I.J. Image, Graphics and Signal Processing, **2022, 4, 13-31**
Published Online August 2022 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2022.04.02

Enhancing Security of Medical Image Data in the Cloud Using Machine Learning Technique

Chandra Shekhar Tiwari
Department of Computer Science of Engg from Birla institute of Technology, Mesra, Ranchi (Jharkhand), India
Email: tiwaridalton@gmail.com

Vijay Kumar Jha
BIT Mesra (Ranchi) Computer Science and Engg department

Received: 15 December 2021; Revised: 15 February 2022; Accepted: 02 March 2022; Published: 08 August 2022

Abstract: To prevent medical data leakage to third parties, algorithm developers have enhanced and modified existing models and tightened the cloud security through complex processes. This research utilizes PlayFair and K-Means clustering algorithm as double-level encryption/ decryption technique with ArnoldCat maps towards securing the medical images in cloud. K-Means is used for segmenting images into pixels and auto-encoders to remove noise (de-noising); the Random Forest regressor, tree-method based ensemble model is used for classification. The study obtained CT scan-images as datasets from 'Kaggle' and classifies the images into 'Non-Covid' and 'Covid' categories. The software utilized is Jupyter-Notebook, in Python. PSNR with MSE evaluation metrics is done using Python. Through testing-and-training datasets, lower MSE score ('0') and higher PSNR score (60%) were obtained, stating that, the developed decryption/ encryption model is a good fit that enhances cloud security to preserve digital medical images.

Index Terms: Arnoldcat mapping, Playfair ciphering, k-means, random forest, image encryption, image decryption, medical image security, security enhancements.

1. Introduction

Cloud services consist of both secure platforms and insecure platforms that could be hacked and manipulated by third parties to bypass basic security like, single level password encryption) and hack crucial information of cloud users namely bank details, personal information, images and other sensitive data [1]. Thus private cloud security providers along with government enhanced the security levels through double-level security, complicated algorithm, higher encryption keys and biometric access for decryption [2] that enhances the accessibility of data. Thus, sender and receiver or sometimes a user alone would be able to access the data. Data encryption in digital era evolved from single-level towards double-level encryption where image encryption involves higher security levels since images are crucial and significant datasets that could be edited/ manipulated by third parties and hackers, where they might misuse the images to manipulate the users for monetary gains [3].Simultaneously medical images are crucial and significant, since it could affect the patients' health physically and mentally; thus nowadays medical practitioners ethically doesn't share confidential images of their patients until its necessary for second opinions [4]. However, while file sharing, the network path or the cloud storage could be bypassed/ hacked and the same could be misused by hackers which is identified as issues in "Caching" and Pre-fetching" [5]. To prevent these issues, the cloud services based companies and private security providers developed advanced and complicated algorithms to enhance the security levels [6] where usage of AES encryption in image encrypt/ decrypt model and auto-encoder model is common but the current security models have evolved along with technological impact [7]. Other machine learning techniques like XOR and Affine Transform, Chaotic system, One-dimensional Random-Pixelscrambling, Block displacement of Explosive n*n with inter-pixel displacement, neural network, novel encryption, double-layer chaotic network, ciphering, advanced cryptography, SD-AEI, multi-dimensional chaotic system, Unified approach, Blowfish algorithm, RC2 color stream chaotic and ciphering, GS-IES, etc [8]. Among other techniques machine learning chaotic mapping was found effective, easier, faster and reliable [8]. Henceforth adopting machine learning and neural network based models for securing datasets especially the digital images in the medical industry seems appropriate.

Researchers [9] argued that, cloud security in healthcare system is easily compromised with singe-level encryption where electronic health-data is shared, transferred, stored/ accessed and reviewed by many personnel and staffs in the healthcare systems; personnel with encryption (password) could access the data like images, statistics, medical records, patient history, health insurance, biomedical research information and also other irrelevant records and information that

could cause chaos. Henceforth monitoring the entry/ exit of every personnel access and their time per access was made necessary in healthcare systems to avoid manipulation and misleading malpractices [10].

Thus to ensure the security and privacy in the healthcare systems through cloud services, enhancing algorithms and developing hybrid models to increase the complexity in encryption/ decryption levels, especially in the digital images was found effective [11]. The proposed aim in the research is to enhance the security of the datasets (medical images) in cloud through machine learning. The study will be adopting the Arnoldcat algorithm and Playfair ciphering algorithm as image security enhanced techniques through Convolutional Neural-Networks (CNN) as deep learning technique.

1.1. Research Contribution

The developed research comprises of ciphering (Playfair) combined with chaotic map (Arnoldcat) based algorithms which maximizes the security of image stored and accessed in cloud than single-level of algorithm. Machine Learning towards digital image security based coding and encoding have been increased recently due to digitization of medical recods (hospital and patient records) few government and many private hospitals have adapted to usage of cloud storage and servers [12]. However, developing and underdeveloped countries lack in security systems that aid hackers and third-parties to gain access of personal information of patients. Thus to enhance the level of security the research has contributed the adaption of double-level encryption of digital images in medical sector.

1.2. Research Problem

The model to be developed will utilize the Playfair with Arnoldcat algorithm to secure digital medical images in cloud. Medical images and information have been at risk where personal information, payment access, data hacking and other risks are higher especially in the underdeveloped countries, poor information technology based countries and developing countries. The existing researches either uses single-level encryption model like chaotic mapping [13], Cyclic-Redundancy-Check [14] or double-level encryption using Arnoldcat and Henon algorithms [15] or combining two or more methods into a hybrid model like Secure-Hash-Algorithm with blowfish model [16]. The lack of PlayFair algorithm in image encryption especially in cloud security enhancement paved-way for the developed model, where Arnoldcat with Playfair could optimize image encryption and security in cloud.

2. Literature Review

Studies by [17,18] focused on cloud service enhancement and privacy of medical images, where authors described that, storage and access of datasets had to focus upon: splitting of data, monitoring and access controlling, enforcing measures of security and integrity along with monitoring of accesses through intranet and internet network paths. This could be achieved only through double-level encryptions, enhanced cloud services, enhanced AES (Advanced-Encryption-Standard) models, SSL (Secure-Socket-Layer) models and similar techniques. Studies concluded that, secured model and framework ensures the dataset's security in cloud, thus, researchers should centre their ideas on developing enhanced and complex model that would provide higher security and adopt double-level encryption, especially in image security in healthcare systems. [19] examined in-depth about algorithms and models in image security that ranges as: cryptography, ciphering, Convolutional networking, artificial intelligence, deep belief-networking, secret sharing, embedding techniques, watermarking mechanisms, chaotic mapping, block coding and image optimization schemes. Among all techniques, the authors found, chaotic mapping, ciphering, cryptography, CNN and ANN as effective techniques that provides reliable/ accurate outcomes when compared to other results.

[20] had examined cloud services in the healthcare systems and how e-health cloud securities tackles cloud based challenges. The authors simultaneously studied how e-healthcare is protected and secured in cloud through examining the existing literature reviews; the research found significant outcome that, cloud services and the service providers offers the healthcare industry with information-centric models that allows 'internal file-sharing access' without restriction from staffs and employees and contrarily restrains file sharing to/from 'external paths' through cloud. Though this model is effective, it affects external accesses when needed by the same employees outside their network paths resulting in hindrance and bypassing the security regulations and sometimes sharing their access code with peers leading to higher security challenges and privacy issues in medical datasets. Hence authors concluded that, advanced technique that allows internal and external accesses with complicated security keys and biometric access would resolve the problem. However it might need the presence of the user for every-access which should be focused and examined.

[21] argued that ciphering and cryptography in image encryption is secured process where as the authors [22, 23] argued that chaotic mapping like Arnold Cat's mapping as scheme with cryptography is more secured than plain cryptography and chaotic mapping. [24] focused on hybrid techniques towards image decryption/encryption and found techniques apart from AES. The model developed by authors includes hybridization of ECC (Elliptic Curve-Cryptography) with HC (Hill Ciphering), DPC (Double Playfair-Ciphering) with ElGamal and ECC combined with AES as three models and evaluated the models through metric evaluation of PSNR (Peak Signal-to-Noise Ratio), UACI (Unified Average-Changing Intensity), NPCR (Number of Pixels-Change-Rate), entropy and encryption/decryption time estimation. The PSNR estimation in Playfair ciphering hybridized with ElGamal as symmetric algorithm stood-out

as higher outcome model that produced good quality images post encryption/ decryption whereas ECC-hybridized-with HC was found to produce better image encryption/ decryption outcome.

