Digital Image Forgery Detection using Machine Learning

Dr. G. S. Bapi Raju*Department of Computer Science and EngineeringGokarajuRangaraju Institute of Engineering & Technology. Telangana, India*gsbapiraju@gmail.com

Praveen Palla*, Department of Computer Science and EngineeringGokarajuRangaraju Institute of Engineering & Technology. Telangana, India*palla.praveen2002@gmail.com

Aryan Sudhagon , Department *of Computer Science and EngineeringGokarajuRangaraju Institute of Engineering & Technology. Telangana, India*aryansudhagoni@gmail.com

Umasai Pothanaboina*, Department of Computer Science and EngineeringGokarajuRangaraju Institute of Engineering & Technology. Telangana, India*Pothanaumasai@gmail.com

Tarunkumar Voggu *Department of Computer Science and EngineeringGokarajuRangaraju Institute of Engineering & Technology. Telangana, India*tarunkumarvoggu9442@gmail.com

Abstract—
A digital picture can be forged by concealing part of the image's significant or crucial data. Identifying the portion of the original image that was altered is typically challenging. The detection of forgery in the image is necessary to maintain the integrity and authenticity of the image. Exploiting digital photos is now simple with the aid of image editing software due to advances in photographic technology and acclimatization to modern living. Therefore, detecting such image forgery operations in the images is crucial. Object addition, object removal, and unexpected size changes in the picture can all be used to detect image forgeries. One of the most effective forms of media for communication is the image. In this project, we have used algorithms such as Support Vector Machine (SVM), Copy Move Technique (CMT), Decision Tree (DT), Convolution Neural Network (CNN), and Random Forest (RF). All are measured and compared in terms of accuracy, proving that SVM performs better than other algorithms.

Keywords-- Digital image, Forgery detection, Machine learning, Support Vector Machines, Detection technique
DOINumber:10.48047/nq.2023.21.5.NQ222048 NeuroQuantology2023;21(5):532-538

I. INTRODUCTION

Unauthorized manipulation of photos or documents is known as a forgery. Images are altered for a variety of purposes, including to provide false evidence or to illegally make money. A picture of an image communicates a concept considerably more effectively than human speech does.The authenticity of digital images is now threatened by the use of applications such as Adobe Photoshop, GIMP, and Corel Paint Shop to manipulate photos due to new advanced technologies.

Forgeries of digital images are becoming more common in criminal cases and public courses. There are currently no well-established methods for automatically determining the authenticity and integrity of digital photographs. The

software created helps us detect the region of modification done in an image provided.

II. LITERATURE SURVEY

- **Yong-In Yun et al:**The image interpolation approach from digital photos is used in this work to provide a unique strategy for identifying digital forgeries. The interpolated picture between the photos is always present in digital forgeries made from digital photographs. Their method involved applying the expectation-maximization (EM) technique to calculate the interpolated coefficient for the pictures. They showed a ratio of recognizing the fake photographs and illustrated how to achieve it using each filter for the Adobe Photoshop tool.
- **H.M. Shahriar Parvez:** An effective region-duplication forgery detection method is suggested in this work. The segment-based region duplication forgery detection techniques are a subset of this study. The approach is built using Gabor descriptors and K-Means clustering and is based on picture segmentation. The normalized cut (NCut) segmentation technique is initially used to segment the picture. Then, using the K-means clustering approach, applied Gabor Filters were used to extract picture information. The legitimacy of the image will next be determined by comparing the clustering areas with the specified threshold value.
- **Saiqa Khan and Arun Kulkarni:**This study offers a copy-move forgery detection method using blind image forensics. Using DWT, the dimension of the fake picture is decreased in this method (Discrete Wavelet Transform). Then, overlapping chunks of a predetermined size are created from the compressed picture. Lexicographical sorting is used to arrange these blocks, and Phase Correlation is used as similarity criteria to spot repeated blocks.
- **Baina Su and Zhu Kaizhen:** This study presents a new approach for detecting picture forgeries based on LPP-SIFT (Locality Preserving Projection- Scale Invariant Feature Transform) characteristics. The approach begins by extracting SIFT key points from an image, combining LPP to provide low-dimensional feature descriptors, and then performing key point matching as the last step. Lines are drawn between each pair of the image's matching key points. If there has been a copy-forging procedure on the picture, these lines will focus on two specific areas.
- **Gunjan Bhartiya and Anand Singh Jalal:**This study presents a technique to identify forgeries in JPEG images. Based on this technique, an algorithm is developed to categorize picture blocks as forgeries or not. The method is more effective in detecting forgeries than earlier methods that used a probability-based approach.

