Utilizing Machine Learning Techniques to Assess Technical Document Quality

Muhammad Junaid Iqbal[1*], Fabio Massimo Zanzotto[2], Usman Nawaz[3]
Department of Enterprise Engineering, University of Roma tor Vergata, Rome, 00133, Italy[1, 2]
Department of Engineering, University of Palermo, Palermo, Italy[3]

***Abstract*—Information is disseminated through images in newspapers, periodicals, the internet, and academic journals. With the aid of various tools such as Adobe, GIMP, and Corel Draw, distinguishing between an original image and a forgery has become increasingly challenging. Most conventional methods rely on constructed traits for detecting image counterfeiting. Image verification plays a crucial role in securing and ensuring the authenticity of individuals' identities in sensitive documents. This research proposes a machine learning approach (Support Vector Machine, SVM, and Histogram of Oriented Gradients, HOG) to identify images and confirm their authenticity. The Histogram of Oriented Gradients (HOG) is employed to extract diverse features including matching, image size, and dimensions for image verification. The training and testing phases are carried out using a Support Vector Machine (SVM). The proposed image verification technique is evaluated using extensive datasets to ascertain image recognition accuracy, alongside metrics such as specificity, sensitivity, and precision. Comparative analysis with existing techniques reveals that the average image verification accuracy of the proposed method stands at 98%, surpassing previous image verification methods.**

Keywords*—Image verification; machine learning; ensemble approach; multi-feature image recognition

I. Introduction

Images are now increasingly one of the primary sources of information and are essential in various disciplines, including medicine, education, computer forensics, sports, and the media. Thanks to tools like Adobe Photoshop, GIMP, Coral Draw, and Android apps like Photo Hacker, creating a fake image is surprisingly simple. When a picture is presented as evidence in court, its veracity becomes extremely important. Any operation performed on digital photographs using ana program is called image manipulation, or "image editing”. Image forging is a technique that alters an image's content to make it inconsistent with historical events. Image manipulation is if the new content is copied from the same image itself, then it is called copy-move tampering, and if the new content is copied from a different image, then it is called image splicing [1, 2]. The methods for detecting picture alteration can be divided into two categories: (i) active and (ii) passive. In an active approach, a person with authorization embeds extra details (such as a digital watermark) into the image either during the acquisition phase or later.

This embedded data is used by the active technique to detect manipulation. The passive methods do not rely on extra information to detect forgeries. These methods are sometimes known as "blind approaches" because they don't require additional information to detect forged documents. The passive methods take the image's features and utilize them [3].

In our proposed method fake images made using pixels detection is to confirm the veracity of electronic pictures without providing access to the source image. We proposed machine learning algorithms for image processing and feature extraction recognition by using the copy move forgery. The objective of duplicate forgery is to replicate or conceal an object by cutting it out of one region of the image and putting it into another [4, 5]. Post-processing on altered photos, however, can make the work of spotting instances of forgeries far more difficult.

The proposed idea deals with image verification performance as well as image verification accuracy.

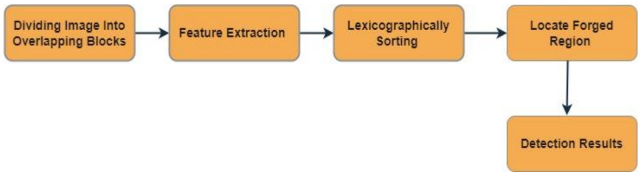

Dividing Image Into Overlapping Blocks
Feature Extraction
Lexicographically Sorting
Locate Forged Region
Detection Results

Fig. 1. Basic steps for pixel-based forgery image detection.

II. Literature Review

With picture editing software, digital photos may readily be altered. It is critical to detect manipulation attempts. Without prior knowledge of the source photos, passive digital picture tampering detection tries to confirm the validity of digital photographs. In recent years, numerous strategies have been proposed in this field [6, 7]. This work presents the three tiers of these methods low-level, mid-level, and high-level. At each level, the essential concepts of the suggested approaches are discussed in detail along with some remarks. The authors in [8] proposed a deep learning-based algorithm to detect fake images in order to recognize the image and retrieve its information, this technique employs a convolution neural network design. What investigation utilized the MICC-F200 dataset? Design process parameters were used to assess the model's performance. The accuracy of the model was 95.5%.

Numerous researchers have already developed several approaches for detecting pixel-based image counterfeiting. In study [9, 10] author provided a framework to recognize fake paper photocopies using the bounding box technique. This technique mainly focuses on identifying copies of documents that have been edited by adding new text above it, smearing whitener over the old text, and then editing the contents using the cut-and-paste technique. The effectiveness of this technique is about 86%. The advantage of this method is that it does not need expensive hardware. It does not function when photocopied documents have background art or dust. The work can be improved by making better use of the bounding box features. For better results, a single strategy rather than a hybrid can be utilized for categorization. One of the most common types of picture forgeries was first presented by Paul and other writers in 2019. The SURF and k-NN algorithms are the foundation of this method. In addition to using K-NN for training and mapping, Paul et AL Speeded-Up's Robust Features method extracts essential details from the image. Compared to SIFT-based approaches, this method promises cheaper processing costs while displaying higher accuracy. SURF- based methods, however, do not consistently follow the edges[11, 12].

