Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Video Forgery Detection using an Improved BAT with Stacked Auto Encoder Model

Girish Nagaraj[1,*], Nandini Channegowda[2]

[1] DSATM Research Center, VTU Campus, Visvesvaraya Technological University, Machhe, Belagavi, Karnataka 590018, India
[2] Department of Computer Science and Engineering (AI), Dayananda Sagar Academy of Technology and Management, Bengaluru, India

ARTICLE INFO

Article history:
Received
Received in revised form
Accepted
Available online

Keywords:

Video forgery; Improved bat optimization; Spatial-temporal averaging method; Unsupervised feature selection; Video forgery detection

ABSTRACT

There are various public and private places such as banks, roads, offices, and homes equipped with cameras for surveillance. The surveillance videos are consisting of a precious source of information related to critical application scopes. The main problem is to aid powerful and accessible software that changes the content present in the video for the forgery creation of a video. The forgery involves region duplication that has a common video tampering. The existing techniques are utilized to detect video tampering from the forged videos that showed complexity in the background. Thus, it is important to overcome the problem of forgery detection in the research. The Spatio-temporal averaging model is carried out for the collection of a video sequence for obtaining the background information. This can detect the moving objects effectively for forgery detection. Next, the ResNet 18 is used for extraction of the feature vectors, and the discriminative feature vectors were reduced and improved the training time and accuracy. The Single Auto Encoder (SAE) is not able to reduce the input features' dimensionality. Thus, the SAE has used 3 encoders stacked on the top for detecting the forgery. It is based on the sequence of videos. In comparison to the existing models, the proposed approach outperformed them with accuracy rates of 98.6%, sensitivity rates of 98.60%, specificity rates of 98.47%, MCC rates of 97.29%, and precision rates of 99.93%.

1. Introduction

The multimedia editing tools show the fast development of the forgers which modifies the contents conveniently without multi-media professional skills [1]. Fake media is distinguished from multimedia forensics; the real version has been examined for several years. The video demonstration has proved a productive way to share thoughts and sentiments [2]. An extensive creation was priced reasonably that captured the portable video through devices that include cell phones and cameras. The cell phones are activated with rapid improvement based on visual data generation [3]. The various key fields like courtrooms and journalism worldwide are used as a means of communication for the videos. The model assures the validation of the content that was never provided. Software

* *Corresponding author.*
E-mail address: girish.pt.6@gmail.com

https://doi.org/10.37934/araset.42.2.175187

tools with high quality were used that alter the content of the videos [4]. The model was interrogated with genuineness and the multi-media tools were used for modifying the content conventionally without professional skills [5-6]. The forgery recognition approaches of two types were used for detecting passive and active approaches. The information is hidden with an image that has extracted active information from it [7]. The active approach can extract the information that has been hidden from an image. The secret information present in the watermarks was formed as a digital signature. The passive methods are used for detecting the duplicate region like splicing and copy-move forgery (CMF) in an image [8]. The tampered video datasets mainly focus on single tampering such as copy, object-based tampering, and move object splicing in the spatial domain. The video splicing is performed by replacing it with the video content portions from other video regions. The copy operations are moved and tampered with the areas for splicing that comes with another video [9]. Therefore, it is important to generate splices with distinct optical pipelines. In the splicing detection task, the research study considers the GRIP dataset. The image forgery scenario that was tampered with was not the same as that of the original region. It usually undergoes the process of post-processing operations like scaling, softening, rotation, blurring, smoothing, and denoising showing a better appearance visually [10]. The research [11] derived concise characteristics from digital videos that encompassed both temporal and spatial data of video segmentation. They utilized I3D and ResNet, achieving an accuracy of 84.05% on the VIRAT dataset and 85.88% on the MFC dataset. The research [12] introduced an approach for identifying inter-frame forgeries through Convolutional Neural Networks (CNN). This method involved the retraining of existing CNN models trained on the ImageNet dataset, enabling the detection of inter-frame forgeries while leveraging spatial-temporal relationships in a video. Their CNN model achieved an 81% accuracy rate. The research [13] introduced two network structures aimed at efficiently detecting forgeries with minimal computational expense through the utilization of deep learning techniques. Their methodology centred on detecting both Deepfake and Face2Face, two methods utilized to produce highly realistic manipulated videos. They attained a successful detection rate exceeding 98% for Deepfake and 95% for Face2Face. Thus, human beings are deceived with tampered images showed difficulty manually for image authentication. Thus, the computational overhead is reduced for the detection method and the feature selection technique reduced the features for improving the prediction rates. The devices are limited by their computational capabilities. The major contribution of this research is mentioned as follows,

Initially, pre-processing technique spatio temporal averaging is done on the collected GRIP dataset to eliminate the unwanted noise.

Then, the feature extraction is completed by using CNN and ResNet18 models.

Followed by that, the feature selection method selects the relevant features from the dataset using IBOA and applied for classification.

Finally, the classifier model based on selected features performs classification using SAE with DNN to improve the accuracy.