III.METHODOLOGY

This section provides extensive information on the data set and the various construction techniques used in building the project.

A. Existing System

Copy Move Technique (CMT):

A section of an image is duplicated and transferred to another region of the same picture in a photo spoofing technique called copy-move forgery. Copy move forgery is used to hide certain details or replicate objects inside a photograph. Copy-Move forgery's main goal is to conceal certain information from the original picture. Since the duplicated region is a part of the original image, its properties—including its color scheme, noise level, dynamic range, and other characteristics—will be consistent with those of the remainder of the image.

Random Forest (RF):

Random forests, also known as random decision forests, are an ensemble learning technique for classification, regression, and other tasks. During the training phase, a large number of decision trees are built, and the output class represents the mean of the classes (for classification) or the mean/average prediction (for regression) of the individual trees. Random decision forests correct the tendency of decision trees to overfit their training set. Random forests often outperform decision trees, but they are less accurate than gradient-boosted trees [1]-[2]. However, data abnormalities may reduce their efficiency [3]-[4].

B. Proposed System

Support Vector Machine (SVM):

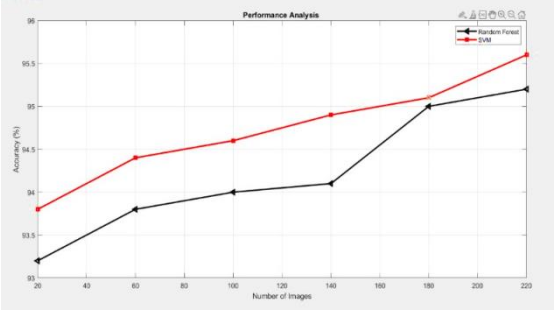
Identifying an optimal hyperplane in a high-dimensional space for several distinct situations is the SVM's fundamental process. Several hyperplanes can implement this model. This strategy relies on the information that is closest to the closed surface and has coordinated with the optimal choice surface—the bolster vector. By creating a hyperplane to partition the data and planning the input vectors into a high-dimensional space, it performs classification. Most commonly, this approach is used to address nonconvex, unrestricted minimization, and quadratic programming issues. The classification process's most effective method is the SVM [5]-[6].

C. Problem statement

Due to the availability of low-cost, high-resolution digital cameras in recent years, there are a lot of digital photos available worldwide. Any non-expert can change the photograph with the use of extremely simple photo editing software [7]-[8]. If an image alteration modifies the original image's semantics, it becomes a counterfeit. There are a variety of reasons why a forger could commit a forgery, including hiding something in an image, emphasizing certain elements, covering over items in an image to generate false proof, improving the image's aesthetics, etc [9]-[10]. Digital image fraud may be categorized in a variety of ways, but the primary ones include enhancing, retouching, splicing, morphing, and Copying/Move [11]-[12]. The following are the problems in the existing system:

1. Less accuracy rate in classifying the Forgery areas in the input image [13]-[14].
2. More time consumption [15] .
3. Cannot be implemented in all datasets.
4. More noise ratio[16] .

D. Comparative study of Existing and Proposed Systems

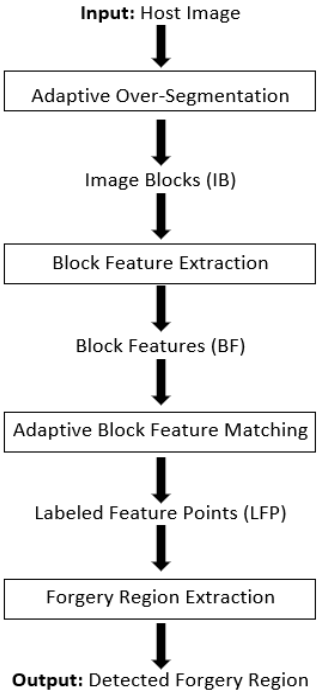


In this project, we have used algorithms such as Support Vector Machine (SVM), Copy Move Technique (CMT), Decision Tree (DT), Convolution Neural Network (CNN), and Random Forest (RF)[17] . All are measured and compared in terms of accuracy, proving that SVM performs better than other algorithms. The below figure is the accuracy analysis[18].

