to be in class j

Table 6 shows accuracy, recall, precision, F1 score, and CPU execution time.

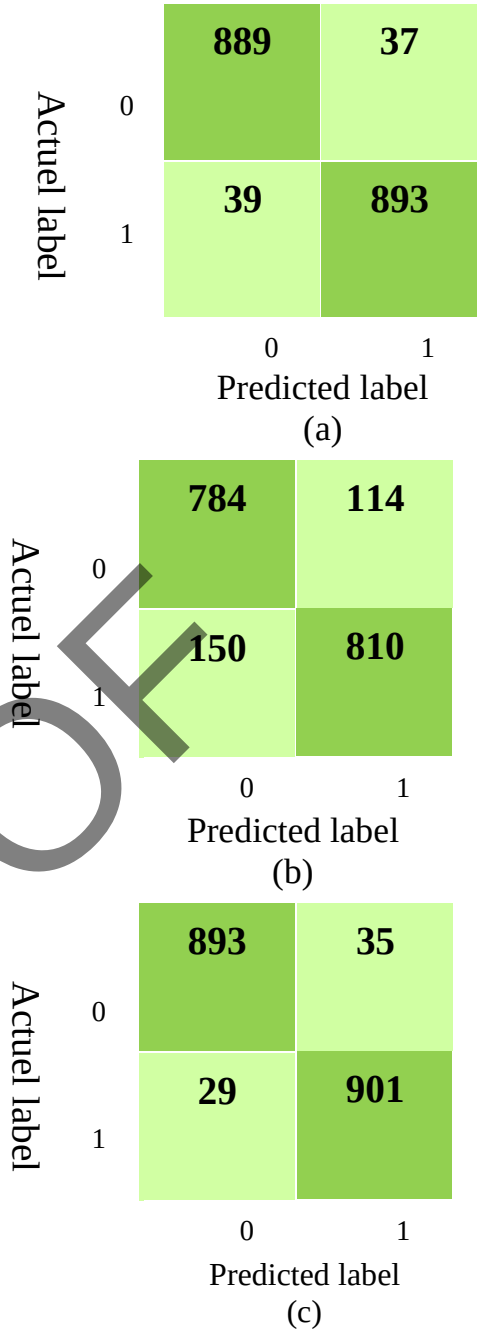


Figure 3. The confusion matrix of the proposed solution for the first classifier with each feature configuration: (a) The variance and the maximum value of the Laplacian operator and SAN classifier; (b) The variance and the maximum value of the Sobel operator and SAN classifier; (c) The variance and the maximum value of the Laplacian and Sobel operators and SAN classifier

Table 7 displays the accuracy, recall, precision, F1 score, and CPU execution time for the second and third classifiers.

Figure 4 shows the confusion matrices of the second classifier (a) and the third classifier (b). The two classes are 'Blurred Image' (class 0) and 'Smoked

Image' (class 1) for the second classifier. For the third classifier, the classes are 'Defocus Blurred Image' (class 0) and 'Motion Blurred Image' (class 1).

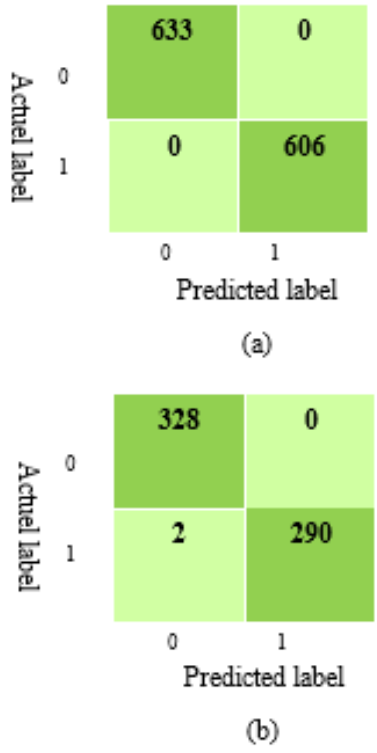


Figure 4. The confusion matrix of the proposed solution for the second and third classifiers: (a) The second classifier with the maximum value and the variance of the Laplacian, the Sobel operators, and the SAN classifier; (b) The third classifier with the maximum value and the variance of both the Laplacian and the Sobel operators

Table 7 displays the accuracy, recall, precision, F1 score, and CPU execution time for the second and third classifiers.