In study [13-15] authors suggested a pixel-based method for spotting fake images, which uses the Columbia DVMM dataset, which is openly accessible. This strategy is based on Hilbert-Huang transforms (HHT) and support vector machines (SVM). SVM was utilized as a classifier, and HHT approaches were applied for feature extraction. The test is conducted in MATLAB, and the evaluation parameters are used to derive results for three metrics: true negative (80.25%), true positive (80.03%), and accuracy (80.15%).

The proposed method will be developed using MATLAB 2013a as a tool, and it is crucial to focus on the libraries and methods required to run the suggested strategy. Framework.NET will be followed Working Approach of Proposed Method:

Step 1: The proposed system architecture will give an image for training.

Step 2: After that, pre-processing functions are applied for image processing and covert RGB image in grayscale image.

Step 3: Feature extract from the images.

Step 4: Train the images on a proposed method based on matching objects, speed, and Edge pixels. Step 5: Verifywhether the image is original or forgery.

A. Significance Research

In every area of life, verifying the picture has grown to be a significant difficulty. Most verification algorithms have low performance and accuracy. What makes the recommended method notable is its utility in highlighting the benefits of TP, TN, FP, and FN. The proposed approach can be utilized to validate the image and demonstrate correctness by utilizing several evaluation criteria.

B. Image Acquisition

The original and fake image is acquired from the gallery at this stage.

C. Pre-processing

The color conversion is carried out in the second stage of our process. The following formula is used to convert the RGB image first into the grayscale image I: I = 0.299R + 0.587G + 0.114B. It stands for the brightness component, where R, G, and B are the input color image's red, green, and blue channels. In pre- processing, images are selected as"original images" or "fake images." The image dataset's extension should be (. JPJ, .PNG, .PGM, .TIFF).Pre-processing of the image can be done using image processing techniques that involve 2D and 3D (R, G, and B)with the size of (Im, 3) = 3. The IF image is a colour image that has been converted into gray. Some functions are pre-processed, such as Im = rgb grey (Im); end. In this study, the Kaggle datasets have been used for pre-processing and feature extraction, which are publicly available.

III. METHODOLOGY

The proposed method has used well-known techniques to make image recognition whether the image is fake or original. The evaluations matric for the proposed method included precision, accuracy and recall. The dataset used to evaluate the suggested strategy is available to the public. However, some random dataset has also been taken for evaluation to show the accuracy of the proposed method. The SVM and HOG schemes have been used to verify the original or fake image, which contains four fundamentalphases.

There are four steps in this image recognition process. In the first step, the image is acquired, pre-processing is donein the second step, and the image features are extracted in step three. The last step is image forgery detection. Fig. 1 explains each of these steps.

- Image Acquisition
- Pre-processing
- Feature Extraction
- Forgery Detection

The research makes use of a dataset of pictures. The image dataset has been converted to CSV format. The dataset is pre-processed before the CNN method is applied. The categorization of images is finished. Lastly, it is possible to determine whether the image is false matched, time may be saved, especially if feature extraction methods like DCT or PCA are used.

The suggested work has the following phases.

- The image is divided into corresponding blocks of a fixed size.
- Features extract of each block using HOG descriptors.
- Similar block pairs correspond then SVM is used to find whether the image is a forgery or genuine.
- Lastly, a bounding box is created for copied areas.

The proposed method is shown through data flow diagrams. The DFD is also known as a bubble chart. It is a simple graphical structure that may be utilized to describe a system in terms of the data input, the various operations carried out on it, and the information created as an outcome of those activities.

The data flow diagram is the most vital modeling tool. It is used to construct the component models for the system. These components include how the system works, the information it uses, how a third party interacts with it, and how data flows through it. DFD displays the system's information flow as well as the numerous modifications that have an impact. It uses graphics to show how information moves and how data is altered as it moves from source to output.

Table I compares a number of image forgery detection techniques. For every entry, there is a list of the recognition technique, feature extraction strategy, datasets used, photo forgery type addressed, recognition parameters, achieved accuracy, and researchers involved. The table shows that the recommended SVM strategy employing HOG features achieved the highest accuracy of 98% in copy-move forgery detection. This suggests significant advancements in digital image forensics.

In the chosen image from the dataset, shown in Fig. 2, several objects captured in the scene are depicted. These images have been submitted for pre-processing to lower noise and unused pixels. After that, features are extracted using the feature extractor function (HOG). Histogram of Ordered Gradients is a pattern extraction method comparable to Scale Invariant and Fourier Transform (SIFT) Canny Edge Detection. It is employed in computer recognition and image processing for object detection. An image dataset is used for both training and testing the SVM classifier. The Confusion matrix is then used to examine the SVM findings.