E.Algorithm of Proposed Work

- Blocks of an image are separated into tiny, overlapping, or non-overlapping segments.
- Use conventional methods to extract the features [19] .
- A matrix is used to hold the values of the extracted features that correspond to each key point[20].
- Use sorting methods to identify features that are similar and nearer to.
- To identify the key point with similar shifting introduce a shift vector [20].
- Set the counter to 1 and use the counter vector to count the recurrence of the same shifting key point.
- Using the threshold value and the previous methods, similar areas are found.

F. Framework of Proposed System



IV. SYSTEM IMPLEMENTATION

A. Image Acquisition

Image acquisition is the procedure used to gather images. These images are downloaded from the online dataset provider called Kaggle.com.

B. Image pre-processing

Grayscale image conversion is an aspect of image pre-processing. An RGB picture preserves the original colors of the photos. Images that are grayscale combine black and white. To improve the available dataset, RGB is converted to grayscale. The precision of the output is increased by converting the photos to grayscale. Grayscale images neutralize the backdrop and assist to decrease noise [21].

C. Image Segmentation

By using image segmentation, the image is divided into useful parts. It separates the digital picture into many section [22]. The objective is to make the representation clearer or transform it into a picture with greater depth [23].

D. Feature Extraction

The divided area of the image is extracted or shown as a feature, making classification simpler. To distinguish between the photos, features are extracted. Nearly all machine vision techniques rely on feature extraction.

E. Matching

To find the duplicated sections, matching is used. A copy region is indicated by a high degree of similarity between two attribute descriptors. Lexical sorting and Best-Bin-First search are two techniques for matching.

F. Filtering

We can lower the likelihood of erroneous matching images by using a filter. Common noise suppression is removed along with the matches between geographically adjacent areas. It appears erroneous forgery detection since most neighboring pixels often have similar brightness.

G. Post-processing

The key objective of the post-processing stage is to save matches that exhibit a typical performance. Let's look at a group of matches that belongs to the duplicated area. Such matches are expected to be played close to one another in both the starting block and the final block (or key points).

H. Key point Based Method

Techniques based on key points operate with the full image. Block-based approaches and point-based methods, on the other hand, only compute their properties on image regions with the extreme disorder.

I. Classification

Here, we apply the classification method's idea. Tensor Flow and machine learning methods will be employed for classification in the final module. Machine learning is made quicker and simpler with TensorFlow, an open-source library for numerical computation that is compatible with Matlab.

V. SODTWARE DEVELOPMENT

A. Introduction

We used MATLAB software to write and use the algorithm that we used in this project. Manipulating Matrices is a good place to start if you are new to MATLAB. Learning how to input matrices, how to utilize the: (colon) operator, and how to call functions are the three most crucial skills.

B. Components of the MATLAB

- Development Environment - explains the MATLAB development environment, describing the MATLAB desktop and its tools.
- Manipulating Matrices - explains how to create matrices using MATLAB and how to manipulate them mathematically.
- Graphics - presents the graphic features of MATLAB, including details on data charting, graph annotation, and dealing with pictures.
- Programming with MATLAB - explains how to manage data structures including cell arrays and multidimensional arrays by writing scripts and functions in the MATLAB programming language.

For technical computing, MATLAB is a high-performance language. It combines calculation, visualization, and programming with an easy-to-use interface, and it uses well-known mathematical language to represent problems and answers.