3.3.2. The Proposed Method Tested on the ICIP2020 Dataset

Our method was tested on the 200 videos in the ICIP2020 testing dataset [32].

We compared the distortion classification results of our method to the ground truth of the ICIP2020 testing dataset and illustrated the outcomes using confusion matrices. Figure 5 presents these confusion matrices categorized by distortion types: defocus blur (a), motion blur (b), and smoke (c).

Table 8 displays the performance parameters (accuracy, recall, precision, and F1 score) of our method tested on 200 distorted videos of the ICIP 2020 testing dataset. The results are categorized by distortion types (defocus blur, motion blur, and smoke).

3.3.3. Comparison of Methods

To showcase the efficiency of the used approach in classifying defocus blur, motion blur, and smoke

Table 6. The performance metrics of the first classifier for each feature configuration

Methods	Accuracy %	Recall %	Precision %	F1-Score %	CPU execution time (s)
The variance and the maximum value of the Laplacian operator and the SAN classifier	95.90	95.81	96.02	95.91	0.93
The variance and the maximum value of the Sobel operator and the SAN classifier	85.79	84.37	87.66	85.98	2.18
The variance and the maximum value of the Laplacian, the Sobel operators, and the SAN classifier	96.55	96.88	96.26	96.56	0.78

Table 7. The performance metrics of the second and third classifiers

Methods	Accuracy %	Recall %	Precision %	F1-Score %	CPU execution time (s)
The second classifier (the maximum value and the variance of the Laplacian and Sobel operators, and the SAN classifier)	100	100	100	100	0.24
The third classifier (the maximum value and the variance of both the Laplacian and Sobel operators)	99.67	99.31	100	99.65	0.14

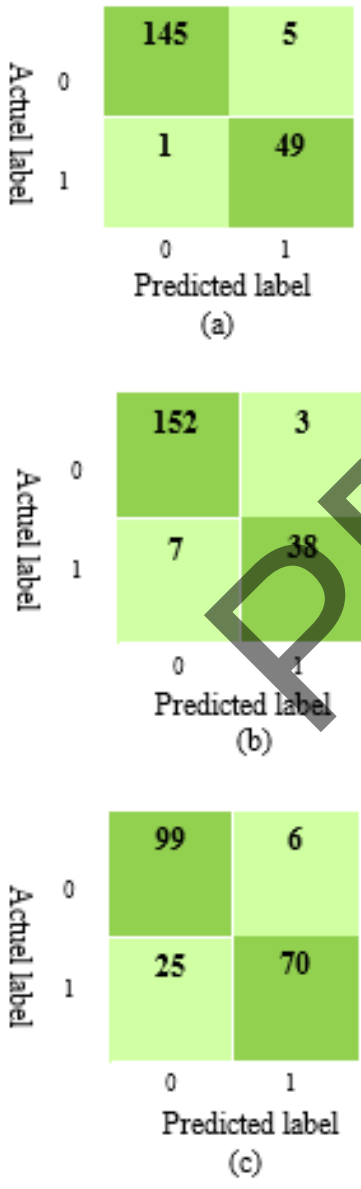


Figure 5. The confusion matrix of the proposed solution for each distortion tested on the ICIP testing dataset: (a) defocus blur, (b) motion blur, and (c) smoke

Table 8. Performance metrics of the proposed method tested on the ICIP testing dataset

Distortions	Accuracy %	Recall %	Precision %	F1-Score %
Defocus blur	97	98	90.74	94.23
Motion blur	95	84.44	92.68	88.36
Smoke	84.5	73.68	92.10	81.86