The structure of the research work is given as follows; Section 2 describes the existing methods involved in forgery detection. Section 3 illustrates the proposed method and Section 4 discusses the results. The conclusion and future work of this research work is given in Section 5.

2. Literature Review

The existing models that are involved in forgery detection using machine learning models are given as follows:

Harpreet Kaur & Jindal [14] developed a Deep CNN (DCNN) model for the detection of forgery in the video. The CNN-based model utilized the pre-processing layer for the detection of repositioned frames. Later, the overfitting was mitigated that added convolutional networks that were highly efficient and exposed the inter-frame tampering using DCNN. The developed algorithm detected the forgery without additional pre-embedding information in the frame. However, a various number of videos showed higher image resolution for a long length obtained lower metrics, and lesser efficiency.

Beijing Chen [15] developed a Fractional Quaternion Zernike Moments for performing a robust colour image for forgery detection. The implementation algorithm used is FrQZMs to speed up the computation for each component with the quaternion signal. The proposed FrQZMs evaluated the performances by considering the colour image copy for determining forgery. The modified PatchMatch algorithm is used for matching features using the algorithm and the FrQZMs are the features considered to match the algorithm. However, a deep learning model was needed to detect the copy-move forgery strategy. So, an effective network to improve the robustness of various kinds of additional operations was needed.

Xiao Jin [16] developed a dual stream network to embed the object-based video forgery detection based on depth information. The dual stream framework was used that jointly discovered and integrated effective features to detect the object based on video forgery. There are two distinct types of branches that were employed in examining the discriminative features. The dual stream features fused the Conditional Random Field (CRF) layer for enhancing the segmentation results. The temporal consistency was required to be incorporated into the video tracking strategy. The training stage cannot find the large-scale video strategy using neural networks. The temporal information was not introduced in training the forged videos.

Kang Hyeon Rhee [17] performed semantic segmentation for Copy-Move forgery detection to classify the image. The developed model generated a novel GT image for solving the problem to detect the Copy-Move forgery detection. The developed scheme generated a GT image that solved the problem of correct detection of Copy-move forgery. The GT images were configured using semantic segmentation and image classification. The GT images generated were adopted by other classification techniques that were helped by using a deep neural network. However, it is necessary for performing advanced research to detect the multiple class for moving the patches further.

Monika [18] utilized a Discrete Cosine Transform approach for performing image forensic investigation. The developed model showed its robustness with post-processing and pre-operations to detect automatically and localize the specific artifacts. The developed model used Discrete Cosine Transform technique that obtained features from each block of images and reduced the block dimensions. The tampered blocks of images were compared with the threshold values based on robust parameters for detecting the blocks which are similar in reduced time. However, the results obtained were having higher geometric disturbance in the image quality.

Abhishek & Jindal [19] developed a Deep CNN model for segmentation based on the localized features to detect image forgeries. The DCNN uses 91 layers for detecting the forgery region that performed image forgery detection and localization better. The developed transfer learning showed a major advantage that worked well with small data and also retrained the model with it. The developed model used block-based image methods to localize the forgery region. However, the block-based method divided the image into overlapping blocks which showed complexity in the time.

Structural Correlation Dependent Methodology for Image Forgery Classification and Localization has been proven by Nam Thanh Pham *et al.,* [20]. The suggested technique makes use of Hamming Embedding (HE)-based image retrieval and Bag-Of-Features (BOF) image representation. The Image Forgery Clustering (IFC) used the structural connections among images to group pertinent images

into clusters. The suggested algorithm extracts the cluster centroid from image clusters, that had only one genuine image in the cluster, by taking advantage of the structural correlation between images. The forged areas were then localised and image forgery was classified. Whenever the duplicated areas vary only slightly from the original, the suggested technique works more effectively compared to other approaches. However, it becomes less effective as the rotation angle and scaling ratio increase.

In this study [21], six Image Quality Assessment (IQA) metrics, including Signal to Noise Ratio (SNR), Peak Signal to Noise Ratio (PSNR), Global Contrast Factor (GCF), Normalized Absolute Error (NAE), Average Difference (AD), and Misclassification Error (ME), were investigated. The research aimed to understand the impact of each IQA metric on tested images. The study provided an overview of the background and related work in IQA, highlighting the suitability of SNR and PSNR for contrast images and ME for segmented images. The IQA metrics were categorized based on the image analysis.

In this study [22] a straightforward and efficient method to conceal information within the bit sequence of a Code Excited Linear Prediction (CELP) speech codec. It suggests employing random sequences produced by chaotic maps for this purpose. Our approach utilizes chaotic maps in two key ways: first, to encrypt the secret image by switching among various random sequences generated by different logistic maps, and second, to identify random bit positions in the least significant bits (LSBs) of speech linear prediction coefficients to hide image data.

3. Proposed Methodology

The present research work collects the data that is undergone for the process of pre-processing. The pre- processed image is undergone for the feature extraction and the extracted features are undergone for the process of feature selection. The selected features have undergone the process of classification that classified into original video and spliced video. The block diagram of the proposed method is shown in Figure 1.

