TuijinJishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 04 (2023)

Performance Comparison of Security Level of Cryptosystems by using Machine Learning

G Samuel Joe Victor[1], K Rajasekhar[2]

[1]*Assistant Professor, Dept. of ECE, Narasaraopeta Engineering College, Narasaraopet, India.*

[2]*Assistant Professor ,Dept.of ECE, University College of Engineering Kakinada, JNTUK Kakinada, India.*

Abstract:-Recent developments in multimedia technologies have raised concerns about digital data security. However, many proposed encryption algorithms have been proven insecure over the past few decades, presenting a serious security risk to sensitive data. The best encryption technology should be used as protection against these attacks, but the type of data to be protected will determine the right algorithm in each case. Comparing different encryption systems one by one to find the best one can take a lot of time. We present a security level determination method for image encryption schemes that uses a support vector machine (SVM) for quickly and accurately selection of appropriate encryption algorithms. Furthermore, the dataset uses common cryptographic security techniques including entropy, contrast, homogeneity, peak-signal-to-noise ratio, mean square error, and energy. These variables are used as features for dividing the encryption algorithms extracted from the different images. Depending on the level of security, dataset labels are divided into three groups: weak, acceptable, and strong. The results show the advantage of Support Vector Machine (SVM) and we also improve the performance of the SVM by using XGBoost to improve the performance of our existing model.

Keywords*: XGBoost, Image Encryption, Support vector Machine (SVM), Cryptosystems.*

1. Introduction

The Encryption techniques chosen must be effective in preventing unwanted access to digital data. Additional to this, encryption techniques must be strong because the amount of security of the encryption algorithm used to encrypt an image affects that image's robustness. Confusion and diffusion are two important characteristics for digital picture encryption (scrambling) [2]. When it comes to digital photos, the diffusion alters the original pixel values and scrambling is done on pixels directly or on the columns and rows.

The data cannot be completely safeguarded during transmission by using encryption. The data will be transferred in an encrypted manner, but it is possible that unauthorized parties might still access it due to the lack of robustness in the encryption methods. A strong encryption technique is used to encrypt the whole image, making it unbreakable and secure against any attempts to compromise its confidentiality, privacy, or integrity. When choosing an encryption technique, time complexity is a crucial consideration when choosing the right encryption method. The type of application to be encrypted will determine which cryptosystem should be used because different forms of data will require varying levels of protection. In order to avoid problems, we classify all of the considered algorithms into three categories using the (SVM) [8], We have carried out many assessments to gauge the effectiveness of our Xgboost (precision and recall, F1 score, accuracy), and we will enhance the SVM prediction accuracy using the Xgboost. A contrast, energy, entropy or homogeneity statistical study had to be conducted on an encryption algorithm to determine its level of security. These jobs can be obtained by analyzing and verifying each encryption algorithm statistics about its security measures. After

engaging in such security evaluations of each encryption technique individually, we can select the option that is the strongest and greatest from those tested. However, this technique frequently consumes excessive time from completing the intended task. Instead, we suggest that a machine learning model can take the place of manual testing, it is capable of choosing the most secure encryption schemes swiftly, simply, and precisely. The security of encryption methods been divided into three tiers (weak, acceptable, strong) is based on standard encryption algorithm security parameter [3]. we separated the encryption algorithms as noted in Table I based on their level of security into three mentioned tiers by using the security features like entropy, homogeneity, contrast, PSNR, MSE, and energy values. All sorts of picture encryption techniques, including the chaotic maps, transforms, and frequency domain, have been taken into concern for the level of security detection. Finding the amount of security of the concerned encryption techniques is the primary goal of the suggested effort. We took into account a number of encrypted images and extracted the feature values to create a dataset. The size of dataset is not limited in any way. The feature values in the dataset should properly reference the acceptable and high security levels. As an example, we'll use entropy values: for these, we've decided to use a step size of 0.0001, [9].