Table 9. Comparing the performance metrics of our method to those of Aldahoul *et al.*'s [27] and Khan *et al.*'s [1] methods for defocus blur

Distortions	Accuracy %	Recall %	Precision %	F1-Score %
Our method	97	98	90.74	94.23
Aldahoul *et al.* method	100	100	100	100
Khan *et al.* method	91	77.5	77.5	77.5

Table 10. Comparing the performance metrics of our method to those of Aldahoul *et al.*'s [27] and Khan et al.'s [1] methods for motion blur

Distortions	Accuracy %	Recall %	Precision %	F1-Score %
Our method	95	84.44	92.68	88.36
Aldahoul et al. method	90	55.55	100	71.42
Khan et al. method	89.5	75	73.17	74.07

distortions in single-label classification, we compare it to the methods proposed by Aldahoul *et al.* [27] and

Khan *et al.* [1]. Both of these methods utilize the ICIP 2020 dataset.

Tables 9, 10, and 11 present the performance parameters (accuracy, recall, precision, and F1 score) of our method compared to Aldahoul *et al.*'s [27] and Khan *et al.*'s [1] methods. The tables are organized by distortion types: defocus blur, motion blur, and smoke. We assessed 200 distorted videos from the ICIP 2020 testing dataset. The results of Khan *et al.*'s method were extracted from the Khan thesis [40].

Table 11. Comparing the performance metrics of our method to those of Aldahoul *et al.*'s [27] and Khan *et al.*'s [1] methods for smoke

Distortions	Accuracy %	Recall %	Precision %	F1-Score %
Our method	84.5	73.68	92.10	81.86
Aldahoul *et al.* method	95.5	90.53	100	95.03
Khan *et al.* method	86.2	95	58.46	72.37

The inference time of our solution took 0.027 seconds, calculated from the time step of applying the video frame at the input until the time step of producing the distortion types at the output. Consequently, the model demonstrated the ability to scan video frames at a speed of 37 Frames Per Second (FPS). In contrast, for Aldahoul *et al.*'s method, the speed was 20 fps. In their study, Khan *et al.* did not specify the frame rate for their method.

Figure 6 shows a graphical comparison of the three methods for each type of distortion: defocus blur (a), motion blur (b), and smoke (c). The comparison is based on accuracy, recall, precision, and F1 score.