In earlier studies by [25,26,27] image encryption especially in the medical and hospital sector lacked security and safety. Algorithm developers enhanced, modified and wrote new algorithms that could encrypt images in similar process to text encryption. However, the accuracy of models was found lesser by the authors along with image quality. Since the models developed in early 20s by Ye and Zhao and Khalil focused upon single-level encryption with chaos-based mapping algorithms. However, authors Bhogal et al., focused on quality of image degrading post coding. From these researches, it was understood that, image quality degrades post coding and decoding unlike text encryption/decryption. Thus in the following year authors [28] focused on double-level encryption/decryption model that has Chaos based and cellular automata (CA) technique. From all four studies, its observed, image coding has more disadvantage with single-level encryption and decoding, where either data loss or data quality is degraded. Hence, to retain the quality with minimal loss (or no-loss) enhancement in existing algorithms was needed rather than developing new algorithms. Thus chaotic mapping algorithms, reversible algorithms, hash algorithms, ACM, and others were modified and existing models were remodelled according to research necessity.

In [29], developed a model that secures the access, storage and protection of using data in cloud-servers in the medical/ hospital sector. The study focused on private clouds and also the public clouds in the Poland based hospitals. The servers were used by hospitals to store the digital data and authors developed the model to find how often they were either breached or accessed illegally. The findings showed, higher the cost in implementing secure data storage in cloud the higher the security provided by the service providers in private clouds. Contrarily, the government cloud servers provided lesser cost and lesser security and hosts (hospital management) preferred lesser cost to avoid maximum loss. Authors conclude that, especially in European, Asian and African countries (like: Pakistan, India, Nigeria, Norway, Austria, and more) the lack of technological implementation and advanced technology in medical sector has positively aided the hackers/ third parties to penetrate the cloud and servers easily. Henceforth, implementing secure clouds and servers in hospital sector is mandatory for digital patient records.

In recent image encryption/decryption model based researches use of the PlayFair have been either rarely identified or couldn't be identified, completely. Similarly, the current researches use SSIM (Structural Similarity Index) as image-processing metric analysis to measure data loss and data quality. Thus by reviewing existing models and techniques, research focuses on:

- PSNR estimation metric evaluation,
- PlayfairCiphering,
- ArnoldCatmapping with
- K-means and Random Forest regression techniques.

3. Research Background of Algorithm Proposed

The research is developed towards examining the decryption and encryption based techniques towards enhancement of the cloud services based medical image security and safety. Henceforth the encryption/ decryption in image security are focused towards examination of better security measures against other security measures. Generally the image encryption is designed and developed through basic algorithm where AES encryption is commonly used. In this research, Arnold's cat as encryption/ decryption is used with Playfair ciphering method for the auto-encoder model.

3.1 Arnoldcat mapping algorithm

The Arnoldcat mapping (ACM) is of chaotic mapping category that generates random number via generator with lesser parameters. Thus the mapping technique is widely used for its speed, accuracy, flexibility and ease-of-use [23]. ACM is used in image encryption where the image pixel position could be altered without changing original information of image.

The research developed the following Arnoldcat algorithm for mapping:

- Initially load the images and transform the images through Arnoldcat;
- Get the rows,columns, channel from the image shape where n==ch;
- Initialize the img_arnold as '0' 'zeros';
- For every 'x' and 'y' within range of columns and rows, respectively, calculate img_arnold[x][y] = img[(x+y)%n][(x+2*y)%n];
- Return the img_arnold.

Through the above algorithm the image in Arnoldcat is encrypted for security enhancement as 1st level through chaotic mapping technique.

3.2 Playfair ciphering algorithm

The Playfiar ciphering is usually adopted for plain text based ciphering like passwords and text-security where the texts are complicated with 'bigrams' (paired letters as encryption technique). It is easier, simpler, complex and robust; hence researchers use Playfair ciphering making plain texts complicated [24]. In image or digital image security the modified or enhanced Playfair is used to enhance the digital images' security. Thus the algorithm is developed as:

- Initialize key, randomly, with ranges of unique values from 0 to 256;
- Encrypt the images with its' key, row and column values;
- Acquire the values of rows, channel and columns from 'img.shape';
- Initialize the Encrypt_img as '0';
- For every i,j,k in the row,col,ch, encrypt the images through Encrypt_img[i][j][k]=key[img[i][j][k]];
- Finally return the 'encrypt_img'.

4. Proposed Model and Approach

The study is designed and developed based on the following approaches and architecture:

4.1 Flow Chart

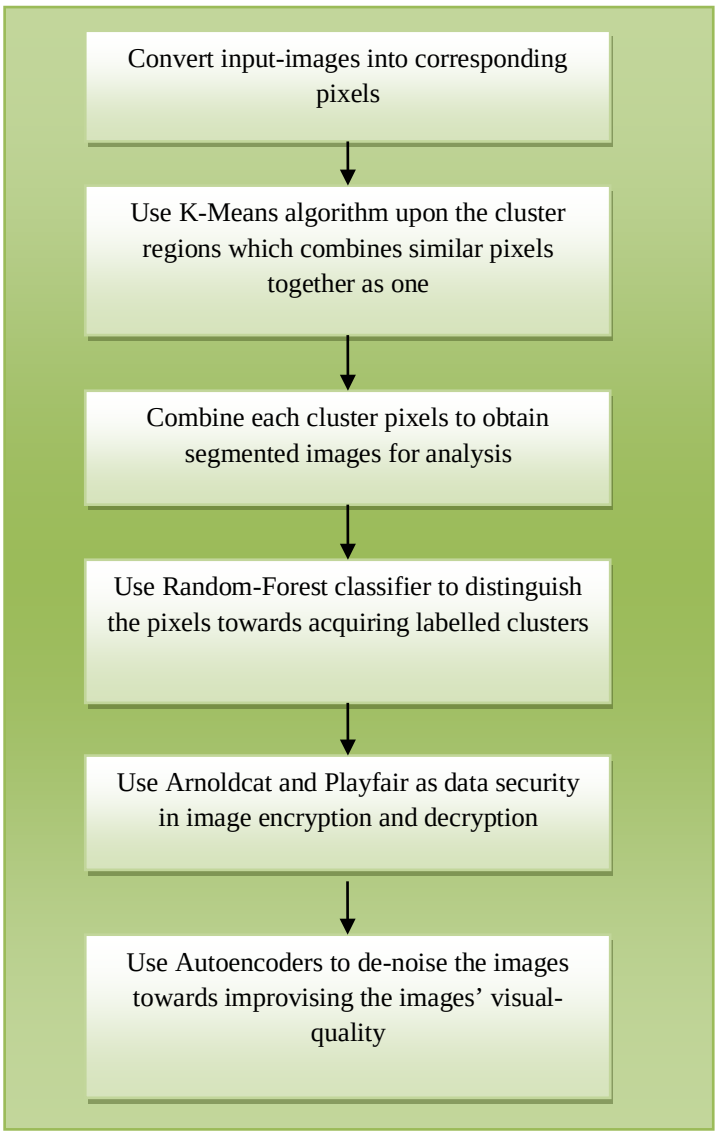


Fig. 1. Flow-chart for the research proposed

The flow of the research (refer fig-1) initiates where the input images are converted into required pixels and pre-processed. Once the images are pre-processed for training they are stored in local cloud so that, the model developed could access the data and apply algorithm for obtaining results for cloud security and image security enhancement. Here K-Means with Random Forest is utilized where K-Means clusters the similar regions towards segmenting the images and Random Forest classifies the segmented images into Covid and Non-Covid datasets, as labels. For encryption/ decryption, Arnoldcat mapping with Playfair ciphering based model is developed with auto-encoders, where de-noising is made and image quality is enhanced for better image as visual outputs. Each step involves varied process and procedures and thus the research will focus upon individual processes in-depth.

4.2 Research approach

The research since focuses upon image security enhancement as protocol, it adopts the following approach:

- Initially acquire the images for the research and use them as inputs;
- Next convert the images with corresponding pixels to the research;
- Make use of 'K-Means clustering' as algorithm to group regions that has similar pixels, collectively;
- Use the 'Elbow' method for determining the optimal no-of-clusters in the processed inputs;
- Next combine the clustered pixels with 'cluster centroid' as base towards segmenting the images that are processed as inputs;
- Now, apply 'Random Forest' as regressor technique for classification of pixels into gathered clusters;
- Design the image encryption/decryption model with 'ArnoldCat' mapping and 'Playfair' ciphering, respectively;
- Finally use the CNN 'auto-encoders' for de-noising the images and also for enhancing the visual image-quality, without compromising overall details using PSNR and MSE evaluation techniques.