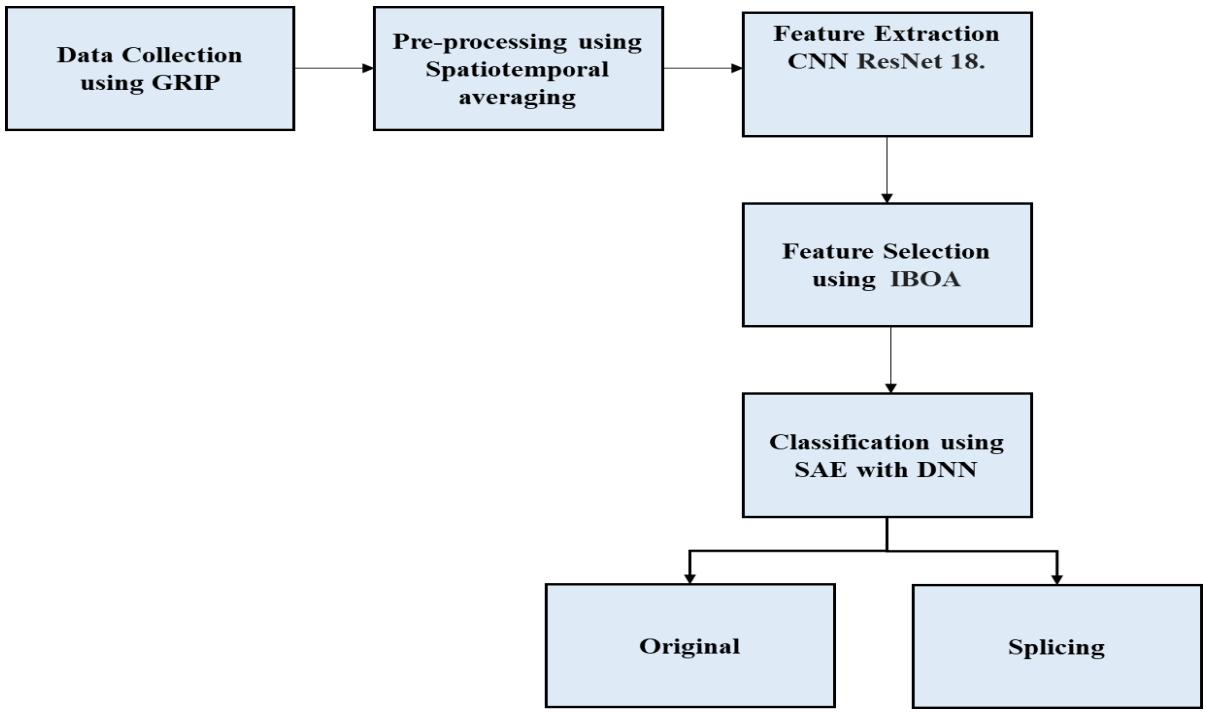

Fig. 1. Block diagram of the proposed research work

3.1 Data Collection

The present research uses Group Research Image Processing (GRIP) dataset that is collected as the test dataset used for marking the bench video forensic techniques. The GRIP dataset has a total 40 number of videos which consists of 10 YouTube, 10 forged, and 10 authentic datasets that comprise compressed videos. The model was compressed with videos that remained with binary masks that were served with the ground truth. The resolution of the video sequence is around 1280×720. The datasets consisted of the most tampered videos which were belonging to the case of the adding object. The dataset was open access that divided the tampers into the temporal and spatial domains. The database consists of a total 33 videos and 26 tampered videos that consisted of 10 splicing, 6 frame swapping, and 10 copy move. The database images are publicly available on the YouTube channel and all of the videos have been shot carefully by considering spatial and temporal video characteristics. Figure 2 shows the GRIP Dataset images.

Fig. 2. GRIP Dataset images

3.2 Pre-processing

The input video is used for the process of pre-processing that divides the video into smaller shots. The shots obtained are non-overlapping sub-sequences where the frames are having z as the size of each of the shot systems. The model was designed based on the number of frames as it showed small variations among the short interval of time. The pre-processing uses the videos which are divided into a number of frames processed to detect the tampered region portions. The spatiotemporal averaging was evaluated to perform the temporal and spatial averaging. At first, spatial averaging is applied to the group operation that estimates the new values for the pixels from the neighborhood pixels based on the template convolution. The spatial average is mathematically defined as shown in Eq. (1).

$$p_i = w \times p_i' \tag{1}$$

Where, p_i is known as the new video frame which is convolving based on the template w that has the frame of the shot p_i'. Here, $w = 3 \times 3$ states with the temporal averaging and weight coefficients that are mathematically expressed as shown in Eq. (2).

$$Spatiotemporal\ averaging = \frac{1}{s}\sum_{i=1}^{s} p_i \tag{2}$$

Spatiotemporal averaging is performed that obtains the resultant frame by combining the spatial and temporal information. The output for the spatiotemporal averaging is performed which is

represented graphically as shown in Figure 2. The spatiotemporal averaging frame consisted of scene background information that has a pale appearance for an object to move through the whole shot.