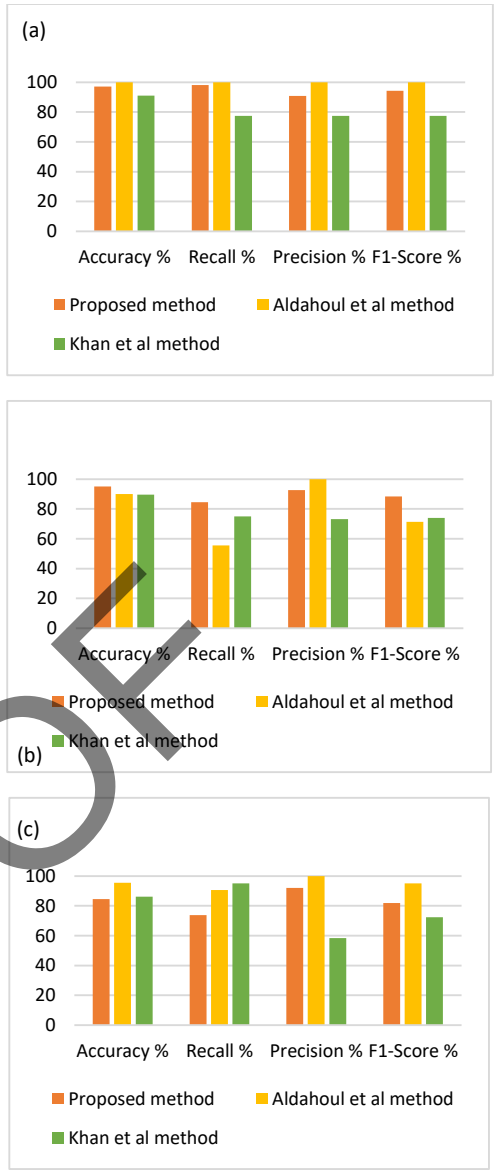


Figure 6. Graphical comparison between the proposed method, Aldahoul *et al.* method, and Khan et al. method [1], tested on the ICIP testing dataset, for each distortion: (a) defocus blur; (b) motion blur; (c) smoke

4. Discussion

The main objective of this study was to develop a method for real-time detection and classification of defocus blur and smoke in laparoscopic video, which can be caused by lens contamination. These distortions can impair surgical visibility. For these tasks, processing speed is crucial to supporting surgeons. To achieve this goal, we introduce a novel cascaded Support Vector Machine (SVM) classifier framework comprised of three sequentially interconnected SVMs. This innovative approach effectively addresses the challenges of real-time blur detection and classification in laparoscopic video. The first classifier is to detect if the video is good or distorted; the second is to find the type of distortion: blur or smoke; and the third is to distinguish between defocus blur and motion blur. We train our method on the LVQ database developed by Khan *et al.* [1].

In Section 3.3.1, we observe that the first classifier achieves optimal performance when utilizing the

variance and maximum values of the Laplacian and Sobel operators, along with the SAN classifier, as features for the image. This configuration yields an accuracy of 96.55%, an F1-Score of 96.56%, and a CPU execution time of 0.78 seconds, as depicted in Figure 3 and summarized in Table 6. In the second position, we find that employing the variance and maximum value of the Laplacian operator along with the SAN classifier results in an accuracy of 95.90%, an F1-Score of 95.91%, and a CPU execution time of 0.93 seconds. These findings underscore the impact of feature extraction methods on the evaluation metric values and, consequently, on the classification results.

In Section 3.3.2, we find that the second classifier has an accuracy of 100%, an F1-Score of 100%, and a CPU execution time of 0.24 seconds, demonstrating that it is highly capable of identifying blurred and smoked frames, as shown in Figure 4 and Table 7. The third classifier has an accuracy of 99.67%, an F1-Score of 99.65%, and a CPU execution time of 0.14 seconds, demonstrating that it is highly capable of distinguishing between defocus-blurred and motion-blurred frames.

In Section 3.3.3, our evaluation of the method on the testing dataset's 200 videos indicates its strong ability to detect defocus-blurred videos, among other distortions. It achieves an accuracy of 97% and an F1-Score of 94.23%, as depicted in Figure 5 and Table 8. Regarding motion-blurred videos, we can see that this method can detect this distortion with a good accuracy of 95% and an F1-Score of 88.36%. The few misclassified videos have a distortion combination between motion blur and uneven illumination or motion blur and smoke. The proposed method classifies smoke with a reasonable accuracy of 84.5% and an F1 score of 81.86%. The high number of videos with smoke that were not detected by the framework (25 false negatives) contributed to the low performance. Notably, 88% of these are distorted by a combination of smoke and defocus blur, highlighting a limitation of our method in accurately detecting smoke in the presence of defocus blur.

In Section 3.3.4, we find that the metrics performance analysis of the compared methods reveals that, in the classification of defocus blur (as shown in Table 9), our method achieves an accuracy of 97%. This accuracy surpasses that of Khan *et al.* (91%) but falls short of that found by Aldahoul *et al.* (100%). Additionally, the F1 score is 94.23%, compared with 100% of the Aldahoul *et al.* method and 77.5% of the Khan *et al.* method, as depicted in Table 10, our method achieves an accuracy of 95%, surpassing both the Aldahoul *et al.* method (90% accuracy) and the Khan *et al.* method (89.5% accuracy). Additionally, the F1 score of our method is 88.36%, which is the highest compared to 74.07% of the Khan *et al.* method and 71.42% of the Aldahoul *et al.* method.