Thus the outcomes are achieved to compare the original data with encrypted/decrypted data where the PSNR values with Loss values are weighed for data loss and image quality post image security enhancement through adopted algorithms and mapping techniques. The image segmenting through K-Means clustering is estimated with R (Red), G (Green) and B (Blue) as centre values with pixel corrections.

4.3 Proposed model Architecture

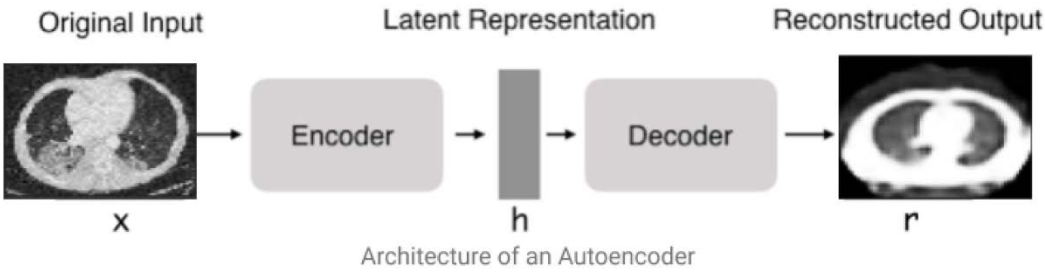


Fig.2. Model Architecture for medical image security enhancement

The model (refer to fig 2) is developed with stacked Convolutional Neural Networks (CNN) as deep learning method where commonly NN (neural networks) based models adopt hidden kernels and layers [30]. The architecture includes 64x64 input, 5concolutional layers, 2maxpooling and 2upsampling layers where in final layer the images are flattened (refer appendix A).

4.4 Proposed Auto-Encoder model

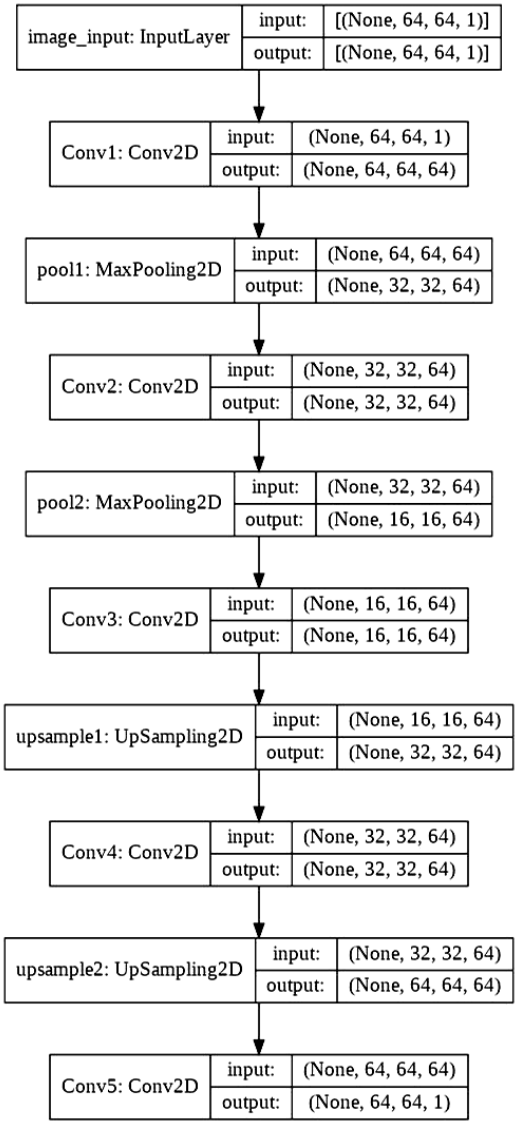
The auto-encoder model along with de-noising algorithm in research is developed for image enhancement to suppress the noise from images that are noise-contaminated. Thus pre-processing images is necessary prior passing them through encryption/ decryption model. For de-noising: Gaussian, Bilateral, Median, Average, Mean, Weiner, etc are performed upon images to preserve the image. Though varied filters perform differently, each method has minimal loss and de-noising algorithm is required to retain the image values without losing the information. Thus algorithm for de-noising in this research is developed as:

4.4.1 Pseudo-code

- Read the 'image' as input datasets;
- Noises are added to images for training the datasets;
- Encoder model is built model that compresses data by utilizing the deterministic mapping;

- A decoder model is built that decompresses the original images through compressed data;
- The PSNR values are calculated for the developed 'auto-encoder model' towards performance evaluation;
- When the obtained PSNR value is higher from original value, reconstructed and de-noised images are better and same as original.

Appendix A



4.4.2 Approach

– *Training datasets by introducing noise*

To train the datasets the researcher purposely corrupted images by introducing noise through 'stochastic behavior', i.e. a random normal-distribution of mean '0' and SD '1', along with 'multiplying factor' value of 0.07, as default dataset that contains noiseless images is carried out. This process significantly stimulates noise which could occur with the datasets acquired i.e. 'medical images' since images are corrupted. Post this process de-noising is done to test the accuracy of the de-nosing algorithm.

– *De-noising using Auto-encoder*

Machine-Learning technique (unsupervised) based 'auto-encoder' is technique which uses artificial neural networks (ANN) is used for representing learning and it consists of two parts:

a) ***Encoder***: Initially it accepts data as input and then maps it as a 'latent-space'that represents the compressed data with deterministic mapping;

b) ***Decoder***: The latent-space representation is accepted here and decoder maps it as reconstruction which resembles the input datasets shape with mapping.

In simpler terms, the auto-encoder accepts input datasets i.e. the corrupted medical CT-Scan images and then locally compresses the data sets to hidden space and reconstructs the data from hidden representation as final outcome.

4.5 Advantages of adopted algorithms

The research adopts K-means clustering algorithm and Random Forest as regression technique to cluster the regions of segmented pixels and then classify the images under Covid and non-Covid categories, respectively.

For mapping images in encryption the Arnold's cat as mapping technique is used which is of cryptography based algorithm and falls under chaotic map category. Contrarily, the Playfair is the ciphering technique in encryption/decryption model that uses 'digraph substitution' as the cipher technique.

4.5.1 Advantages of K-Means

Though other clustering techniques (hierarchical, k-medoids, db-scan) are also used by researchers for image segmenting, K-Means is selected here for its better advantages, where: simpler to implement; guarantees convergence; large datasets are scaled better in K-Means over other clustering; easy adaptation with new examples; centroids' position could be warm-started; manual choice of datasets and calculations; initial values are dependent; generalization for clusters natures, shapes and sizes.

4.5.2 Advantages of Random-Forest

The Random-Forest as regressor have been selected here over SVM, logistic regression, decision tree and other algorithms since it has more positive advantages in image classification. The advantages are: it is very flexible in classifying data and also in regression problems; it automatically computerizes and allocates values for missing data with given datasets; reduces significantly the decision-tree over-fitting thus assists researchers to improvise the accuracy; it works great in continuous and categorical values and data normalization is not needed since it utilizes rule-based approach.

4.5.3 Advantages of ArnoldCat

TheArnoldcat is used for digital image securing; securing images with large pixel ratios without decreasing its quality; securing files and texts in clouds without decreasing its size and value and digital images are secured and retrieved with higher visual quality unlike other encryption/ decryption mapping models.

4.5.4 Advantages of Playfair

ThePlayfair is complex; hard to penetrate; it uses vast range of digraphs up-to 625 (25x25) than monographs (25); frequency analysis; requires more ciphers towards cracking the codes and encryption in Playfair.

Thus the study aims at higher and enhanced security level in managing and storing medical digital-images in cloud that cannot be compromised easily.

5. Methodologies and Statistical Approach

The study is based on quantitative datasets and seeks outcomes from existing methods and approaches. It adopts positive paradigm, quantitative approach and experiment design. Thus the following sampling techniques and dataset configurations were examined and adopted for the developed research.

Task: The research is based on three main tasks: (a) to use Arnoldcat mapping along with Playfair algorithm to encrypt/ decrypt the medical images as input-data; (b) to use K-means algorithm to cluster similar regions into clusters for image segmentation and Random Forest regressor for classifying the segmented datasets into Covid and Non-Covid labelled files in local cloud and (c) to use the auto-encoder for de-noising or removing noise in images post encryption/decryption.