The values of entropy were split into the three intervals, there are in the range of 7.9221 to 7.9147 for robust security. Likewise, there are values in the range of 7.8221 to 7.8022 for the acceptable security level. Every other result below will indicate a weak security status. The parameter values of other features were similarly separated into three ranges by choosing a suitable step size.

Principles for classification: The considered model must obey by the following rules in order to classify encryption algorithms into three groups (weak, acceptable, strong).The decision regarding each category's classification will be made based on the security parameter's values. Each parameter's range has been separated into three categories designated for weak, acceptable, and good. Below 50% of the feature values must fall inside the permitted interval values for the weak security level. At least 65 % of feature values must fall inside permitted interval values for acceptable security. More than 80 % of the feature values must fall inside the interval range for strong security.

2. Objectives

The objective of this research is to address the growing concerns surrounding digital data security, particularly in the context of multimedia technologies. Many existing encryption algorithms have proven to be insecure over the years, posing significant risks to sensitive data. Our primary goal is to develop and evaluate image encryption techniques that offer robust security. To expedite the process of selecting the most suitable encryption algorithms, we aim to implement a Support Vector Machine (SVM) model, supplemented with XGBoost for performance enhancement, to determine the security level of each algorithm. This automation will streamline the selection process, crucial as different types of data require tailored encryption methods. Additionally, we will evaluate encryption techniques for their resistance to unauthorized access during data transmission, considering time complexity and ensuring that the chosen encryption methods are both secure and efficient. Through comprehensive feature analysis and the creation of a well-structured security dataset, we intend to categorize encryption algorithms into three tiers: weak, acceptable, and strong. These objectives are designed to provide a framework for enhancing digital data security and encryption in the realm of multimedia technologies, with a focus on image protection.

3. Methods

Support vector machine (svm)

To determine the level of security offered by various encryption techniques, we have suggested a new dataset. The proposed dataset includes security criteria for the assessment of terms "strong," "acceptable," and "weak" are used to denote three different security levels, while encryption techniques are viewed as characteristics. We had fostered an original model with the help of support vector machine (SVM) [3] to decide the level of safety of various cryptosystems. We perform experiments and data analysis for variables including recall, precision, F1 score, and accuracy, then utilize the results to assess the value of the work.

A. Features as security parameters

Contrast: The difference in pixel values can be seen using contrast analysis. The image will have more contrast if the difference is more between pixels values. Better security is correlated with higher contrast, whereas a small differ between the original and manipulated pixel values is indicated by a lower contrast value. Contrast can be mathematically written as

$$Contrast = \sum \quad |x-y|^2 z(x,y) \qquad (1)$$

Where $z(x,y)$ represents the gray level co-occurrence matrices (GLCM).

Entropy:How much randomness an encryption algorithm has introduced into cipher image is revealed via entropy analysis. It depends on the number of bits in the image, different images have varied maximum entropy values. For instance, the picture is 8 bits, the maximum entropy value for that specific image will be 8. Similarly, the entropy value for a binary image of a single bit will never be greater than 1. The cipher image's entropy value needs to be near to the maximum value for effective encryption. Compute the entropy as

$$Entropy = \sum_{m=1}^{M} \quad p(s_m) s_m \qquad (2)$$

In Eqn. (2), $\boldsymbol{p(s_m)}$ represents the probability of occurrence of message $\boldsymbol{s_m}$ and M signifies the total number of pixels in an image.

Energy: This property is used to determine how much information an image contains. More information is included in the image when the energy values are higher. Simply said, original images have a larger energy value than cipher images because they have more information than cipher images, which have less information

$$Energy = \sum_{k=1}^{L} \quad [im(x,y)]^2 \qquad (3)$$

Where, Lrepresents the No. of pixels present in the original image $im(x,y)$,is the pixel dimension Xthrow and Ythcolumn.

Homogeneity: The gray level occurrence matrix (GLCM) presents a representation of the brightness of pixels. Homogeneity levels should be lower for a secure encryption. Calculating homogeneity is as easy as

$$Homogeneity = \sum_{a} \quad \sum_{b} \quad \frac{p(a,b)}{1+|a-b|} \qquad (4)$$

Where $p(a,b)$ represents the gray level co-occurrence matrices.