These results highlight the robust capabilities of our model, not only in accurately detecting blur but also in effectively distinguishing between defocus blur and motion blur.

In the classification of smoke, as displayed in Table 11, the Aldahoul *et al.* method achieves the highest accuracy (95.5%), outperforming the Khan *et al.* method (86.2%) and our method (84.5%). Additionally, the F1 score was 95.03%, which is the highest, compared to 81.86% of our method and 72.37% of the Khan *et al.* method, confirming the remark in the previous paragraph.

With a frame rate of 37 fps, our method surpasses the Aldahoul *et al.* approach, which operates at 20 fps in terms of processing time.

This result underscores the feasibility of implementing our algorithm in real-time systems, including low-cost computers, facilitating the identification of distortions in live-captured laparoscopic videos. Consequently, the real-time solution can effectively identify distortions in live-captured videos with a speed of 37 FPS.

5. Conclusion

During laparoscopy, the visual field can be obstructed by condensation, debris, blood, and smoke, which generate defocus blur and smoke distortions in the output real-time laparoscopic video. Processing speed is crucial to supporting surgeons in the detection and elimination of lens contamination.

The advantage of our method is that it is simple and does not require high-performance computer hardware.

In this work, we used machine-learning classification to automatically detect distortions in

laparoscopic video. The results show a high level of performance in terms of accuracy and processing speed.

One drawback of the current approach is its dependence on manually crafted features. Looking ahead, there is potential for leveraging deep learning methods to address this limitation. Deep learning algorithms present several advantages over traditional machine learning methods, including automatic feature learning, robust handling of extensive and complex datasets, improved performance, adept management of non-linear relationships, versatility in processing structured and unstructured data, and skillful predictive modeling. Moreover, platforms such as Google Colab and Kaggle simplify the implementation of these advanced algorithms.

Acknowledgments

There are no conflicts to declare.

No funding was received to conduct this research and publication of this article.

All authors have seen and approved the manuscript.

References

1- Z. A. Khan *et al.*, "Towards a Video Quality Assessment based Framework for Enhancement of Laparoscopic Videos." 2020/3// (2020).

2- P. Sánchez-González *et al.*, "Laparoscopic video analysis for training and image-guided surgery." in Minimally Invasive Therapy and Allied Technologies Vol. 20, ed, (2011), pp. 311-20.

3- J. Cartucho, S. Tukra, Y. Li, D. S. Elson, and S. Giannarou, "VisionBlender: a tool to efficiently generate computer vision datasets for robotic surgery." *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, Vol. 9 (No. 4), pp. 331-38, (2021).

4- S. Voros, J.A. Long, and P. Cinquin, "Automatic detection of instruments in laparoscopic images: A first step towards high-level command of robotic endoscopic holders." in *International Journal of Robotics Research*, (2007), Vol. 26, pp. 1173-90.

5- D. Owen, M. Grammatikopoulou, I. Luengo, and D. Stoyanov, "Automated identification of critical structures in laparoscopic cholecystectomy." *International Journal of Computer Assisted Radiology and Surgery*, Vol. 17 (No. 12), pp. 2173-81, 2022/12// (2022).

6- E. Padovan *et al.*, "A deep learning framework for real-time 3D model registration in robot-assisted laparoscopic surgery." *International Journal of Medical Robotics and Computer Assisted Surgery*, Vol. 18 (No. 3), 2022/6// (2022).

7- M. Carstens *et al.*, "The Dresden Surgical Anatomy Dataset for Abdominal Organ Segmentation in Surgical Data Science." *Scientific Data*, Vol. 10 (No. 1), pp. 3-3, 2023/1// (2023).