A. Tools and Technology

The suggested technique has been implemented in MATLAB. MATLAB R2013A was used to experiment with the suggested technique of system implementation while running under Windows 7's 64-bitoperating system. The desktop computer has a Pentium processor and 1 GB of physical memory (RAM). Dual Core processor central processing unit (CPU). A keyboard and mouse are used for input. The suggested system was made using a variety of programs and libraries.

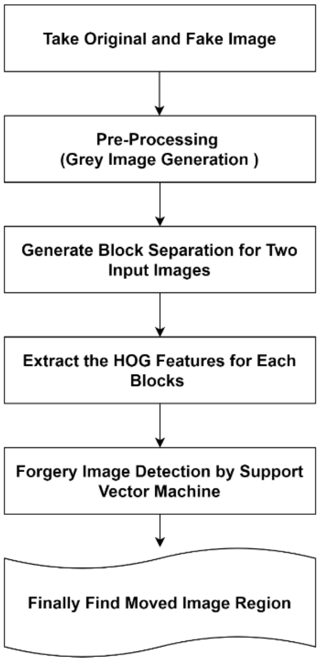

Take Original and Fake Image
Pre-Processing
(Grey Image Generation)
Generate Block Separation for Two Input Images
Extract the HOG Features for Each Blocks
Forgery Image Detection by Support Vector Machine
Finally Find Moved Image Region

Fig. 2. Image recognition process of suggested algorithm.

TABLE I. COMPARATIVE ANALYSIS OF EXISTING APPROACHES

Sr.no	Recognition Techniques	Feature ExtractionMethod	Datasets	Image Forgery Technique	Recognition parameters	Accuracy	Researchers
1	Improved Relevance Vector Machine (IRVM) Used for forgery detection.	Biorthogon al Wavelet Transform with Singular Value Decomposition (BWT SVD)-based feature extraction	The inputdataset has been downloaded from thewebsite	Copy-Move based image Forgery (Blocked)	Variance, mean, skewness, energy, etc.	Accuracy rate of 92.22%	Rathore, Neeraj Kumar, et al., 2021
2	The suggestedtechnique uses the Scale Invariant Feature Transform (SIFT) and Fuzzy C-means (FCM) for clustering. .	SIRF	MICC-220 Dataset 25 3 Images	Copy-Move based image Forgery (Ke y point based)	Number of Clusters Maximum no of Iteration	Standards for accuracy and small improvement in somecases.	Alberry, Hesham A.,Abdelfatah A. Hegazy, and GoudaI.Salama. 2018
3	Shallow Convolution al Neural Network (SCNN)	Extractfeature Vectors with dimension	CASIA 2.0 Dataset51 23 images	Image Splicing	Dimensions	80.91% Accuracy	Zhang, Zhongping,et al., 2018
4	Novel similarity metric combiningcosine	To extract the facial landmarks, ORB is utilized.	PASCALVOC MIC-F22072 Images	Dimension	Similarity Translation, rotation, noise, Illumination and JPEG compression.	83.33 % Accuracy	Tian, Xiuxia, Guoshuai Zhou, andMan Xu 2020
5	Enhancement of Relevance Vector Machine	Singular Value Biorthogonal Wavelet Transform Decomposition (BWT-SVD)	http://www.vcl.fef.hr /comofod/dowdownl .ht ml Datasets	Principle Points	Not clear	92.22 %	Rathore, N. K., Jain, N. K., Shukla, P. K., Rawat, U., & Dubey, R. (2021).
Proposed method	SupportVector Machine (SVM)	HOG	MICC_F60 MICC-F220 MICCF8multiCoMo FoD_small_v2 Local Dataset	Copy Move (Pixels)	Similarity, Translation rotation, Pixel value	98%	

TABLE II. TOOLS AND LIBRARIES USED IN PROPOSED METHOD

MATLAB 2013 A languages	**Description**
FbgTrainMem	This information is necessary for performingrecognition.
Normalized image.m	This method is used to determine the image sizeand twice its size
Divide DB.m	Each feature is represented as a matrix with thedimensions feature-length x total features.
Calculate results .m	Do the True Positive & Negative calculations. Thenumber of true classes in classes 1 and 2 of the confusion matrices

Table II lists some significant functions that MATLAB utilizes to carry out algorithms. Other significantMATLAB 2019a libraries are used to support the suggested system

TABLE III. SOME OTHER LIBRARIES USED IN THE PROPOSED METHOD

MATLAB Libraries	Description
LIBSVM	For picture training and testing, the SVM classifier is applied in MATLAB.
Sklearn	This library is used for implementing the Support vector machine
y = f(x)	Display the findings in columns and rows

Table III shows the MATLAB libraries and its descriptions. The suggested solution was developed in MATLAB 2013A, a more productive programming language, Application-specific software. It can be considered an external library and is used to implement our approach. It offers many library functions that are simple to use for personal authentication the image recognition. The suggested techniques can be implemented in a variety of computer languages, but for our study, we chose MATLAB2013A.