Peak signal to noise ratio and Mean square error: Any two images can be used to calculate the peak signal to noise ratio value. It is important to compute the mean square error value between the desired images of two prior to computing the PSNR value. The obtained image is highly similar to the original image if the (PSNR) value between the (original and cipher) image is high. MSE is inversely proportional to PSNR. Therefore, for a strong encryption, the PSNR value difference between the plain and encrypted picture must be at least one. There should not much of a difference between original and cipher image.

$$PSNR = 20\left(\frac{maximum\ value}{\sqrt{Mean\ square\ eerror}}\right) \qquad (5)$$

Where, maximum value refers the greatest value present in the original image.

$$MSE = \frac{1}{XY}\sum_{a=1}^{X}\sum_{b=1}^{Y}\left(p_{im}(a,b) - c_{im}(a,b)\right) \quad (6)$$

Where, p_{im} represents the plain image, c_{im} represents the cipher image and XY is the dimensions of the image pixel. In light of this, describe how the security parameters are derived from the cipher images and contrasted with the original images. Table.I, lists the values of the derived security parameters for cipher images that have been encrypted using various encryption techniques.

B. Support vector machine algorithm

It is mainly used to solve classification issues. SVM algorithm are used to offer the optimal line or decision border (Hyper plane) in a high or boundless layered space for use in grouping, relapse, and exception recognition exercise. Hyper planes that are uttermost from the closest preparation important piece of information for any class (alluded to as useful edge) accomplish a fair division since the grater the edge, the lower the classifier's speculation error. When training a support vector machine (SVM), each training sample is given a specific position in space in order to optimize the separation between two classes. Then, we anticipate whether a new set of instances belongs to a certain class by looking at which side of the chasm they land on. SVM employed in this case to evaluate the security level of several encryption algorithms, classifying them as strong, acceptable, or weak based on criteria of security parameter extracted values of cipher images.

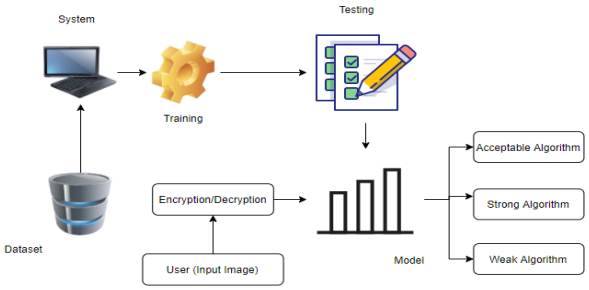

Fig. 1: Architecture of SVM

This process needs many inputs that can be considered as feature vectors, suppose some samples contains (X1,Y1), (X2,Y2) …(Xm,Ym), in which Yj represents the output and Xj represents the input. This type of data depends upon the Number of features, as demonstrated below.

For 2-D dataset: Y = (X1, X2) and for n-D dataset: Y = (X1, X2, ….Xm), where X1 and X2 both are independent features, through this SVM classifies the output (Yj).

Analysis of Existing Model:

We have done some experimental analysis to see how the current model works, and the results are below.

A. Confusion Matrix

A confusion matrix, often referred as an error matrix, is a particular table structure that makes it possible to see how well an algorithm performs when applied to the statistical classification issue in the context of machine learning. To provide precision, recall, and accuracy, the confusion matrix might be arranged in a two-dimensional array.

B. True positives

If the system indicates that security is strong, then "strong security" was also the end result.

C. True negative

If the outcome is “acceptable security" the system correctly foresaw "acceptable security". Another example is when the system predicts as of "poor security", yet the actual outcome was likewise "weak security.

D. False positives

If the system predicts "strong security," but the actual result was "weak or acceptable security".

E. False negative

If the system forecasts "weak security " or "acceptable security," yet the actual outcome was "high security."

Alternatively, if the system predicts "poor security" actual result was "adequate security."