3.3 Feature Extraction

The splicing analysis plays an important role in the process of feature extraction. The main challenge in the area is to extract the opinions for identifying the forged area to classify it as forged or not. The pre-processed frames are processed for the feature extraction using a Convolutional Neural Network (CNN) which forms a suitable combination of low and high-level features for the extraction of features from the frames. The main aim of the model is to reduce the feature numbers from the dataset that creates a set of new features. These features from the existing model improved the accuracy and also overcomes the overfitting risk. The ResNet is used as a Residual Network the classic neural network that acts as a backbone to perform various tasks. The ResNet model has trained DCNN extremely with 150 layers that perform fundamental breakthroughs successfully. The ResNet training before the deep neural network showed difficulty because of the vanishing gradient problems. The AlexNet model has started for focusing on deep learning models that consisted of convolutional layers. ResNet training is a deep neural network that showed difficulty with the problem of vanishing gradients. The AlexNet model started to focus on deep learning as it consisted of 8 convolutional layers, the VGG network consisted of 19 layers, Inception or GoogleNet had 22 layers, and ResNet had 152 layers showed repeating multiplication showed gradient small.

The ResNet-18 is a type of CNN model which has a total 18 number of layers and is pre-trained on more than a million images using the ImageNet database. The ResNet-18 network learns the rich features to represent a variety of wide range of images. The image is having input size of 224-by-224 in the network which is a pre-trained network by using MATLAB for pre-training the DNN. The images have been classified the images using the ResNet-18 model which Classified the image based on GoogLeNet that can replace the GoogLeNet with ResNet-18. The network is retrained based on the new classification task which follows by training a deep learning network to classify the images and ResNet.

3.4 Feature Selection

Before processing for classification, it is important for assessing the feature space which is created from the data. The high-dimensional feature spaces showed difficulty for the classifiers for identifying the data patterns. The useful features from the smaller subset are performed by the selection of those features showed improvement in the classifier performances. The model reduced the computational requirements and storage during classifier training. The most significant step is performed by reducing the feature space which quantifies the underlying patterns in the data.

In the present research work, the Bat Algorithm (BA) is used as a meta-heuristic algorithm that works mainly based on bat characteristics. It finds the echolocation widely with respect to the bat characteristics that overcome the problems of optimization when appeared. The BA has insufficient local search and showed poor performances due to high dimensional optimization issues, lack of diversity, and insufficient local search. An improved BA utilizes an extremal Optimization algorithm that improved the performance of BA. The IBA-EO model has improved the updated strategy obtained solutions to generate the randomly selected bats for enhancing the global search capacity. The ability to exploit is shown by updating the strategy in the proposed work which has gained a solution for randomly selecting the bats to enhance the search capability. The ability to exploit the EO has shown excellence with the local search. The EO algorithm with BA is used for enhancing the

exploitation ability. The aforementioned techniques are used for improving and monitoring the mechanism. It is used for keeping a suitable balance between exploration and exploitation ability.

3.4.1 An improved updating strategy

The position of BA is updated in two ways where the new position is generated based on frequency and speed that is adopted with the pulse rate decides X_i. It develops towards an optimal solution and both of them have been determined correctly and effectively. The exploration of the capacity is promoted that played a better role for low dimensional space. The high dimensional complex finds the search spaces that focused mainly on the search direction explored with a globally optimal solution. The condition after acceptance obtains a new solution that needs a smaller objective function when compared with the current optimal solution. The generated random number satisfies smaller values than the policy parameter. The acceptance strategy has caused the loss of excellent candidate solutions which has seemed with the search space that is not expanded fully. The positions of the bats are updated as the new strategy is set up. The detailed expression is provided in Eq. (3) and Eq. (4).

$$X_i = X_i^{t-1} + v_i^t + (X_i^{t-1} - X_k) \times \delta \tag{3}$$

$$\delta = I_{max} - \frac{t}{I_{max}} + 0.1 \tag{4}$$

From the above Equations, subscript i is known as the i^{th} bat and k mean random integer having 1 to NP that is number having the bat population$i$. I_{max} is known as the maximum iteration number and t is the number of the current iteration, δ is known as the control parameter. For the early search, large δ is used for enhancing the diversity among the population for producing an optimal solution. Later, the step size become smaller and showed randomness was weaker. The exploration of the optimal solution was strengthened and a new selection strategy was developed for accepting it with a better solution. The worse solution is accepted for certain probability functions when the two cases were operating. In case the function value is better, then unconditional things are accepted. If the obtained solution is better, then the unconditional solutions are generated. If in case the objective function value is worse, then the solutions are accepted. The poor candidates have achieved the expanding effect of a search field that is accepted moderately. The obtained model makes the solution fall into the local optimal jump-through predicament.

3.5 Classification

Once the best features are found from the Improvised Bat Algorithm, the data is classified using a classification algorithm. The datasets are having a relationship complexity with features. Using a single Auto encoder is not sufficient and a single autoencoder is unable to reduce the input features' dimensionality. Thus, in such cases, a stacked autoencoder is used that has multiple encoders that are stacked from the top. The SAE has 3 encoders that are stacked on one another and the SAE has 3 encoders that are stacked as shown in the description. The output obtained by encoder 1 and input of the autoencoder 1 is provided with an input for autoencoder 2. The output from autoencoder 2 and the input from autoencoder 2 is provides input to encoder 3 as an input. Therefore, the input length of autoencoder 3 will be double autoencoder 1 and 2’s input. Thus, the model solves the insufficient data problem to some extent.