B. Forgery Detection

A feature match is finished after the two photos' traits have been extracted. After that, the noise in the area containing the counterfeit was removed using a wavelet transform. The final region is sent as input to the SVM for forgery detection. According to the SVM output, a score of 0 denotes authenticity, and a score of 1 indicates forgery. Overfitting issues can be resolved using Support Vector Machines (SVM), even though this technique is normally associated with the classification. There is no problem handling many continuous and categorical variables. SVM is used to create a hyperplane in multidimensional space to divide different classes. SVM iteratively creates an ideal hyperplane to decrease error. The primary objective of SVM is to locate an MMH that best classifies the dataset [16].

The most crucial task of a duplicate image forgery identification system is finding out whether a given image has duplicated portions. Intelligently speaking it is challenging to evaluate each pair of areas individually, pixel by pixel, because of post-processing processes like rotation, scaling, blur deterioration, and changes, in contrast. In comparison, cannot be known in advance [17, 18].

An image produced by a system is recognized by the specific standards diagram shown in Fig. 3. Despite the effectiveness of various approaches for validating photos, they may not be able to identify complicated false images. It does not make the owner's verification secure. The design of the proposed system is divided into three groups.

C. Datasets

For image recognition datasets, there are numerous ones that are openly accessible. We used images as a document. The suggested method is assessed using the datasets. The dataset has been utilized in other research studies. MICC_F600, MICC-F220, MICC- F8multi, CoMoFoD_small_v2 Dataset have been used in the proposed method.

The goal of the recommended technique was to effectively enhance an image recognition system so that it could determine if a picture was real or fake. The SVM classifier is utilized for object-based picture recognition. The proposed classifier uses various image recognition factors that demonstrate the viability of the provided approaches. Sensitivity, specificity, accuracy, and precision are among the various variables that can be utilized for recognition purposes.

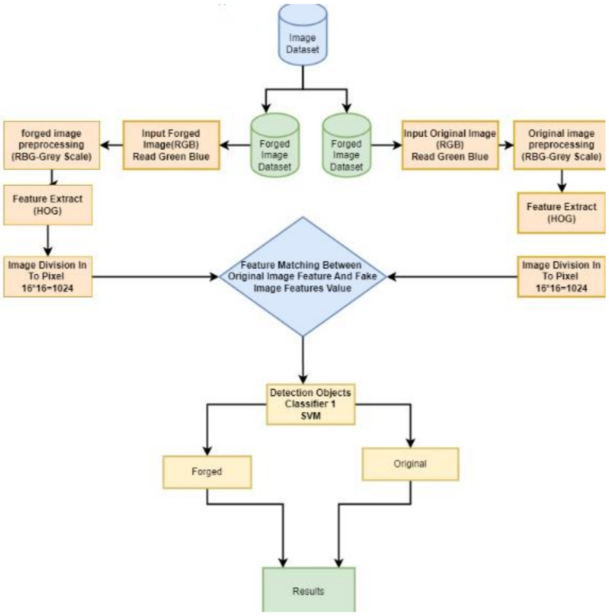


Image Dataset
forged image preprocessing (RBG-Grey Scale)
Input Forged Image(RGB) Read Green Blue
Forged Image Dataset
Forged Image Dataset
Input Original Image (RGB) Read Green Blue
Original image preprocessing (RBG-Grey Scale)
Feature Extract (HOG)
Feature Extract (HOG)
Image Division In To Pixel 16*16=1024
Feature Matching Between Original Image Feature And Fake Image Features Value
Image Division In To Pixel 16*16=1024
Detection Objects Classifier 1 SVM
Forged
Original
Results

Fig. 3. The complete design of the proposed method.

D. Evaluation Metrics

These settings are thought of as conventional for picture recognition. Many researchers employed this parameter for pixel-based picture recognition [19]. The following formulas are used to calculate evaluation metrics.

$$\text{Confusion matrix} = \frac{TP + TN}{TP + TN + FP + FN} * 100$$

Most researchers have utilized this formula to validate picture recognition in their studies. An overall valuation is mostly used to assess the efficacy and verification rates of the suggested approach. In the confusion matrix (FN), four terms are employed: true negative (TN), true positive (TP), false positive (FP), and false negative (FN). In essence, the confusion matrix is a table with rows and columns using a row, the training ratio is displayed. The number of photos used as a dataset is determined using the training ratio. The TP, TN, FP, FN, accuracy, precision, specificity, scores, and sensitivity are represented by columns [20].

$$\text{Verification rate} = \frac{\text{Total number that verify } * 100}{\text{Total number of images}}$$

A genuinely positive image that is stored in a database is divided by all positive photos, including false positives, to determine precision. The accuracy is detailed below.

$$\text{Precision} = \frac{\text{True Positive}}{\text{Total Positive } + \text{False Positive}}$$

To determine the specificity, the number of real positive images in a database is divided by the total of all negativeand false-positive photos [21]. A genuinely positive image stored in a database is divided by all positive photos, including false positives and negatives, to determine sensitivity. The sensitivity described below.