By the usage of confusion matrix, accuracy expressed as:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \quad (7)$$

$$Percentage\ Accuracy = \frac{True\ positive + True\ negative}{Total\ samples} * 100\% \quad (8)$$

F. Precision and Recall

Precision is defined as the proportion of true positive predicted observations to all positive observations. In mathematics, this is equivalent to

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \quad (9)$$

The sensitivity of the model is referred to as recall. The more recall points, the response of the model will be more. This is how the ratio of true positive observations to all false negative and true positive observations is stated.

Recall computed mathematically as

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \quad (10)$$

G. F1 Score

When assessing the effectiveness of machine learning models, accuracy and F1 score are both crucial variables. The F1 score is crucial when F.N and F.P samples are significant, accuracy is crucial when T.P and T.N samples are there. One method of calculating F1score defined as

$$F1\ Score = \frac{2\ (Recall * Precision)}{Recall + Precision} \quad (11)$$

Proposed method:

Extreme Gradient Boosting, often known as XGBoost, is a prominent boosting method in which each prediction corrects the mistake of its predecessor’s. At first, a model is constructed utilizing the preparation informational index. Then, trying to address the weakness of the main model, a subsequent model is created. This course of adding models go on until the insignificant number of models has been added or the entire preparation informational collection has been accurately predicted [7].

XGBoost is a gradient boosting decision tree (GBM) plug in created specifically to increase speed and effectiveness. Deccision trees, bagging, random forests, and internal boosting make up XGBoost. The performance features are same for the proposed method also. Accuracy, F1 score, Precision and Recall achieved by our proposed method using XGBoost are noted in the Table II. We preferred to choose XGBoost

to exhibit better results of accuracy of classification. It provides a parallel tree boosting to efficiently and effectively address a range of data science challenges.

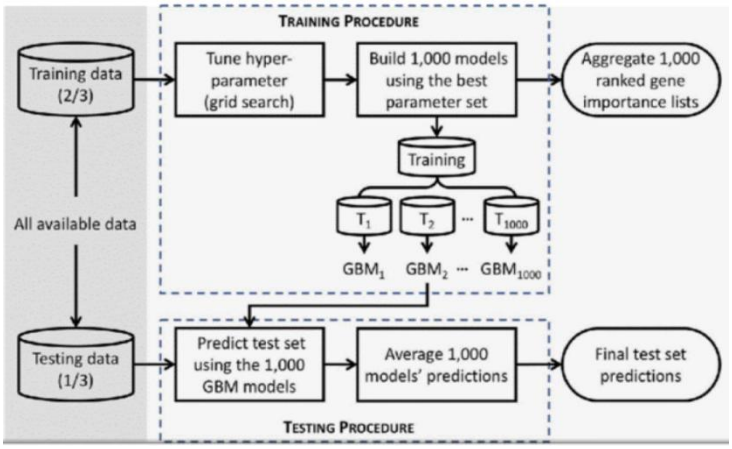

Fig. 2: Schematic view of XGBoost

In Fig.2 shows the schematic view of XGBoost where for training it takes (2/3) of the data and for the testing(1/3) of the available data. And then it goes to training procedures it takes many decision trees usually known as Random forest and then it goes through the gradient boosting model and then it goes for testing here also while predictions we do the gradient boosting and by optimization the final predictions will be out as a final result.

XGBoost has a strong mathematical background which can be available as a library in python .We can easily get XGBoost library using the pip installer by writing command in command prompt cmd as shown below.

```
pip install xgboost
```

The below command is written to import the proposed model for easy analysis and can be helpful while writing the code in a simple manner.

```
Import xgboostas xgb.
```

4. Results

The simulation results are shown in Figs. 3-6. Table II, shows the comparison of XGBoost with SVM in terms of F1 score, precision, accuracy and recall.