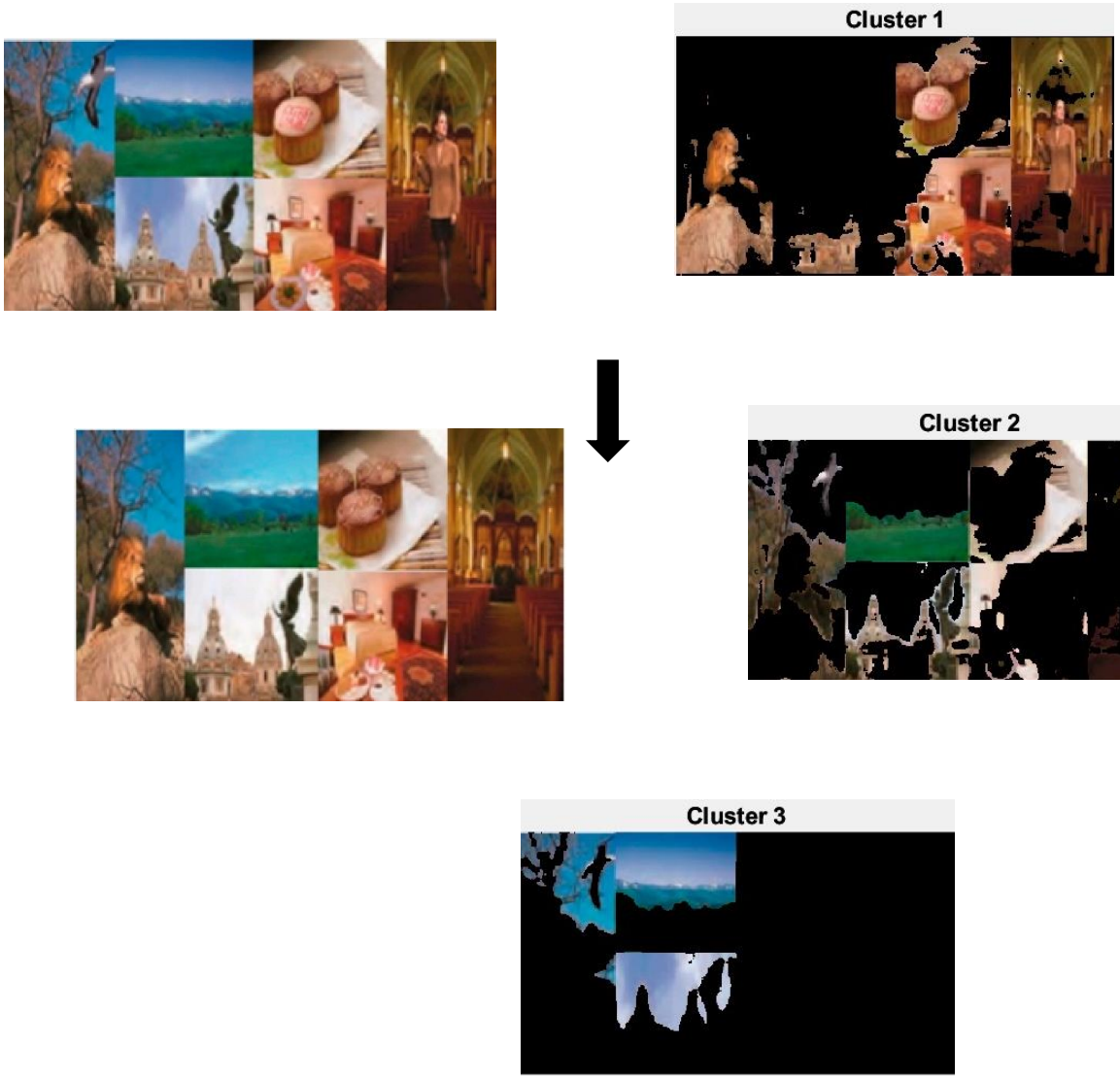
C. Typical uses of MATLAB

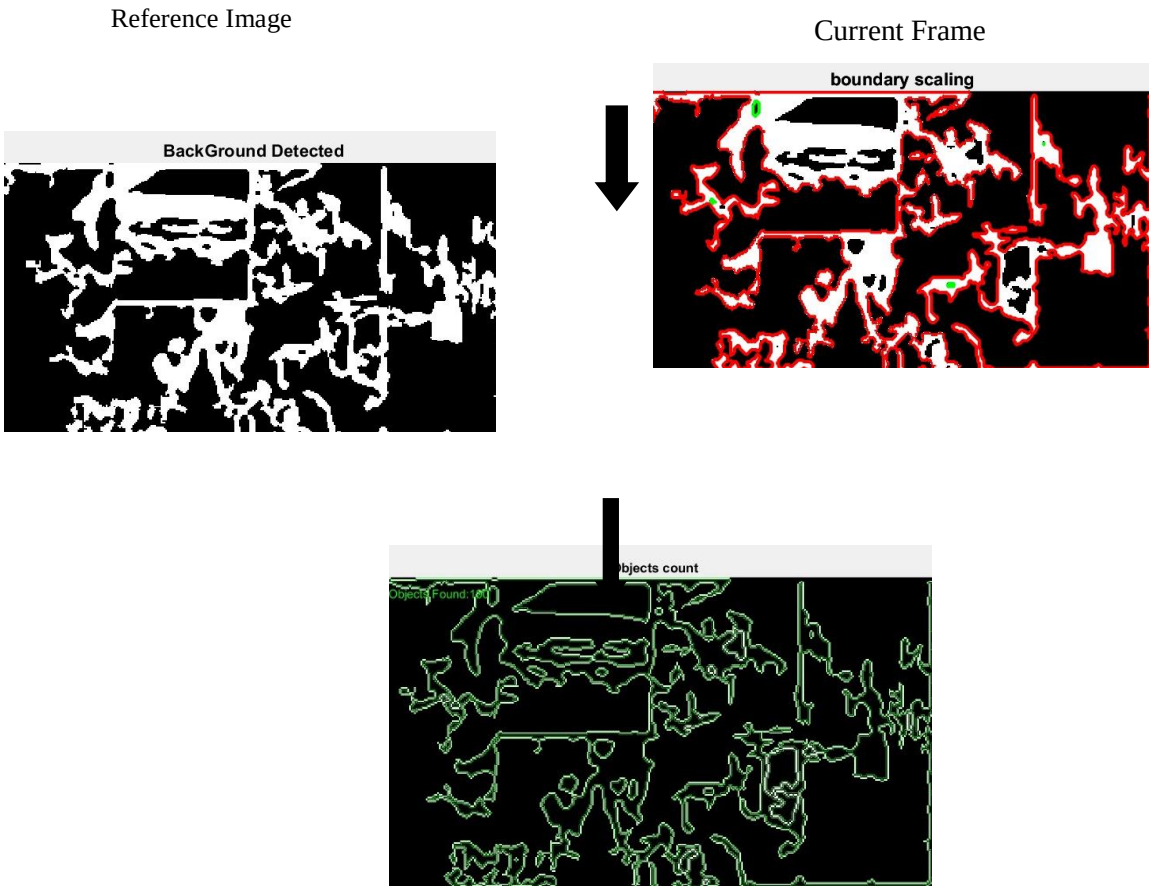
- Development of algorithms.
- Computation and mathematics.
- Prototyping, simulation, and modeling.
- Exploration, analysis, and visualization of data.
- Graphics used in science and engineering.
- The creation of GUIs for applications.

VI. APPLICATIONS

- Forensics: detecting and investigating tampering in criminal cases and legal proceedings.
- Banking: detecting tampering in the signatures made.
- Media and journalism: verifying the authenticity of photos, videos, and other multimedia.
- E-commerce: preventing the sale of fake or counterfeit products by detecting manipulated images.
- Art authentication: detecting alterations invaluable works of art.
- Election fraud: detecting manipulation of election results through forged images and videos.

VII.OUTPUT





VIII. CONCLUTION AND FUTURE SCOPE

A. Conclusion

The proposed method for detecting image forgery makes use of morphological operation and feature point extraction. By indicating the affected pixel, it may split the forged region. The algorithm implemented in the described experiment may perform well under a variety of difficult circumstances, including geometrical transform and JPEG compression. Therefore, without using any pre-existing data sets for the forged picture, the system can identify forgery with accuracy and efficiency.

B. Future Scope