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{Total Positive } + \text{False Negative}}$$

A database's existing false negative image is divided by all positives plus false negative photos to determine the false rejection rate. The FAR is explained below.

$$\text{FRR} = \frac{\text{False Negative}}{\text{Total Positive } + \text{False Negative}}$$

FAR is calculated by dividing a false-positive image from a database by the sum of false-positive and true-negative images [22]. Below is a description of the false mistake rate

$$\text{FAR} = \frac{\text{False Positive}}{\text{False Positive } + \text{True Negative}}$$

AER is calculated by adding the false error rate and the non-match rate. Below is an AER description [23].

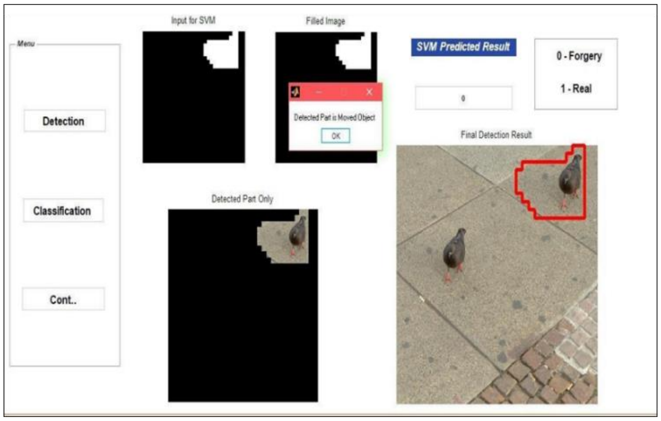
$$\text{AER} = \frac{\text{FAR} + \text{FRR}}{2}$$

IV. Result and Discussion

The technique required the installation of a MATLAB software application, downloaded from their website, which included a graphic user interface developed by Miao et al. for image examination and verification. The GUI interface was constructed using MATLAB and was used for verifying and analyzing feature extraction results, as well as assessing the performance of two algorithms. Feature extraction was performed by selecting the HOG button in the GUI windows, using a library with train weights that serve as changeable weights for component reversal [24].

To test the proposed method, a dataset of 69 photographs, including 24 authentic and 16 beautifully faked photos, was used to obtain results based on each pattern's intensity and similarity. Many earlier scientists had already utilized these datasets in a proposed manner for high accuracy. Block comparison was used in this research to detect matching blocks and to suspect fabricated sections. According to the suggested scheme, matching blocks were found by calculating the Vectors of features' Euclidean distances. To correctly detect fabricated regions, the distance threshold Td and similarity criteria must be predetermined.

The predict button is used to predict the image as forged or genuine as shown in Fig. 4.



Detected Part is Moved Object

0

Fig. 4. Detect object using the feature matching.

The image is fragmented into 16X16 numbers of blocks. The total value of one picture is 1024. The picture is divided into blocks of overlapped squares to help find forged sections. To calculate the HOG descriptors, the grayscale image I of M N is first split into overlapping sub-blocks of L. Then, overlapping (M L + 1) (N L + 1) pieces of the image are created. GUI is used to access the project. The purpose of pressing the train button is to facilitate machine training.

The image is fragmented into 16X16 numbers of blocks. The total value of one picture is 1024. The picture is divided into blocks of overlapped squares to help find forged sections. To calculate the HOG descriptors, the grayscale image I of M N is first split into overlapping sub-blocks of L. Then, overlapping (M L + 1) (N L + 1) pieces of the image are created. GUI is used to access the project. The purpose of pressing the train button is to facilitate machine training. Table IV shows the HOG feature of each block.

Table V shows the original image pixel values of each block. Table VI shows the HOG Feature of Each Block (Original Image).

TABLE IV. HOG Feature of each Block (Original Image) (1)

0.0399	0.02039	0.026578	0.025309	0.089255	0.18337	0.28502
0.13172	0.053051	0.31601	0.04851	0.013364	0.026087	0.028498
0.060453	0.17684	0.075901	0.31601	0.032584	0.028347	0.031878
0.13566	0.31601	0.17209	0.11828	0.072399	0.035738	0.27992
0.016298	0.001963	0.10184	0.31601	0.14993	0.30773	0.17204
0.31601						