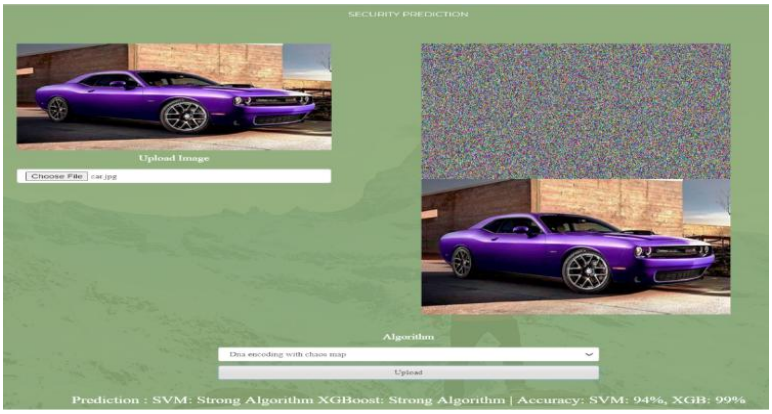

Fig. 3: DNA Chaos Map Encryption

The existing method (SVM) accuracy is 94% and the proposed method (XGBoost) enhance the accuracy as 99%.By the security Status it is considered as the strong algorithm.

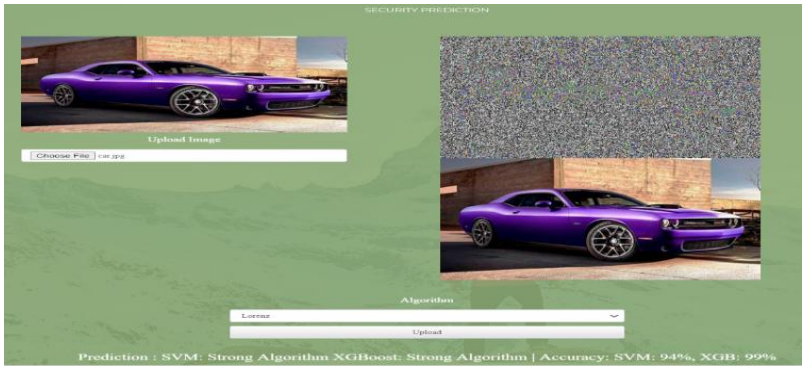

Fig. 4: Lorenz Image Encryption

The existing method (SVM) accuracy is 94% and the proposed method (XGBoost) enhance the accuracy as 99%.By the security Status it is considered as the strong algorithm.

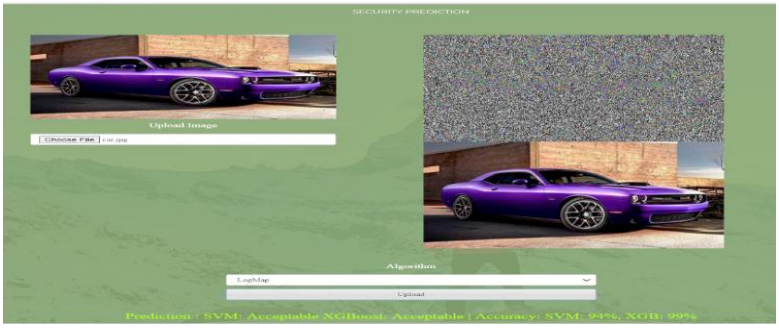

Fig. 5: Logistic Map Encryption

Table-I: Security Status of Encryption Algorithm by using proposed method

Existing Encryption Schemes	**Entropy**	**Contrast**	**Energy**	**Homogeneity**	**PSNR**	**MSE**	**Security status**
DNA encoding with chaos map [1]	7.90478	10.1975	0.0058	0.017044	27.91018	105.211163	**Strong Algorithm**
Lorenz image encryption [4]	7.90921	10.1975	0.00509	0.020194	27.998358	103.096655	**Strong Algorithm**
Logistic Map [5]	7.90611	9.1835	0.00508	0.019779	27.993255	103.21786	Acceptable Algorithm
Rubik's Cube image encryption [6]	7.82217	10.1975	0.00516	0.014846	27.912469	105.155856	**Strong Algorithm**
No encryption	7.00070	8.2463	0.02794	0.0299902	28.130804	100	weak Algorithm

The existing method (SVM) accuracy is 94% and the proposed method (XGBoost) enhance the accuracy as 99%.By the security Status it is considered as the Acceptable algorithm.