Future research may improve the suggested algorithm's forgery detection accuracy in photos and videos. Another potential route for the suggested system is to employ overlapping blocks of different sizes for the morphological procedures. The resilience and time required to identify the forgery can both be improved by using the variable size of the block for detection. This method is now only used for forensics, but it may one day be used to filter out dangerous and fake information on social media.

IX. REFERENCES

[1] A.C. Popscu, and H. Farid. “Statistical Tools for Digital Forensics”. Toronto, Canada 2019.
[2] Shivani Thakur, RamanpreetKaur, Dr. Raman
[3] Chadha, JasmeetKaur. “A Review Paper on Image Forgery Detection in Image Processing”. IOSR Journal of Computer Engineering. Volume 18, Issue 4, Ver. I (Jul.-Aug. 2019), Pg 86-89.
[4] M. Qiao, A. Sung, Q. Liu, and B. Ribeiro. “A novel approach for detection of copy-move forgery”. ADVCOMP, 2019.
[5] Gajanan K. Birajdar, Vijay H. Mankar. “Digital image forgery detection using passive techniques: A survey”. Digital Investigation, volume 10, 2019, Pg 226-245.
[6] GagandeepKaur, Manoj Kumar. “Study of Various Copy Move Forgery Attack Detection in Digital Images”. International Journal of Research in Computer Applications and Robotics, Vol.3, Pg 30-34 September 2019.
[7] Devanshi Chauhan, Dipali Kasat, Sanjeev Jain, VilasThakare. “Survey on KeyPoint-based Copymove Forgery Detection Methods on Images”. Volume 85, 2019.