Originality: The model developed for image encryption/decryption uses the chaotic mapping with 'Arnoldcat' algorithm. Unlike existing models from the studies [13] (ICA: Independent-Component-Analysis with ACM: ArnoldCat algorithm), [15] (ACM confusion mapping with Henon diffusion mapping), [31] (5 dimensional 3 leaf chaotic model) and [32] (PWLCM: Price-Wise-Linear-Chaotic-Map), the developed research uses ***Playfair***ciphering combined with ***Arnoldcat*** mapping. Thus the developed model is unique among other medical image encryption/decryption models, especially in cloud storage based security enhancement.

5.1 Methodology

a) Dataset acquisition

The datasets are pre-existing data acquired from *"kaggle"* website through the link:

https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset

The datasets includes 1252 tested positive SARS-CoV-2 (Covid-19) patients' CT-Scans; 1230 tested negative SARS-CoV-2 (Non-Covid-19) patients' CT-Scans totalling ***2482 CT-Scans***. The data is public and could be accessed by anyone globally and thus the research would be appropriate to choose these datasets as samples. The patients were from Sao-Paulo (Brazil) of gender category male and female, belonging to age-group of 15years and above.

b) Training datasets

The datasets are trained for testing the model prior through:

Through the developed model, ***20images*** as samples are trained initially. The samples are pre-processed and ratio of each individual image is set correspondingly to 10*8as size. The training datasets are passed through K-Means for clustering and Random Forest for classification of categories "Covid" and "Non-Covid" labelling. Then the images are passed through auto-encoder model with Arnoldcat and Playfair technique for encryption/decryption where the Loss is estimated through mean-square error (MSE) along with the peak signal-to-noise ratio (PSNR) values and when the PSNR is higher than original value with low MSE value, the model is saved for testing the reaming datasets (refer to fig 3).

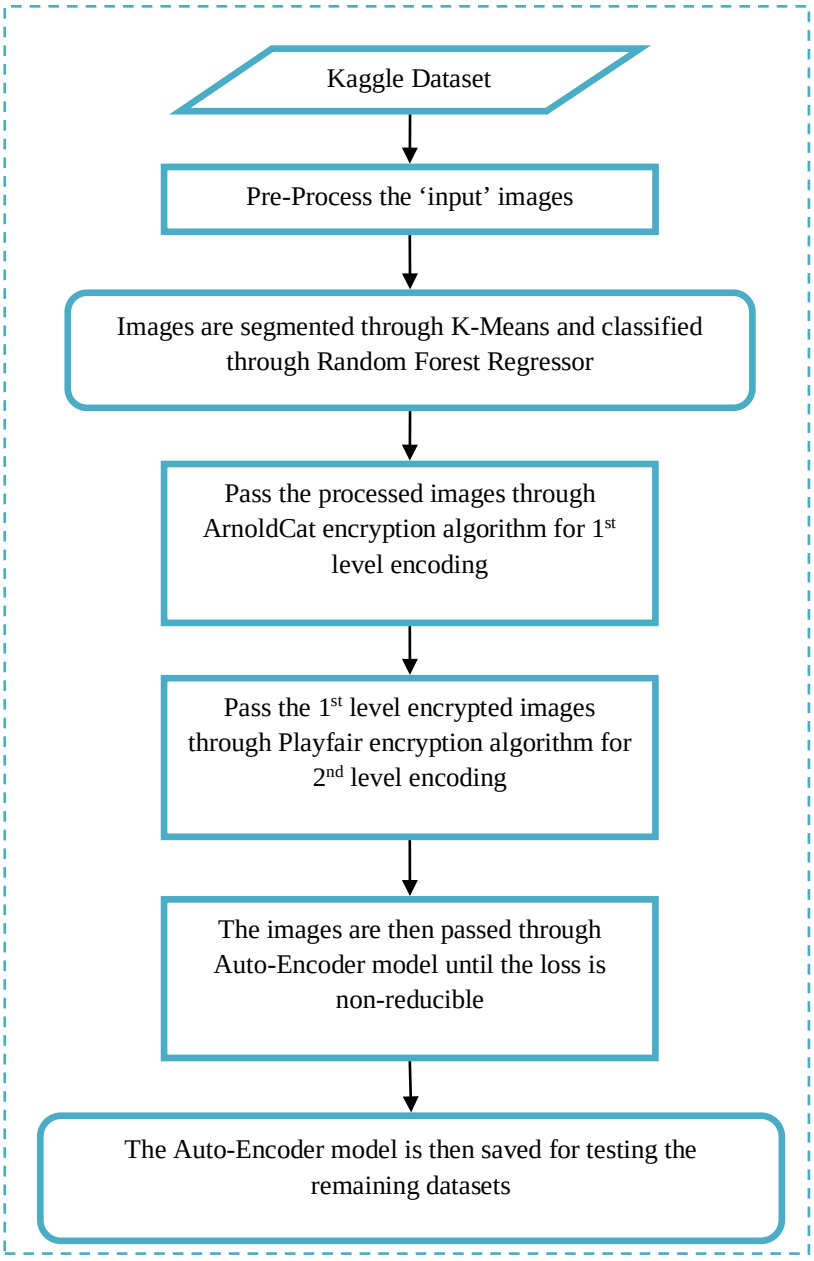


Fig.3. Training datasets through auto-encoder model

5.2 K-Means Clustering

In K-Means, with specified set-of 'n' observations (n_1, n_2, ..., n_x) that has dimensional real vector (*d*), the '*x*' observations are portioned as *c (≤ x)* sets where S={S_1, S_2, ..., S_c}. Henceforth to minimize clusters-within, the squares are added together (*WCSS: Variance*) and primary objective is to obtain:

$$ave_S mini \sum_{a=1}^{c} \sum_{n \epsilon S_a} \|n - \mu_a\|^2 = ave_S mini \sum_{a=1}^{c} |S_a| \, Var \, S_a \tag{1}$$

where, μ_a denotes the mean of S_j points, which is evidently equivalent to minimize teach points' squared deviations (i.e. pair-wise) with same cluster.

$$ave_S mini \sum_{a=1}^{c} \frac{1}{2|S_a|} \sum_{n_1 m \epsilon S_a} \|n - m\|^2 \tag{2}$$

By using the following identity below, the equivalence could be deduced:

$$\sum_{n \epsilon S_a} \|N - \mu_a\|^2 = \sum_{n \neq m \epsilon S_a} (n - \mu_a)^T \, (\mu_a - m) \tag{3}$$

Since the total-variance remains constant, it is equivalent to the maximization of total sum of squared-deviations that lies between points of different clusters (*BCSS: law-of-total variance*)

5.3 Random Forest Regressor

The Random Forest as regress or (RFR) is a tree-decision based technique used by researchers for noise based analyses. The RFR algorithm to train datasets in research adopts the bagging (i.e. bootstrap aggregating) technique where generally the beginner-level to expert-level decision-tree learners adopt this technique since it provides more accurate and higher precise outcomes.

In training model, the given datasets, say, Y=y_1, y_2,...,y_n where responses Z= z_1, z_2,...,z_n the bagging repeatedly opts for random samples along with alternate replacements (*A* times) for training datasets towards fitting the tress for the samples.

For a = 1, ..., *A*:

- Sample, with alternate replacements, training examples *'n'* from *Y, Z*; are denoted as Y_a, Z_a.
- Train the regression tree or the classifications f_a, on Y_a, Z_a.

Post training datasets, predictions upon unseen samples z'could be made through calculating or averaging every individual regression based predictions from trees on z', either through the following formula or through majority votes from classification trees:

$$\hat{f} = \frac{1}{A} \sum_{a=1}^{A} f_a(z') \tag{4}$$

This procedure of bagging or bootstrapping leads towards better model-performance since it decreases model's variance without increasing bias; which states, though single-tree predictions are highly sensitive towards noise within training datasets, average of tree-clusters are not sensitive when the trees are non-correlated. Generally, in single-training set, training many trees would offer researchers with correlated trees that are robust (or if a researcher adopts and developed a deterministic training algorithm same tree is returned as correlated sets many-times); the bootstrap sampling represents the de-correlated tress through different training sets. Thus to estimate predictions' uncertainty, SD (Standard Deviation) through each tress' individual regression, on z' is calculated, where:

$$\varphi = \sqrt{\frac{\sum_{a=1}^{A} (f_a(z') - \hat{f})^2}{A-1}} \tag{5}$$

Samples/trees '*A'* is estimated as free parameter; where generally few hundred or even sometimes several thousand as trees are utilized depending upon nature and size of training sets. Optimal number *(A)* of trees could be gained by utilizing cross-validation or even through observing out-of-bag errors. Mean-prediction-error upon individual training sample z_i is estimated through trees that has no bootstrap sample z_i. Similarly the errors from testing and training lean towards levelling off post number-of-trees are fit.