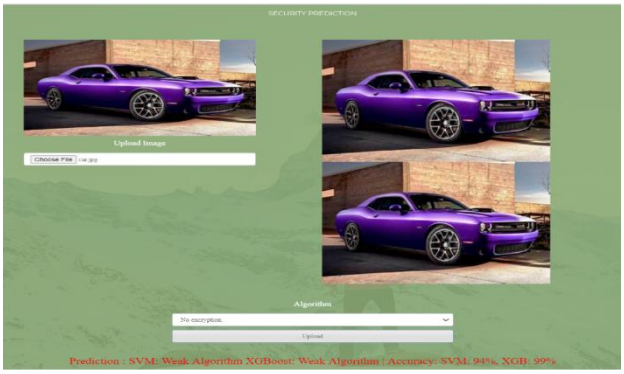

Fig. 6: No Encryption

The existing method (SVM) accuracy is 94% and the proposed method (XGBoost) enhances the accuracy as 99%. By the security Status it is considered as the weak algorithm.

Table-II: Accuracy comparison of (SVM) and XGBoost

Method	Accuracy	F1 Score	Precision	Recall
Existed method (SVM)	94%	0.93	1.00	0.83
Proposed Method (Xgboost)	99%	1.00	0.98	0.99

5. Discussion

In this work enhanced the Support Vector Machine(SVM) algorithm's accuracy by including the XGBoost and to evaluate the performance of the classifier we have analyzed the parameters which has noted in the Table II. Also, we have labeled the security status in the Table I by taking the security features into consideration like (Entropy, Contrast, Homogeneity, Energy, PSNR, MSE).We have enhanced the accuracy of the classifier with the proposed method (XGBoost) up to 5%.

Refrences

[1] J. S. Khan. Boulila, J. Ahmad, S. Rubaiee, A. U. Rehman, R. Alroobaea, and W. J. Buchanan, "DNA and plaintext dependent chaotic visual selective image encryption"," *IEEE Access*, vol. 8, pp. 159732_159744, 2020.

[2] A.Qayyum, J. Ahmad, W. Boulila, S. Rubaiee, Arshad, F. Masood, F. Khan, and W. J. Buchanan, ``Chaos-based confusion and diffusion of image pixels using dynamic substitution," *IEEE Access*, vol. 8, pp. 140876_140895, 2020.

[3] M. S. Anwar, J. Wang, W. Khan, A. Ullah,S. Ahmad, and Z. Fei, ``Subjective QoE of 360-degree virtual reality videos and machine learning predictions," *IEEE Access*, vol. 8, pp.148084-148099, 2020.

[4] WassimAlexan, Mohamed ElBeltagy, AmrAboshousha, "Lightweight Image Encryption: Cellular Automata and the Lorenz System", In: International Conference on Microelectronics (ICM), pp.34-39, 2021.

[5] MarwaTarek, WassimAlexan, Hisham Hussein, "Logistic Sine Map Based Image Encryption",In: 2019 Signal Processing: Algorithms, Architectures,Arrangements, and Applications (SPA), IEEE, pp.290- 295, 2019.

[6] Valeriu Manuel lonescu, Adrian-ViorelDiaconu, "Rubik's cube Principle based image encryption Algorithm implementation on mobile devices",In: 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). IEEE, p.-31, 2015.

[7] Hua li, Yumengcao, Siwen Li, Jianbin Zhao, Yutong sun, "Xgboost model and its application to personal credit Evaluation", IEEE Intelligent Systems, Vol. 35, Issue 3,pp. 52-61, 2020.

[8] J.Ker, Y. Bai, H.Y. Lee, J. Rao, and L.Wang, "Automated brain histology classification using machine learning", Journal of Clinical Neuroscience, Vol. 66, pp. 239-245, 2019.