3.5.1 Implementing stacked autoencoders using python

The SAE uses the Fast Fourier Transform (FFT) for the vibrated signal that is used for fault diagnosis with various applications. The data consisted of complex patterns and thus a single autoencoder does not reduce the data dimension. The amplitude of the FFT is ranging between 0 and 1. The autoencoder is designed over dense layers with respect to both the decoder and encoder sides. The number of neurons in the encoder and the decoder is the same. The autoencoder reduces the number of features that build the model compiled and fitted to train to reduce the features to 200 from 4000. The model is built and compiled for fitting the training data. The autoencoder targets the output which is the same as that of the input. The model is trained at the first by the autoencoder and all the output generated is concatenated. The auto-encoder 2 is ready for considering the input. The model is built compiled and trained to autoencoder 2 on a new dataset. The autoencoder 2 is trained that moves toward training 3rd autoencoder and for the next autoencoder, the input to the 3rd encoder is used for the last 2 encoders which compile, trains the data generated at the new. In this research, the undertaken classification method classifies the types of class such as splicing and original as shown in Figure 3.

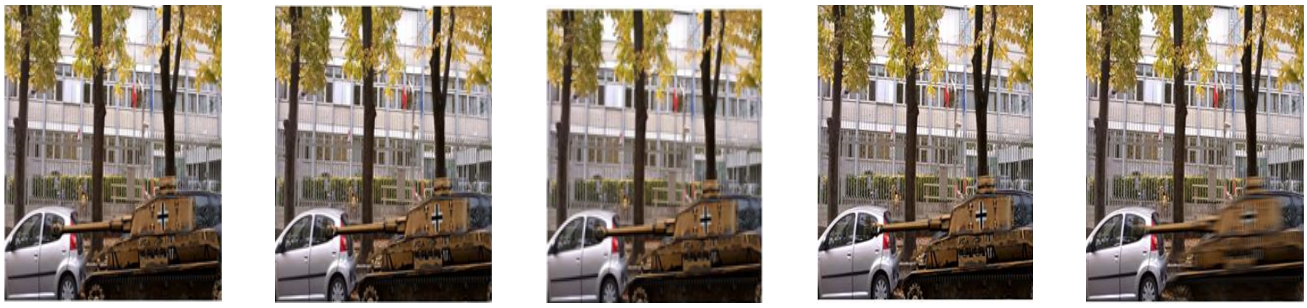

Fig. 3. Classified forgery images

4. Experimental Results

The metrics for the proposed research evaluate the performances using image forgery detection. The Forged images (fi) are classified correctly TP and the wrongly classified images are termed FP (False Positive). The images tampered with are falsely missed that is known as the False Negatives (FN) and which is authenticated under correctly classified images expressed as shown in the below Eq. (5) to Eq. (9)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

$$Precision = \frac{TP}{TP+FP} \times 100 \tag{6}$$

$$MCC = \frac{TN \times TP - FN \times FP}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{7}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{8}$$

$$Specificity = \frac{TN}{TN+FP} \tag{9}$$

Where, TP=True Positive, TN=True Negative, FP= False Positive, FN=False Negative. The K-fold cross-validation is performed that splits data into different folds as$k$. The present research validated results for $k = 3,5,8,10$ and the subsets train the data to leave the last fold as the test data. The model is averaged against all numbers of folds for model finalization. The training and testing percentages for the present research work are provided in Table 1.

Table 1
Training and testing percentage for distinct K values

K values	3.00	5.00	8.00	10.00
Training	66%	80%	85%	90%
Testing	34%	20%	15%	10%

Figure 4 shows the original image and forgery image.

Fig. 4. Original image and forgery image

Figure 5 shows the classified forgery images.

Fig. 5. Classified forgery image

4.1 Quantitative Analysis

Table 2 shows the results obtained for the proposed method where the results are obtained for accuracy, precision, sensitivity, specificity, and MCC. The classifiers considered are MSVM, LSTM, RNN, Sparse auto encoder and DNN which obtained various accuracy values. The classifiers such as MSVM, LSTM, RNN, Sparse auto encoder and DNN outperform well to obtain an accuracy of 78.51%, 95.16 %, 94.38%, 95.85% and 98.95%. The existing SVM does not perform well as the dataset consisted of more noises and the target classes were overlapping. When the feature numbers for each data point exceed the number of training data samples, the SVM underperforms. Similarly, the KNN model doesn't work well with a large dataset which is having a high number of dimensions obtained lower than 94.8%. The RF failed to work with large trees which makes the algorithm too slow and ineffective for real world applications. The developed algorithm trained the model fast but was slow for predicting the trained accuracy of 96.54%. However, the existing Neural Network required to reduce the error for the sample means that completed the training without providing optimum results.