TABLE V. THE PIXEL VALUE OF EACH BLOCK (ORIGINAL IMAGE) 16X16=1024

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	104	77	82	85	70	96	128	133	137	152	132	132	118	103	119	85
2	78	76	127	8	125	135	138	134	117	120	128	124	114	75	74	56
3	73	113	104	118	128	115	113	118	139	119	69	70	63	88	130	112
4	112	122	119	69	70	63	88	130	112	122	119	109	118	113	107	119
5	114	123	125	89	66	102	119	126	123	108	121	133	117	105	111	107
6	112	126	77	75	75	114	121	128	122	123	110	123	108	108	113	129
7	131	63	78	95	116	133	127	135	135	127	132	135	120	118	125	114
8	116	67	100	122	117	124	118	127	137	139	143	134	129	120	131	120
9	119	72	119	135	110	112	126	115	123	135	138	123	114	126	118	116
10	115	92	131	124	117	119	125	138	138	129	139	128	118	128	113	103
11	112	117	124	120	129	120	129	137	138	122	123	131	127	123	118	128
12	113	116	67	100	122	117	124	118	127	137	139	143	134	129	120	131
13	123	119	72	119	135	110	112	126	115	123	135	138	123	114	126	118
14	114	123	125	89	66	102	119	126	123	108	121	133	117	105	111	114
15	112	126	77	75	75	114	121	128	122	123	110	123	108	108	113	112
16	131	63	78	95	116	133	127	135	135	127	132	135	120	118	125	131
6	31	3	8	5	16	33	27	35	35	27	32	35	20	18	25	31

TABLE VI. HOG FEATURE OF EACH BLOCK (ORIGINAL IMAGE) (2)

0.03909	0.02039	0.026578	0.025309	0.089255	0.18337	0.28502
0.13172	0.053051	0.31601	0.04851	0.013364	0.026087	0.028498
0.060453	0.17684	0.075901	0.31601	0.032584	0.028347	0.031878
0.13566	0.31601	0.17209	0.11828	0.072399	0.035738	0.27992
0.016298	0.001963	0.10184	0.31601	0.14993	0.30773	0.17204
0.31601						

TABLE VII. THE PIXEL VALUE OF EACH BLOCK (FAKE IMAGE) 16 X16=1024 PIXEL

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1	104	77	82	85	70	96	128	133	137	152	132	132	118	103	119	85
2	78	76	127	8	125	135	138	134	117	120	128	124	114	75	74	56
3	73	113	104	118	128	115	113	118	139	119	69	70	63	88	130	112
4	112	122	119	69	70	63	88	130	112	122	119	109	118	113	107	119
5	114	123	125	89	66	102	119	126	123	108	121	133	117	105	111	107
6	112	126	77	75	75	114	121	128	122	123	110	123	108	108	113	129
7	131	63	78	95	116	133	127	135	135	127	133	135	120	118	125	114
8	116	67	100	122	117	124	118	127	137	139	143	134	123	120	131	120
9	119	72	119	135	114	112	126	115	123	135	138	123	114	126	118	116
10	115	92	131	124	117	119	125	138	135	129	139	128	118	128	113	103
11	112	117	123	120	129	120	129	137	138	122	123	131	127	123	118	128
12	113	116	67	100	122	117	124	118	127	137	139	143	134	129	120	131
13	123	119	72	119	135	110	112	126	115	123	135	138	123	114	126	118
14	114	123	124	89	66	102	119	126	123	108	121	133	117	105	111	114
15	112	126	77	73	75	114	121	128	122	123	110	123	108	108	113	112
16	131	63	78	95	116	133	127	135	135	127	132	135	120	118	125	131

Table VII shows the fake values of each block of the images. Each connected block is represented by a HOG descriptive matrix that is the same length as the block after HOG has been applied to each block. The local histogram is considered in the following with four bits. Each histogram bin corresponds to a 45-degree orientation interval because of the histogram's uniformly spaced channels between 0 and 180. Table VIII and Table IX respectively shows Hog feature and feature matching between original and fake images.

TABLE VIII. HOG FEATURE OF EACH BLOCK (FAKE IMAGE)

0.03909	0.02039	0.026578	0.025309	0.089255	.18337	0.28502	0.13172
0.053051	0.31601	0.04851	0.013364	0.026087	.028498	0.060453	0.17684
0.075901	0.31601	0.032584	0.028347	0.031878	.13566	0.31601	0.17209
0.11828	0.07239	0.035738	0.27992	0.016298	.0019632	0.10184	0.31601
0.14993	0.30773	0.17204	0.31601				

TABLE IX. FEATURE MATCHING BETWEEN (ORIGINAL AND FAKE IMAGE)

-0.003029 0.1072 0.060453 0.019606 0.086935 0.025834 0.0021474	0.008449 0.023697 0.050846 0.05198 0.031916 0.004362 -0.0070	- .016158 -0.17684 -0.059107 0.021136 0.030952 0.019225	0.02865 0.21214 0.10184 -0.071829 0.21693 -0.09204	0.084415 -0.12757 0.0061755 -0.0594 -0.030912	0.012647 -014036 -0.043751 0.050846 -0.022

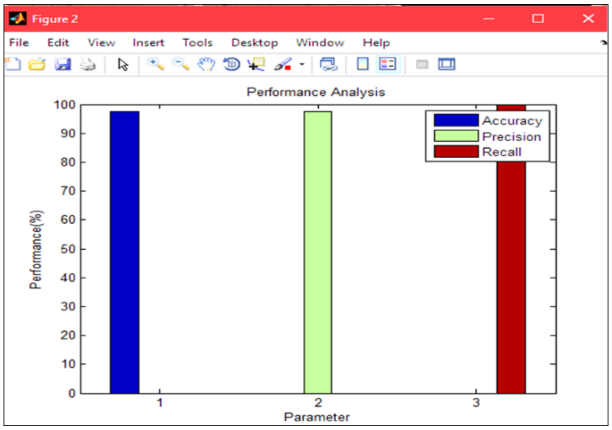


Fig. 5. Final accuracy of suggested algorithm.