8- B.F. Allen, F. Kasper, G. Nataneli, E. Dutson, and P. Faloutsos, "Visual tracking of laparoscopic instruments in standard training environments." in *Studies in Health Technology and Informatics*, (2011), Vol. 163: *IOS Press*, pp. 11-17.

9- D. Bouget, M. Allan, D. Stoyanov, and P. Jannin, "Vision-based and marker-less surgical tool detection and tracking: a review of the literature." in Medical Image Analysis Vol. 35, ed: Elsevier B.V., (2017), pp. 633-54.

10- M. Ali, G. Ochoa-Ruiz, and S. Ali, "A semi-supervised Teacher-Student framework for surgical tool detection and localization." 2022/8// (2022).

11- J. Cartucho, C. Wang, B. Huang, D. S. Elson, A. Darzi, and S. Giannarou, "An enhanced marker pattern that achieves improved accuracy in surgical tool tracking." *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, Vol. 10 (No. 4), pp. 400-08, (2022).

12- S. Bernhardt, S.A. Nicolau, L. Soler, and C. Doignon, "The status of augmented reality in laparoscopic surgery as of 2016." in Medical Image Analysis Vol. 37, ed: Elsevier B.V., (2017), pp. 66-90.

13- F. Joeres, F. Heinrich, D. Schott, and C. Hansen, "Towards natural 3D interaction for laparoscopic augmented reality registration." *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, Vol. 9 (No. 4), pp. 384-91, (2021).

14- J.R. Abbing, F.J. Voskens, B.G. A. Gerats, R.M. Egging, F. Milletari, and I.A. M. J. Broeders, "Towards an AI-based assessment model of surgical difficulty during early phase laparoscopic cholecystectomy." *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, (2023).

15- T.A. Alshirbaji, N.A. Jalal, L. Mündermann, and K. Möller, "Classifying smoke in laparoscopic videos using SVM." *Current Directions in Biomedical Engineering*, Vol. 3 (No. 2), pp. 191-94, 2017/9// (2017).

16- A. Leibetseder, M.J. Primus, S. Petscharnig, and K. Schoeffmann, "Real-time image-based smoke detection in endoscopic videos." in *Thematic Workshops 2017 - Proceedings of the Thematic Workshops of ACM Multimedia 2017, co-located with MM 2017*, (2017): *Association for Computing Machinery, Inc*, pp. 296-304.

17- A. Leibetseder, M.J. Primus, and K. Schoeffmann, "Automatic smoke classification in endoscopic video." in *Lecture Notes in Computer Science (including subseries*

Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), (2018), Vol. 10705 LNCS: *Springer Verlag*, pp. 362-66.

18- W. Reiter, "Improving endoscopic smoke detection with semi-supervised noisy student models." *Current Directions in Biomedical Engineering*, Vol. 6 (No. 1), 2020/5// (2020).

19- S. Salazar-Colores, H. M. Jiménez, C.J. Ortiz-Echeverri, and G. Flores, "Desmoking Laparoscopy Surgery Images Using an Image-to-Image Translation Guided by an Embedded Dark Channel." *IEEE Access*, Vol. 8pp. 208898-909, (2020).

20- N.A. Jalal, T.A. Alshirbaji, L. Mündermann, and K. Möller, "Features for detecting smoke in laparoscopic videos." *Current Directions in Biomedical Engineering*, Vol. 3 (No. 2), pp. 521-24, 2017/9// (2017).

21- J. Immerkaer, "NOTE: Fast Noise Variance Estimation." in "COMPUTER VISION AND IMAGE UNDERSTANDING," (1996), Vol. 64.

22- A. Chetouani, A. Beghdadi, and M. Deriche, "A new reference-free image quality index for blur estimation in the frequency domain." in *IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2009*, (2009), pp. 155-59.

23- R. Bansal, G. Raj, and T. Choudhury, "Blur image detection using Laplacian operator and Open-CV." in *Proceedings of the 5th International Conference on System Modeling and Advancement in Research Trends, SMART 2016*, (2017): *Institute of Electrical and Electronics Engineers Inc.*, pp. 63-67.