Table 2
Results obtained by the proposed classifiers without feature selection algorithms

Without feature selection					
Classifiers	Accuracy	Sensitivity	Specificity	MCC	Precision
MSVM	65.8572	64.2253	66.6769	65.0606	67.8615
LSTM	91.069	92.5065	91.57	90.46	90.65
RNN	89.06	90.24	91.75	89.98	88.25
Sparse auto encoder	92.54	92.06	91.6638	90.011	88.65
DNN	90.56	90.26	91.77	89.90	91.82

Figure 6 and 7 shows the performance analysis in the stage of feature selection process.

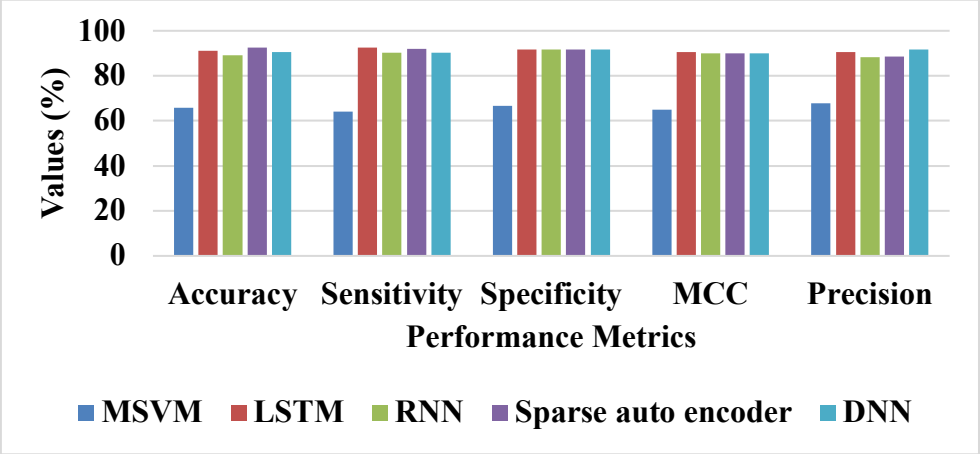

Fig. 6. Performance analysis without using feature selection process

Table 3 shows the results obtained for the proposed method with feature selection. The classifiers such as SVM, KNN, RF, NN, and DNN obtained lower accuracy values without the feature selection algorithm but obtained better accuracy with the feature selection algorithm.

Table 3
Results obtained by the proposed classifiers with feature selection algorithms

With feature selection					
Classifiers	Accuracy	Sensitivity	Specificity	MCC	Precision
MSVM	78.51	77.28	79.50	76.77	78.67
LSTM	95.16	94.85	94.87	95.81	93.17
RNN	94.38	93.87	92.09	94.06	91.74
Sparse auto encoder	95.85	96.64	96.96	95.98	97.73
DNN	98.95	99.46	98.47	97.29	99.93

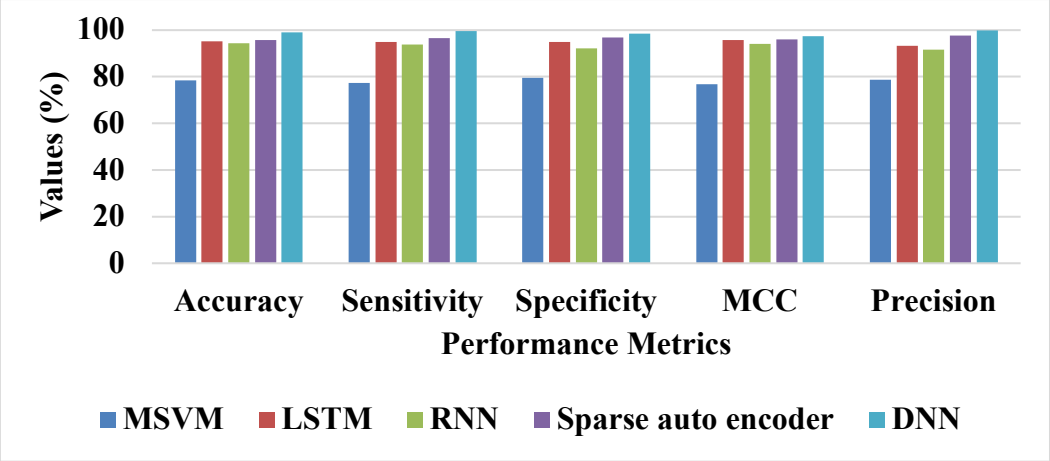

Fig. 7. Performance analysis with feature selection method

Table 4 shows the results obtained by the proposed research for the evaluation of distinct optimization algorithms to various k-fold validations. The present research work utilizes PSO, ACO, and IBA optimization algorithms for validating the performances. The accuracy performances for k fold validation performed obtained 96.6% of accuracy for the proposed method. The existing particle swarm optimization (PSO) algorithm used for finding the local optimum showed higher performances. However, the ACO algorithm overcomes the optimization problems for further improvements and obtained better accuracy values. The IBA showed relatively larger values with the higher convergence rate and the delayed phenomenon at the later stage speeded and slowed down the performances. The developed model does not provide a solution but required a smaller objective function when compared with the currently obtained optimal solution. The model has satisfied the random numbers generated and showed lesser accuracy values compared to the policy parameter.