6. Analysis

The analysis is done through K-Means and RFR techniques, where the outcomes are:

6.1 K-Means Clustering

6.1.1 K-Means

K=3: The pixel segmenting of image is done through K-means at different levels, where in table 1 below, ***k=3***as optimal value is tested:

Table 1. K-Means with optimal number as '3'

Post Pixel	B	G	R	Label	B-center	G-center	R-center
0	2	1	3	0	79.789114	45.285024	20.562078
1	2	1	3	0	79.789114	45.285024	20.562078
2	2	1	3	0	79.789114	45.285024	20.562078

The table 1 represents the R, G and B center values 20.562078, 45.285024 and 79.789114 respectively, with 'k=3' as optimal number and provides the following output (refer to fig 4) with original and segmented result:

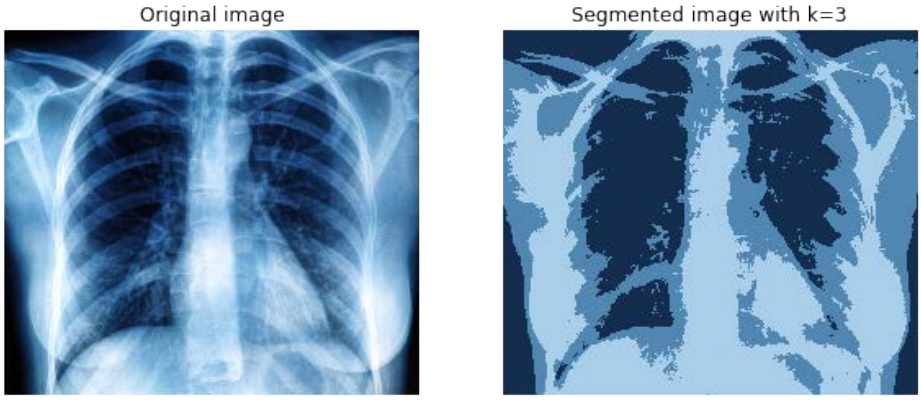


Fig.4. K=3 as optimal value

K=4: In table 2 below, ***k=4*** as optimal value is tested:

Table 2. K-Means with optimal number as '4'

Post Pixel	B	G	R	Label	B-center	G-center	R-center
0	2	1	3	2	64.198793	34.362724	15.442403
1	2	1	3	2	64.198793	34.362724	15.442403
2	2	1	3	2	64.198793	34.362724	15.442403

The table 2 represents the R, G and B center values 15.442403, 34.362724 and 64.198793 respectively, with 'k=4' as optimal number and provides the following output (refer to fig 5) with original and segmented result:

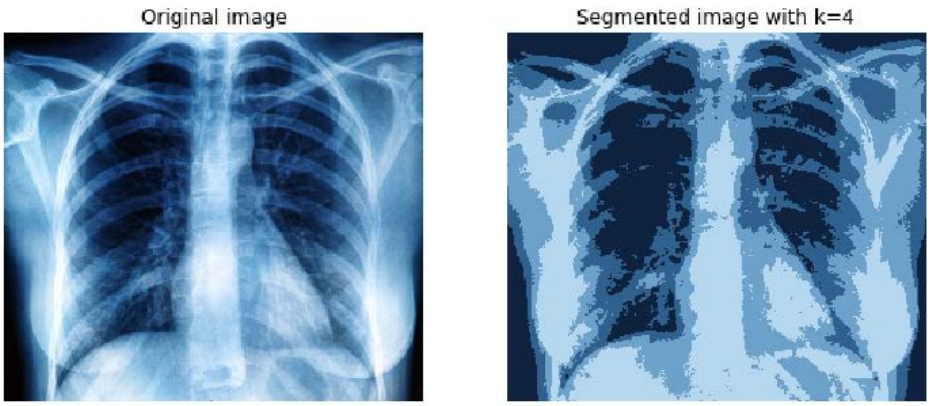


Fig.5. K=4 as optimal value

K=6: In table 3 below, ***k=6***as optimal value is tested:

Table 3. K-Means with optimal number as '6'

Post Pixel	B	G	R	Label	B-center	G-center	R-center
0	2	1	3	2	45.216023	22.958172	10.800741
1	2	1	3	2	45.216023	22.958172	10.800741
2	2	1	3	2	45.216023	22.958172	10.800741

The table 3 represents the R, G and B center values 10.800741, 22.958172 and 45.216023 respectively, with 'k=6' as optimal number and provides the following output (refer to fig 6) with original and segmented result:

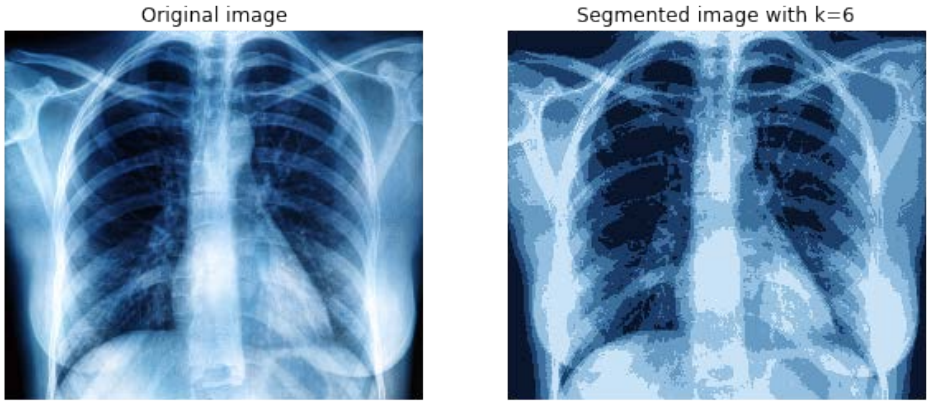


Fig.6. K=6 as optimal value

6.1.2 Optimal Cluster through Elbow method

Elbow method is used in K-means for average cluster distortion scores:

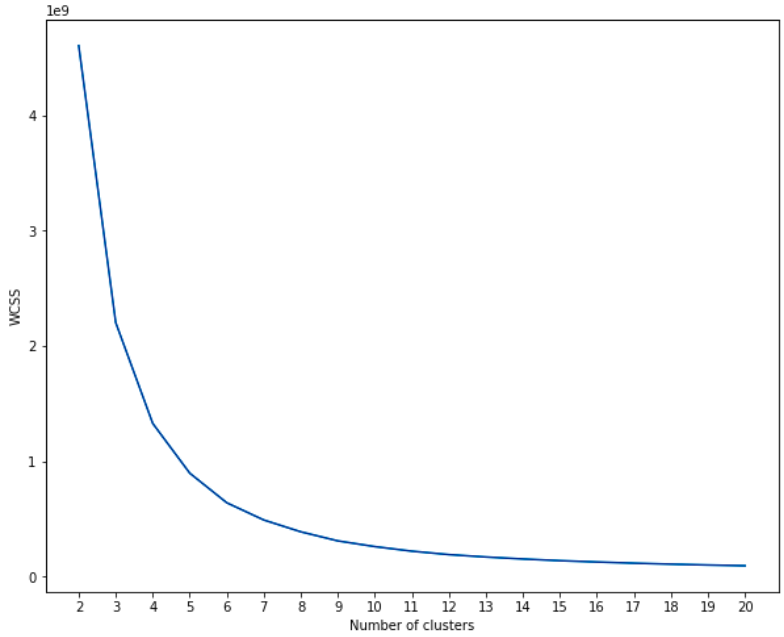


Fig.7. Elbow method

Inference:

Through figure 7, the optimal values of WCSS are plotted and at k=6the K-Mean optimal is achieved.