A separate collection of entirely separate images is created for testing purposes, and the features matrix is extracted from these. A supervised machine learning technique called SVM is frequently employed for classification issues. This technique uses the value of a certain set of coordinates as a feature value and plots it in relation to a position in n-dimensional reality. A decision boundary (hyperplane) that isolates the two-class datasets from one another as much as feasible is created by the SVM classifier. SVM classifier will more accurately determine if the image is real or fake. SVM classifier detects the object that is used for image recognition whether the image is fake or not. If the image is a forgery. If the image is authentic or original, it displays message 1, and else it displays message 0. Fig. 4 and show a test image that is fake identifying the testing sample's forgery or duplicate move regions. Fig. 5 shows the accuracy that deals with how closely the calculated values are to the true values, and it must be high.

The suggested method is having great similarity levels to distinguish the copy-move sections in the testing datasets. The accuracy of the proposed algorithm is 98%. Table X shows comparison between different algorithms with the proposed method.

TABLE X. COMPARISON BETWEEN DIFFERENT ALGORITHMS WITH PROPOSED METHOD

Sr.no	**Datasets**	**Dimension**	**Recognition parameters**	**Accuracy**	**Researchers**
1	Local datasets	512 x234	Image texture, light strength Matching Points	Accuracy 94 %.	Umamahes wari,D.&Karthikeyan2022
2	The input dataset has been downloaded from the website	412 X314	Variance, mean, skewness, energy, etc.	Accuracy rate of 92.22%	Rathore, Neeraj Kumar,et al., 2021
3	MICC-220 dataset253 Images	722 X 480	Number of Clusters Maximum no of Iteration	Accuracy standards and minor enhancement in some cases.	Alberry, Hesham Abdelfattah A. Hegazy, and Gouda I. Salama.2018
4	CASIA 2.0 dataset 5123 images	452X 434	Dimensions	80.91% Accuracy	Zhang, Zhongping, et al., 2018
5	PASCAL VOCMICC-F220 72 Images	560 X 450	Similarity Translation, rotation, noise, illumination and JPEG compression.	83.33 % Accuracy	Tian, Xiuxia, Guoshuai Zhou, and Man Xu2020
6	MICCF8 multi,MICC- F220 benchmark dataset	160X340	lock-based methods Edge Images	80%	William, Y., Safwat, S., &Salem, M. A. M. (2019, September)
7	CASIAv1.0. Datasets	412X340	Block of the images	86.62 %	Kanwal, Navdeep, et al.2019
8	MICC-F2000MICC-F220	415X 412	Matching Refinement Objects	94.45%	Elaskily, M. A.,Elnemr,H.A.,Dessouky,M.M.,& Faragallah,O. S. (2019).
9	CoMoFoD dataset CMHD	412 X412	Matching	91%	Yang, J., Liang, Z., Gan, Y.,& Zhong, J.(2021).
10	MICC-220 dataset 253 Images, MICC-F2000 MICC- F220,	512x512 412X313 725 x735	MatchingPixels Dimensions	98%	**Proposed** Method

V. Conclusion and Future Work

Using the Histogram of Oriented Gradients (HOG) and Support Vector Machine (SVM) algorithms, we provide a unique method to improve picture verification. The findings demonstrate that the suggested strategy distinguishes between real and fake photos with a remarkable accuracy rate of 98%. The usefulness and promise of the HOG-SVM combo for image verification tasks are shown by the accuracy, which exceeds numerous other comparable techniques. The research underlines the value of picture authentication systems in several fields and draws attention to the shortcomings of current approaches for identifying intricately faked images. The suggested method overcomes these difficulties and provides a significant boost in recognition accuracy, making it an important addition to the area of image verification. It does this by using HOG and SVM.

Overall, the study offers insightful information on the application of feature extraction methods and supervised machine learning algorithms to picture recognition. The suggested method's high accuracy raises the confidence and dependability of picture verification systems, possibly resulting in greater security and credibility in various image authentication-related applications.

A. Future Work

The current system indicates the importance of image verification. This research shows an accuracy of 98% but more amount of research and present methods can be added by using other image datasets and Implementations. Develop a working model and Record observations based on the dataset. More clear images must be used in the dataset for extracting the image features.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Funding

This work was funded by Project ECS 0000024 Rome Technopole, CUP B83C22002820006, NRP Mission 4 Component 2 Investment 1.5, Funded by the European Union Next Generation EU.

References

[1] R. Thakur and R. J. F. s. i. Rohilla, "Recent advances in digital image manipulation detection techniques: A brief review," vol. 312, p. 110311, 2020.