Table 4
Performance metrics evaluation for different optimization algorithms for different k-fold validation

Performance metrics	Optimization algorithms	K-fold validation			
		3	5	8	10
Accuracy	PSO	82.36	83.14	82.22	79.43
	ACO	85.79	86.31	84.71	82.28
	IBA(Existing)	88.89	90.52	89.12	86.08
	Proposed	96.60	98.60	97.49	94.05
Sensitivity	PSO	83.18	84.73	81.31	80.98
	ACO	85.73	86.72	83.17	82.74
	IBA(Existing)	89.42	90.66	90.15	87.60
	Proposed	96.77	99.46	97.18	94.30
Specificity	PSO	82.91	83.00	83.38	79.12
	ACO	84.01	85.68	84.20	83.16
	IBA(Existing)	88.54	89.77	89.49	87
	Proposed	97.15	98.47	96.10	93.03
MCC	PSO	82.81	84.01	83.92	80.65
	ACO	83.86	84.58	82.29	82.15
	IBA(Existing)	89.29	90.80	90.35	88.82
	Proposed	96.20	97.29	95.90	94.44
Precision	PSO	82.29	85.57	84.50	78.66
	ACO	83.11	85.30	83.36	81.31
	IBA(Existing)	90.27	91.41	89.94	89.07
	Proposed	96.43	99.93	96.10	93.59

4.2 Comparative Analysis

The Group Research Image Processing (GRIP) dataset, which is gathered as a test dataset for marking the bench video forensic methods, is used in the current study. While considering the GRIP dataset, existing Discrete Cosine Transform [15] obtained better accuracy of 98%, and also a higher geometric disturbance of image quality was needed. However, DCNN [11] block-based method divided the image into blocks resulting in complexity in the time obtained with an accuracy of 91.1%. Whereas, the proposed method was forged exposure for the videos that showed higher resolution images with long lengths. While existing IFC [17] has achieved 98.8% sensitivity and 96.3% precision. Overall, the proposed method showed better performances of achieving higher accuracy as 98.60% in GRIP dataset. Table 5 shows the comparative analysis of the proposed and the existing models on GRIP dataset.

Table 5
Comparative analysis on GRIP dataset

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	MCC (%)	Precision (%)
DCNN [11]	91.1	-	-	-	-
Discrete Cosine Transform [15]	98	93	-	-	96
IFC [17]	-	98.8	-	-	96.3
Proposed improved BAT with SAE	98.60	98.60	98.47	97.29	99.93

5. Conclusions

The present research uses Spatio-temporal averaging model which was used for collecting the video sequences extracted the background information. It is having a pale of moving objects which showed an effective forgery detection in the video. The ResNet 18 was used for performing the feature extraction for obtaining the feature vectors. The discriminative feature vectors were reducing the training time that improved the accuracy of detection. The single Autoencoder was unable to reduce the dimensionality of input features. Therefore, this research introduced stacked autoencoders which consist of multiple encoders that were stacked on one another to improve the classification accuracy. The proposed method obtained 98.6% accuracy, sensitivity (98.60%), specificity (98.47%), MCC (97.29%) and precision (99.93%) which is better when compared with the existing models. In the future, this research will be further extended by using hybrid deep learning models to enhance the classification performances.

Acknowledgement
This research was not funded by any grant.

References
[1] Singh, Gurvinder, and Kulbir Singh. "Video frame and region duplication forgery detection based on correlation coefficient and coefficient of variation." *Multimedia Tools and Applications* 78 (2019): 11527-11562. https://doi.org/10.1007/s11042-018-6585-1

[2] Al-Sanjary, Omar Ismael, Nurulhuda Ghazali, Ahmed Abdullah Ahmed, and Ghazali Sulong. "Semi-automatic methods in video forgery detection based on multi-view dimension." In *Recent Trends in Information and Communication Technology: Proceedings of the 2nd International Conference of Reliable Information and Communication Technology (IRICT 2017)*, pp. 378-388. Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-59427-9_41

[3] Jia, Wei, Li Li, Zhu Li, Shuai Zhao, and Shan Liu. "Scalable hash from triplet loss feature aggregation for video de-duplication." *Journal of Visual Communication and Image Representation* 72 (2020): 102908. https://doi.org/10.1016/j.jvcir.2020.102908