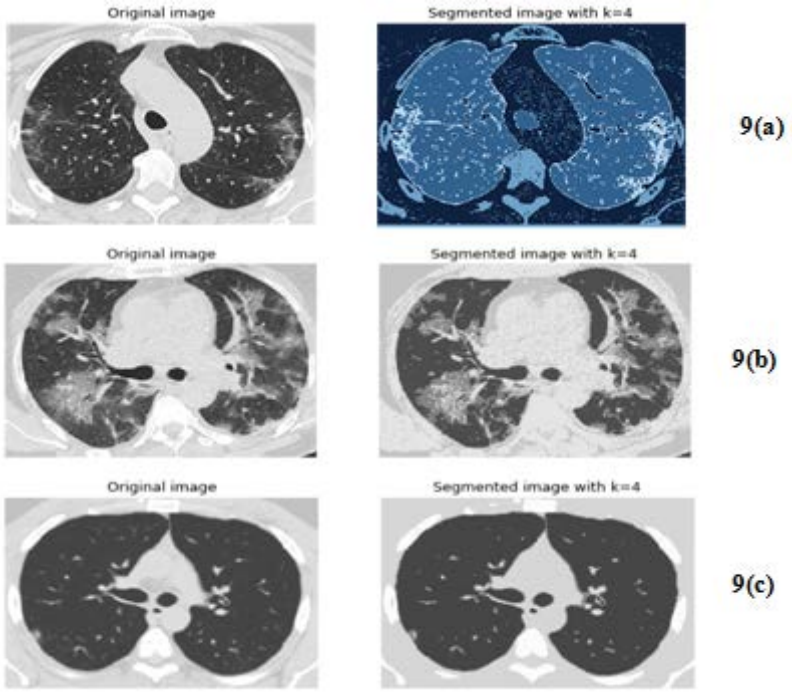


Fig.8. Image segmentation where k=4

Inference:

The figure 8 represents the original image (on left-side) without segmentation and the segmented (on right-side) through K-Means where the optimal value is k=4. It could be inferred that the quality of segmented image from original image post segmenting varies.

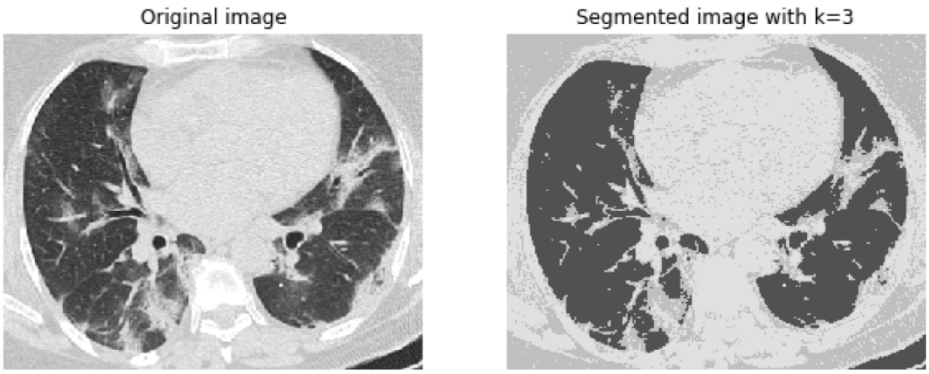


Fig.9. Image segmentation where k=3

Inference:

The figure 9 represents segmented image and original image where k=3 with low quality of segmented image from original.

Henceforth enhancing segmented images to its original image quality is necessary and thus Arnoldcat and Playfair algorithms are carried out post K-Means and Random Forest applications on datasets.

6.2. Random Forest Regressor (RFR)

The RFR technique is used for classifying the datasets post segmentation of images. Here, classification is done as: Covid and Non-Covid post clustering and segmenting of images.

Table 4. RFR Classification values

	Precision	Recall	F-1 score	Support
0	1.00	1.00	1.00	50624
1	1.00	1.00	1.00	43441
2	1.00	1.00	1.00	42798
3	1.00	1.00	1.00	53397
micro avg	1.00	1.00	1.00	190260
macro avg	1.00	1.00	1.00	190260
weighted avg	1.00	1.00	1.00	190260

Inference:

The table 4 represents that, the accuracy of the predictions from RFR technique is 100% stating that, the model trained classifies the datasets with 100% precision, recall and f-1 score.

6.3 Arnoldcat mapping

The outcomes post transforming images through Arnold mapping technique:

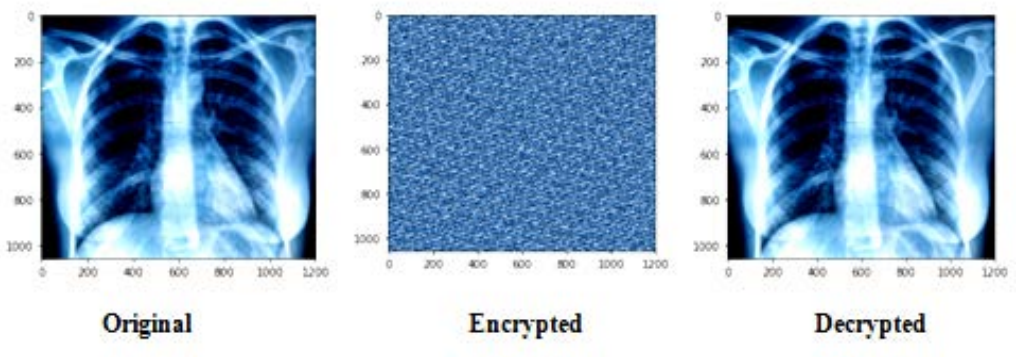


Fig. 10. Encryption/ decryption through Arnoldcat

Inference:

From figure 10 it's found that, post encryption the quality of image is preserved through Arnoldcat mapping algorithm.

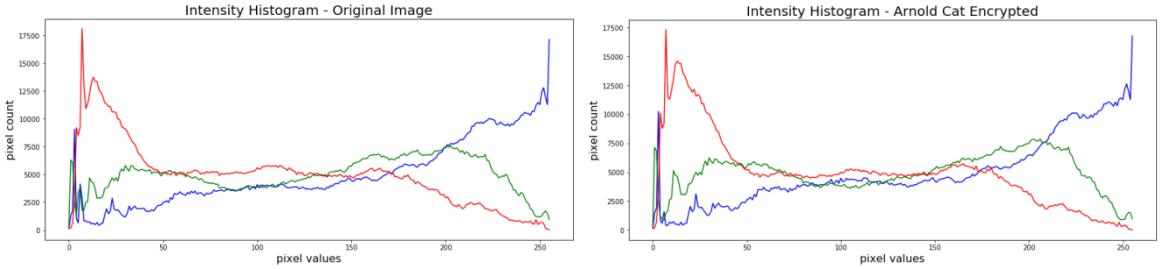


Fig. 11. Intensity histogram pre and post Arnoldcat encryption

Inference:

From the figure 11, it is seen, the pixel intensity of the images are similar which concludes that, the quality of image remains same post encryption/ decryption.

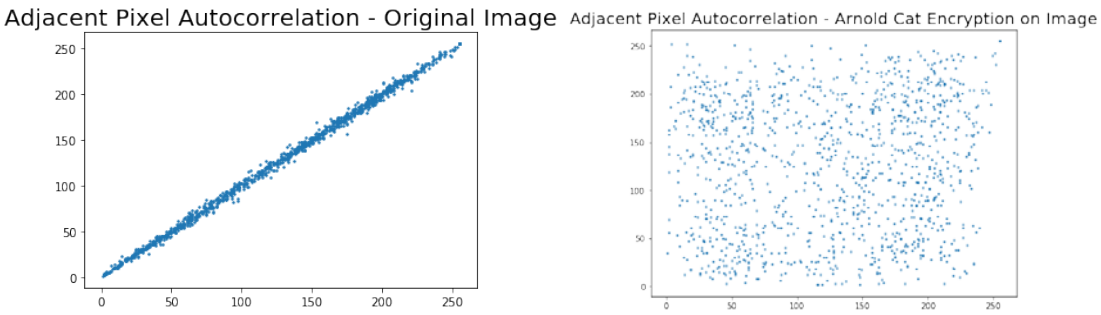


Fig. 12. Analysis of Adjacent pixel auto-correction (APAC)

Inference:

Figure 12 represents that, post applying image ciphering upon the original image, the adjacent pixels broke into several pixels by enhancing the cipher robust.

Thus, Arnoldcat mapping is found effective in the model developed.

6.4 Playfair Ciphering

The Playfair algorithm uses ciphering technique that makes the encryption level complex and more robust.