[8] Ali Qureshi, M., and M. Deriche. "A review on copy move image forgery detection techniques". IEEE, 2019.

[9] Qazi, Tanzeela. "Survey on blind image forgery detection". IET, 2019.

[10] Rohini.A.Maind, AlkaKhade, D.K.Chitre. "Image Copy Move Forgery Detection Using Block Representing Method". International Journal of Soft Computing and Engineering, Volume-4, May 2019.

[11] R.C. Gonzalez, R.E. Woods. "Digital Image Processing". Addison-Wesley, 2019.

[12] L. Kang, X.-P. Cheng. "Copy-Move Forgery Detection in Digital Image". 3rd International Congress on Image and Signal Processing, IEEE Computer Society, 2019, Pg 2419-21.

[13] Kailasam, S., Achanta, S. D. M., Rao, P. R. K., Vatambeti, R., & Kayam, S. (2021). An IoT-based agriculture maintenance using pervasive computing with machine learning technique. International Journal of Intelligent Computing and Cybernetics.

[14] Koppula, N., Sarada, K., Patel, I., Aamani, R., & Saikumar, K. (2021). Identification and Recognition of Speaker Voice Using a Neural Network-Based Algorithm: Deep Learning. In Handbook of Research on Innovations and Applications of AI, IoT, and Cognitive Technologies (pp. 278-289). IGI Global.

[15] Rao, K. S., Reddy, B. V., Sarada, K., & Saikumar, K. (2021). A Sequential Data Mining Technique for Identification of Fault Zone Using FACTS-Based Transmission. In Handbook of Research on Innovations and Applications of AI, IoT, and Cognitive Technologies (pp. 408-419). IGI Global.

[16] Raju, K., Pilli, S. K., Kumar, G. S. S., Saikumar, K., & Jagan, B. O. L. (2019). Implementation of natural random forest machine learning methods on multi spectral image compression. Journal of Critical Reviews, 6(5), 265-273.

[17] Garigipati, R. K., Raghu, K., & Saikumar, K. (2022). Detection and Identification of Employee Attrition Using a Machine Learning Algorithm. In Handbook of Research on Technologies and Systems for E-Collaboration During Global Crises (pp. 120-131). IGI Global.

[18] Mythreya, S., Murthy, A. S. D., Saikumar, K., & Rajesh, V. (2022). Prediction and Prevention of Malicious URL Using ML and LR Techniques for Network Security: Machine Learning. In Handbook of Research on Technologies and Systems for E-Collaboration During Global Crises (pp. 302-315). IGI Global.

[19] Saikumar, K., Rajesh, V., Babu, B.S. (2022). Heart disease detection based on feature fusion technique with augmented classification using deep learning technology. Traitement du Signal, Vol. 39, No. 1, pp. 31-42. https://doi.org/10.18280/ts.390104

[20] Kailasam, S., Achanta, S.D.M., Rama Koteswara Rao, P., Vatambeti, R., Kayam, S. (2022). An IoT-based agriculture maintenance using pervasive computing with machine learning technique. International Journal of Intelligent Computing and Cybernetics, 15(2), pp. 184–197.

[21] Saikumar, K., Rajesh, V. A machine intelligence technique for predicting cardiovascular disease (CVD) using Radiology Dataset. Int J Syst Assur Eng Manag (2022). https://doi.org/10.1007/s13198-022-01681-7.

[22] Nagendram, S., Nag, M. S. R. K., Ahammad, S. H., Satish, K., & Saikumar, K. (2022, January). Analysis For The System Recommended Books That Are Fetched From The Available Dataset. In *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 1801-1804). IEEE.

[23] Shravani, C., Krishna, G. R., Bollam, H. L., Vatambeti, R., & Saikumar, K. (2022, January). A Novel Approach for Implementing Conventional LBIST by High Execution Microprocessors. In *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 804-809). IEEE.

[24] Kiran, K. U., Srikanth, D., Nair, P. S., Ahammad, S. H., & Saikumar, K. (2022, March). Dimensionality Reduction Procedure for Bigdata in Machine Learning Techniques. In *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 836-840). IEEE.