[2] M. M. Islam, G. Karmakar, J. Kamruzzaman, M. Murshed, G. Kahandawa, and N. Parvin, "Detecting splicing and copy-move attacks in color images," in 2018 Digital Image Computing: Techniques and Applications (DICTA), 2018, pp. 1-7: IEEE.

[3] S. Tyagi and D. J. T. V. C. Yadav, "A detailed analysis of image and video forgery detection techniques," vol. 39, no. 3, pp. 813-833, 2023.

[4] R. Gupta, P. Singh, T. Alam, S. J. M. T. Agarwal, and Applications, "A deep neural network with hybrid spotted hyena optimizer and grasshopper optimization algorithm for copy move forgery detection," vol. 82, no. 16, pp. 24547-24572, 2023.

[5] K. J. I. J. o. A. C. S. Arai and Applications, "Image restoration based on maximum entropy method with parameter estimation by means of annealing method," vol. 11, no. 8, 2020.

[6] S. Walia and K. J. A. J. o. F. S. Kumar, "Digital image forgery detection: a systematic scrutiny," vol. 51, no. 5, pp. 488-526, 2019.

[7] W. D. Ferreira, C. B. Ferreira, G. da Cruz Júnior, F. J. C. Soares, and E. Engineering, "A review of digital image forensics," vol. 85, p. 106685, 2020.

[8] K. J. I. J. o. A. C. S. Arai and Applications, "Wavelet multi resolution analysis based data hiding with scanned secrete images," vol. 13, no. 9, 2022.

[9] A. J. I. S. R. N. Piva, "An overview on image forensics," vol. 2013, no. 1, p. 496701, 2013.

[10] L. J. I. J. o. S. T. i. S. P. Verdoliva, "Media forensics and deepfakes: an overview," vol. 14, no. 5, pp. 910-932, 2020.

[11] C. D. M. Henderson, "Large Scale Pattern Detection in Videos and Images from the Wild," Queen Mary University of London, 2017.

[12] F. P. W. Lo, Y. Sun, J. Qiu, B. J. I. j. o. b. Lo, and h. informatics, "Image-based food classification and volume estimation for dietary assessment: A review," vol. 24, no. 7, pp. 1926-1939, 2020.

[13] S. Gupta, N. Mohan, and P. J. A. I. R. Kaushal, "Passive image forensics using universal techniques: a review," vol. 55, no. 3, pp. 1629-1679, 2022.

[14] Y. Malhotra, "Image forgery detection using textural features and deep learning," 2021.

[15] P. Capasso, G. Cattaneo, M. J. A. T. o. M. C. De Marsico, Communications, and Applications, "A Comprehensive Survey on Methods for Image Integrity," 2023.

[16] M. Alirezaei, S. T. A. Niaki, and S. A. A. J. E. S. w. A. Niaki, "A bi-objective hybrid optimization algorithm to reduce noise and data dimension in diabetes diagnosis using support vector machines," vol. 127, pp. 47-57, 2019.

[17] S. Sadeghi, S. Dadkhah, H. A. Jalab, G. Mazzola, D. J. P. A. Uliyan, and Applications, "State of the art in passive digital image forgery detection: copy-move image forgery," vol. 21, pp. 291-306, 2018.

[18] A. K. Rai, S. J. C.-C. M. i. E. Srivastava, and Sciences, "A Thorough Investigation on Image Forgery Detection," vol. 134, no. 3, 2023.

[19] N. K. Rathore, N. K. Jain, P. K. Shukla, U. Rawat, and R. J. N. A. S. L. Dubey, "Image forgery detection using singular value decomposition with some attacks," vol. 44, no. 4, pp. 331-338, 2021.

[20] D. Banumathy, O. I. Khalaf, C. A. T. Romero, P. V. Raja, and D. K. J. C. S. S. E. Sharma, "Breast Calcifications and Histopathological Analysis on Tumour Detection by CNN," vol. 44, no. 1, pp. 595-612, 2023.

[21] J. Sujin and S. J. S. C. Sophia, "High-performance image forgery detection via adaptive SIFT feature extraction for low-contrast or small or smooth copy–move region images," vol. 28, no. 1, pp. 437-445, 2024.

[22] N. Kanwal, A. Girdhar, L. Kaur, and J. S. Bhullar, "Detection of digital image forgery using fast fourier transform and local features," in 2019 international conference on automation, computational and technology management (ICACTM), 2019, pp. 262-267: IEEE.

[23] Y. Kortli, M. Jridi, A. Al Falou, and M. J. S. Atri, "Face recognition systems: A survey," vol. 20, no. 2, p. 342, 2020.

[24] M. J. Zedan, M. A. Zulkifley, A. A. Ibrahim, A. M. Moubark, N. A. M. Kamari, and S. R. J. D. Abdani, "Automated glaucoma screening and diagnosis based on retinal fundus images using deep learning approaches: A comprehensive review," vol. 13, no. 13, p. 2180, 2023.