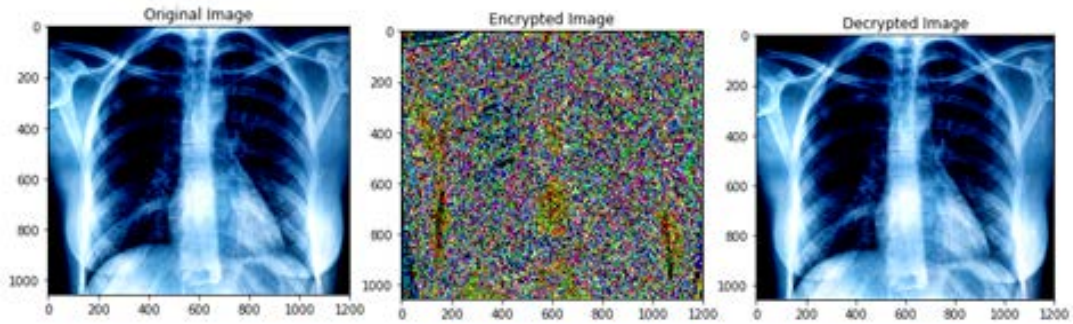


Fig. 13.Encryption/ decryption through Playfair

Inference:

Figure 13 shows, the encrypted image from original image is transformed byte-by-byte where the pixels are broke into complex blocks that scrambles the image completely; it is found, post encryption the quality of image is preserved through decryption.

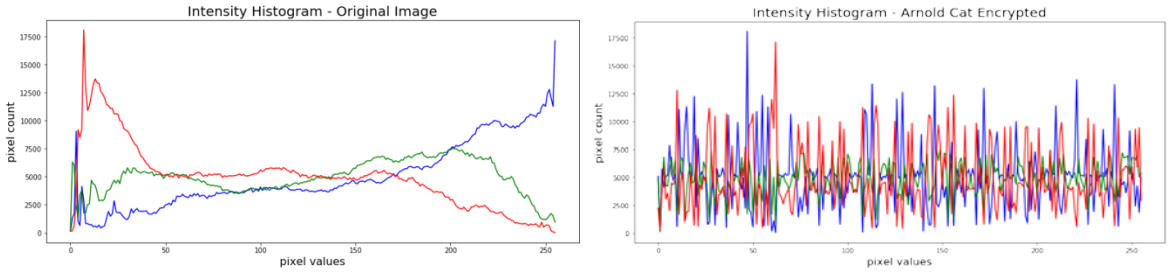


Fig. 14. Intensity histogram pre and post Playfair and Arnoldcat encryption

Inference:

The histogram graph (figure 14) shows vast pixel intensity post Arnoldcat and Playfair encryption from original image pixel intensity; it is thus noted that, Playfair breaks the image pixels into complex pixels that cannot be put-together by third-parties without original value/ image.

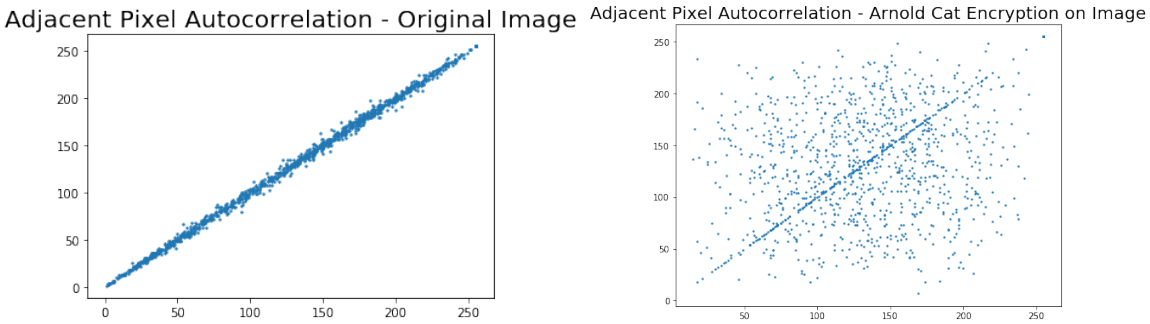


Fig. 15. APAC Analysis

Inference:

Figure 15 represents the pixels of original image is broken into several pixels and few adjacent pixels remained in same position. Thus the image is secured and its security is enhanced through double-level encryption.

The sample datasets are trained and to enhance the quality of images through the developed model; remaining datasets are passed for testing, through the auto-encoder model for encryption/decryption and the outcomes are later compared with original image value with the MSE loss and PSNR values.Thus both Arnoldcat and Playfair algorithms are found effective as double-level encryption that enhances the digital image security in clouds through developed model.

7. Evaluation Metrics

The evaluation metrics for research developed is calculated through estimating MSE (Mean Squared Error) loss and by using the obtained MSE results to calculate the PSNR (Peak signal-to-noise ratio) values. The image's quality, pre and post application of the developed algorithms through obtained outcomes, against existing similar models, is evaluated through PSNR. In PSNR, the ratios of original images (noiseless) with obtained decompressed/ deconstructed images are compared, post de-noising using the de-noising algorithm. The obtained PSNR value should be higher than the original image value to gain better quality of the images, post deconstructing. Thus higher the signal ratio, the better images' quality will be.

The error loss is estimated through, the following equation:

$$MSE = \frac{1}{ab}\sum_{x=0}^{a-1}\sum_{y=0}^{b-1}[M(x,y) - N(x,y)]^2 \tag{6}$$

To estimate the PSNR value, the following equation is applied and results are obtained:

$$\begin{aligned} PSNR &= 10.\log_{10}\left(\frac{MAX_M^2}{MSE}\right) \\ &= 20.\log_{10}\left(\frac{MAX_M}{\sqrt{MSE}}\right) \\ &= 20.\log_{10}(MAX_M) - 10.\log_{10}(MSE) \end{aligned} \tag{7}$$

7.1 Results of PSNR

Generally, the maximum and effective PSNR value range between 30 dB to 50 dB in digital image compression and video compression models with 8-bit data, whereas for 16-bit it ranges from 60-80 dB [33]. However, in other transmissions (wireless), the range could lie between 20 dB to 25 dB [34] based on the bit-size of an image processed.

Table 5.PSNR values

Image	**PSNR Value (in dB)**
Auto-encoder Image	67.75125793845174
Median Filter Image	54.20359258619512
Gaussian Filter Image	54.41328453137318
Average Filter Image	54.53644647654686
Bilateral Filter Image	54.69327483885693

Inference:

The model developed (auto-encoder model) provides the best among the other image encryption/ decryption with 60% (67.75125793845174 in dB), where the outcomes of other filters obtained are above 50dB (refer table 5) stating that 'outcomes provided post encryption/decryption has the better quality of images and preserved the original value of the images.

8. Comparative Analysis

The earlier and existing image encryption models and cloud security models have been discussed and compared. The comparative analysis shows (refer to table 6) existing models used ACM method majorly and provided security for digital images.

Table 6. Comparative analysis of existing image encryption models

S. No.	Author	Year	Image encryption Model	Result: PSNR in dB
1	Proposed model	2022	Playfair ciphering and ArnoldCat map	67
2	Markandey et al.,	2014	RDH (Reversible data hiding)	60
3	Abbas	2016	ICA and ACM	9.5
4	Masood et al.,	2021	Lightweight Chaos-based encryption	7.7
5	Sarosh et al.,	2022	PWLCM based 3D chaotic system	7.5

The developed model uses both PlayFair and ACM for higher-level security and achieved 60% (67dB) in auto-encoding images and other filtering techniques of images acquired approximately ±47%. Thus the proposed model is found reliable and efficient than existing models (refer to fig 16).