[4] Fayyaz, Muhammad Aizad, Adeel Anjum, Sheikh Ziauddin, Ahmed Khan, and Aaliya Sarfaraz. "An improved surveillance video forgery detection technique using sensor pattern noise and correlation of noise residues." *Multimedia Tools and Applications* 79 (2020): 5767-5788. https://doi.org/10.1007/s11042-019-08236-2
[5] Vinolin, V., and M. Sucharitha. "Dual adaptive deep convolutional neural network for video forgery detection in 3D lighting environment." *The Visual Computer* 37 (2021): 2369-2390. https://doi.org/10.1007/s00371-020-01992-5
[6] Jia, Shan, Zhengquan Xu, Hao Wang, Chunhui Feng, and Tao Wang. "Coarse-to-fine copy-move forgery detection for video forensics." *IEEE Access* 6 (2018): 25323-25335. https://doi.org/10.1109/ACCESS.2018.2819624
[7] Aloraini, Mohammed, Mehdi Sharifzadeh, and Dan Schonfeld. "Sequential and patch analyses for object removal video forgery detection and localization." *IEEE Transactions on Circuits and Systems for Video Technology* 31, no. 3 (2020): 917-930. https://doi.org/10.1109/TCSVT.2020.2993004
[8] Saddique, Mubbashar, Khurshid Asghar, Usama Ijaz Bajwa, Muhammad Hussain, and Zulfiqar Habib. "Spatial Video Forgery Detection and Localization using Texture Analysis of Consecutive Frames." *Advances in Electrical & Computer Engineering* 19, no. 3 (2019). https://doi.org/10.4316/AECE.2019.03012
[9] Fadl, Sondos, Amr Megahed, Qi Han, and Li Qiong. "Frame duplication and shuffling forgery detection technique in surveillance videos based on temporal average and gray level co-occurrence matrix." *Multimedia Tools and Applications* 79 (2020): 17619-17643. https://doi.org/10.1007/s11042-019-08603-z
[10] Su, Lichao, Cuihua Li, Yuecong Lai, and Jianmei Yang. "A fast forgery detection algorithm based on exponential-Fourier moments for video region duplication." *IEEE Transactions on Multimedia* 20, no. 4 (2017): 825-840. https://doi.org/10.1109/TMM.2017.2760098
[11] Long, Chengjiang, Arslan Basharat, Anthony Hoogs, Priyanka Singh, and Hany Farid. "A Coarse-to-fine Deep Convolutional Neural Network Framework for Frame Duplication Detection and Localization in Forged Videos." In *CVPR workshops*, pp. 1-10. 2019.
[12] Nguyen, Xuan Hau, Yongjian Hu, Muhmmad Ahmad Amin, Gohar Hayat Khan, and Dinh-Tu Truong. "Detecting video inter-frame forgeries based on convolutional neural network model." *International Journal of Image, Graphics and Signal Processing* 10, no. 3 (2020): 1. https://doi.org/10.5815/ijigsp.2020.03.01
[13] Afchar, Darius, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. "Mesonet: a compact facial video forgery detection network." In *2018 IEEE international workshop on information forensics and security (WIFS)*, pp. 1-7. IEEE, 2018. https://doi.org/10.1109/WIFS.2018.8630761
[14] Kaur, Harpreet, and Neeru Jindal. "Deep convolutional neural network for graphics forgery detection in video." *Wireless Personal Communications* 112 (2020): 1763-1781. https://doi.org/10.1007/s11277-020-07126-3
[15] Chen, Beijing, Ming Yu, Qingtang Su, Hiuk Jae Shim, and Yun-Qing Shi. "Fractional quaternion Zernike moments for robust color image copy-move forgery detection." *IEEE Access* 6 (2018): 56637-56646. https://doi.org/10.1109/ACCESS.2018.2871952
[16] Jin, Xiao, Zhen He, Yongwei Wang, Jiawei Yu, and Jing Xu. "Towards general object-based video forgery detection via dual-stream networks and depth information embedding." *Multimedia Tools and Applications* 81, no. 25 (2022): 35733-35749. https://doi.org/10.1007/s11042-021-11126-1
[17] Rhee, Kang Hyeon. "Generation of novelty ground truth image using image classification and semantic segmentation for copy-move forgery detection." *IEEE Access* 10 (2021): 2783-2796. https://doi.org/10.1109/ACCESS.2021.3136781
[18] Monika, Dipali Bansal, and Abhiruchi Passi. "Image Forensic Investigation Using Discrete Cosine Transform-Based Approach." *Wireless Personal Communications* 119 (2021): 3241-3253. https://doi.org/10.1007/s11277-021-08396-1
[19] Abhishek, and Neeru Jindal. "Copy move and splicing forgery detection using deep convolution neural network, and semantic segmentation." *Multimedia Tools and Applications* 80 (2021): 3571-3599. https://doi.org/10.1007/s11042-020-09816-3
[20] Pham, Nam Thanh, Jong-Weon Lee, and Chun-Su Park. "Structural correlation based method for image forgery classification and localization." *Applied Sciences* 10, no. 13 (2020): 4458. https://doi.org/10.3390/app10134458
[21] Mustafa, W. A., H. Yazid, M. Jaafar, M. Zainal, A. S. Abdul-Nasir, and N. Mazlan. "A review of image quality assessment (iqa): Snr, gcf, ad, nae, psnr, me." *Journal of advanced research in computing and applications* 7, no. 1 (2017): 1-7.
[22] El-Khamy, Said E., Noha O. Korany, and Marwa H. El-Sherif. "Chaos based secure image hiding in variable bit rate CELP speech coding systems."