24- U. Ali and M.T. Mahmood, "Analysis of blur measure operators for single image blur segmentation." *Applied Sciences (Switzerland)*, Vol. 8 (No. 5), 2018/5// (2018).

25- Shun Zhang *et al.*, "Concerns of quality, utility, and reliability of laparoscopic gastrectomy for gastric cancer in public video sharing platform." *Annals of Translational Medicine*, Vol. 8 (No. 5), p. 196, (2020).

26- X. Liu, X. Tao, M. Xu, Y. Zhan, and J. Lu, "An EEG-Based Study on Perception of Video Distortion Under Various Content Motion Conditions." *IEEE Transactions on Multimedia*, Vol. 22 (No. 4), pp. 949-60, (2020).

27- A. Onan, "Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification." *Journal of King Saud University - Computer and Information Sciences*, Vol. 34 (No. 5), pp. 2098-117, 2022/5// (2022).

28- Z.A. Khan, A. Beghdadi, M. Kaaniche, and F.A. Cheikh, "RESIDUAL NETWORKS BASED DISTORTION CLASSIFICATION AND RANKING FOR LAPAROSCOPIC IMAGE QUALITY ASSESSMENT." in "2020 IEEE International Conference on Image Processing (ICIP)," (2020).

29- N. Aldahoul, H. A. Karim, M.J.T. Tan, and J.L. Fermin, "Transfer Learning and Decision Fusion for Real Time Distortion Classification in Laparoscopic Videos." *IEEE Access*, Vol. 9pp. 115006-18, (2021).

30- Z.A. Khan, A. Beghdadi, M. Kaaniche, F.A. Cheikh, and O. Gharbi, "A Neural Network based Framework for Effective Laparoscopic Video Quality Assessment." 2022/2// (2022).

31- M. Siddaiah-Subramanya, M. Nyandowe, and K.W. Tiang, "Technical problems during laparoscopy: A systematic method of troubleshooting for surgeons." *Innovative Surgical Sciences*, Vol. 2 (No. 4), pp. 233-37, 2017/12// (2017).

32- A. Beghdadi, "Real-Time Distortion Classification inLaparoscopic Videos ICIP." ed, (2020).

33- A. Onan, "Consensus Clustering-Based Undersampling Approach to Imbalanced Learning." *Scientific Programming*, Vol. 2019(2019).

34- G. Mahak and K. Jaspreet, "Automation in Blur Image Detection with Segmentation using Machine learning." *IJARCCE*, Vol. 8pp. 1-7, 03/30 (2019).

35- S. Lakshmi Bhavani, "Detection and Classification of Blur Images using Multi-Class Support Vector Machine; Detection and Classification of Blur Images using Multi-Class Support Vector Machine." [Online]. Available: www.ijert.org.

36- Andreas Leibetseder, Manfred Jürgen Primus, Stefan Petscharnig, and Klaus Schoeffmann, "Image-Based Smoke Detection in Laparoscopic Videos." in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, (2017), Vol. 10550 LNCS: *Springer Verlag*, pp. 70-87.

37- B. Schölkopf, "SVMs - A practical consequence of learning theory." *IEEE Intelligent Systems and Their Applications*, Vol. 13 (No. 4), pp. 18-21, 1998/7// (1998).

38- A. Onan, "An ensemble scheme based on language function analysis and feature engineering for text genre classification." *Journal of Information Science*, Vol. 44 (No. 1), pp. 28-47, 2018/2// (2018).

39- G. T. Reddy *et al.*, "Analysis of Dimensionality Reduction Techniques on Big Data." *IEEE Access*, Vol. 8pp. 54776-88, (2020).

40- Z.A. Khan, "Learning based quality assessment for medical imaging in the context of liver cancer treatment." PhD thesis, *Sorbonne Paris Nord University Paris XIII*, 2021PA131004, (2021).