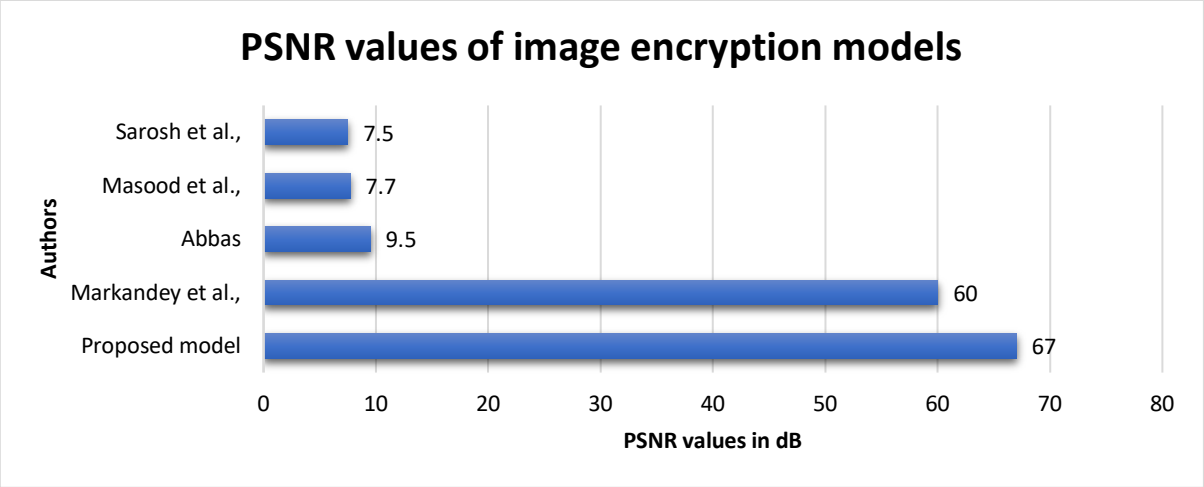


Fig. 16. Comparison of models with PSNR value

Thus, the Playfair ciphering with ArnolCat as chaotic map based model developed achieved higher PSNR value (67.75125793845174dB) as outcome where the quality of image is retained and so the developed model is a success. Though other models used ArnoldCat map the usage of PlayFair ciphering differs from other symmetric encryption and biometric encryption models.

9. Discussion

The image encryption models by authors [35] with RDH (Reversible data hiding) algorithm achieved 60dB in PSNR value. Whereas other studies, like Abbas in 2016 (ICA and ACM model) achieved just 9.5dB, [36] through their developed model of Lightweight Chaos-based encryption achieved 7.7dB and authors Sarosh et al., in 2022 through PWLCM based 3D chaotic system obtained 7.5dB based PSNR values. However, major existing models used chaos based system in encrypting images and ArnoldCat as mapping algorithm. The PlayFair ciphering (digraph) method is user friendly, easily calculable, rapid and needs no additional equipment [37].

It uses 5x5grid where the text, pixel, input is coded, swapped and represented as encrypted message. In each splits 'Z' is added at end-of-coded message and a bogus (counterfeit) letter is added if the digraph pair is incomplete and the encryption is done. In image encryption the same concept is adapted where the pixels of the images are the inputs and the modified/ enhanced PlayFair ciphering algorithm for image processing is utilized. The developed model with ACM mapping and PlayFair ciphering obtained 67db PSNR value as the outcome. according to [20] an outcome pf PSNR value that is higher than 50 is considered as better outcome and those values that are higher than 60 are considered as good outcomes that retains the image quality and originality more. Thus the developed model acquired higher PSNR value, where it uses PlayFair and ACM.

10. Conclusion

Developed research had adopted medical images as datasets from Kaggle, where 2482 images (as large datasets) were acquired and processed. The image encryption/decryption model developed with auto-encoder is found effective with 60% PSNR value and '0' MSE value stating that quality of the images post encryption/decryption is higher with 0 loss. Thus the aim to enhance the security of cloud services in storing and securing medical images (CT-Scans) is enhanced by segmenting images through K-Means clustering and classification through Random Forest as the regressor technique. Post segmenting and classifying the auto-encoder model with Arnold mapping and Playfair ciphering algorithms is implemented to encrypt/ decrypt datasets. Gaussian, Median, Bilateral and Average filters were used in the study to compare the outcomes with the developed auto-encoder model. The MSE loss and PSNR is applied to validate and evaluate the model's performance and it's found that, model is effective and secures data efficiently.

Results obtained from analysis shows the average PSNR value for the developed model is 57.11957127, which is higher than 50dB. Similarly, MSE ('0') value being lesser states that the developed model is effective and produces better quality based outcomes in image processing (coding/decoding) and with secure image storing (cloud).

10.1. Limitations

The research developed an image encryption model in cloud. Thus models of other criteria are not studied. Similarly, author used PSNR value since the recent studies adapt SSIM (Structural-Similarity-Index) analysis to differ from other researches and analyses in image processing.

The study is limited with medical field and other sectors would not be considered. The ciphering technique with Arnoldcat mapping in chaos-based systems is focused and thus other algorithms will be ignored.

10.2. Future enhancements

In future, different filter techniques and regression techniques could be used to test and train the same datasets to observe the variations in outcomes and accuracy of model. Similarly different datasets with same techniques and filters could be tested and trained for larger comparison. In future the model could be adapted with SSIM analysis for examining the difference and comparing the similarities and variations of the adopted techniques. The higher the outcome with different datasets and techniques, the higher the model's accuracy will be. Thus based on reliability and consistency of the model, enhancements and modification of the developed model will be proposed in future.

References

[1] He. C.G, Fan. X.M and Li. Y, (2013), "Toward ubiquitous healthcare services with a novel efficient cloud platform", IEEE trans. Biomed Eng., 60(1), 230-234.

[2] Hallett. Sh, Parr. G, McClean. S, McConnell. A and Majeed. B, "Cloud-based Healthcare: Towards a SLA Compliant Network Aware Solution for Medical Image Processing", In Cloud Computing-2012: The 3rd International Conference on Cloud Computing, GRIDs, and Virtualization, 219-223.

[3] Mirarab. A, Fard. G. N and Shamsi. M, (2014), "A cloud solution for medical image processing", Int. Journal of Engineering Research and Applications, 4(7.3), 74-82.

[4] Peng. C, and Jiangb. Z, (2011), "Building a Cloud Storage Service System", 2nd Edition, 691-696.

[5] Arka. I.H and Chellappan. K, (2014), "Collaborative Compressed I-Cloud Medical Image Storage with Decompress Viewer", Procedia – ScienceDirect, 42, 114-121.

[6] QingZang. H, Lei. Y, MingYuan. Y, FuLi. W and RongHua. L, (2011), "Medical Information Integration Based Cloud Computing,"Network Computing and Information Security (NCIS)", 2011 International Conference on Cloud Computing-IEEE, 1, 79-83.

[7] Mahjoub. M, Mdhaffar. A, Halima. R.B and Jmaiel. M, (2011), "A Comparative Study of the Current Cloud Computing Technologies and Offers", 1st International Symposium on Network Cloud Computing and Applications (NCCA), 131-134.

[8] Kumar. M, Aggarwal. A and Garg. A, (2014), "A review on various digital image encryption techniques and security criteria", International Journal of Computer Applications (0975-8887), 96(13), 19-28.

[9] Prabha. R.J and Prabakaran. S, (2019), "Security in Cloud Health Care", International Journal of Recent Technology and Engineering (IJRTE), 8(4), 6164-6171.

[10] Raval. D and Jangale. S, (2016), "Cloud based information security and privacy in healthcare", International Journal of Computer Applications (0975-8887), 150(4), 11-15.

[11] Shini .S.G, Thomas. T and Chithraranjan. K, (2012), "Cloud Based Medical Image Exchange-Security Challenges", Procedia Engineering - SciVerse ScienceDirect, 38, 3454-3461.
[12] Ogbodo. I.A. and Bakpo. F.S, (2020), "Patient-Centric Cloud-Based EHR System for Government Hospitals in Developing Countries", IOSR Journal of Computer Engineering (IOSR-JCE), 22(3.4): 53-61.
[13] Abbas. N. A-M, (2016), "Image encryption based on Independent Component Analysis and Arnold's Cat Map", Egyptian Informatics Journal, 17: 139-146.
[14] Thilagavathy. R and Murugan. A, (2018), "Cloud storage security scheme for image encryption using modified morse and zigzag pattern", International Journal of Engineering and Technology, 7(4): 6290-6293.
[15] Ratna. A.A.P, Surya. F.T, Husna. D, Purnama. I.K.E, Nurtanio. I, et al., (2021), "Chaos-Based Image Encryption Using Arnold's Cat Map Confusion and Henon Map Diffusion", Advances in Science, Technology and Engineering Systems Journal, 6(1): 316-326.
[16] Kakkad, V, Patel. M and Shah. M, (2019), "Biometric authentication and image encryption for image security in cloud framework", Multiscale and Multidisciplinary Modeling, Experiments and Design, 2: 233-248.
[17] Marwan. M, Kartit. A and Ouahmane. H, (2017), "A Cloud-based Framework to Secure Medical Image Processing", Journal of Mobile Multimedia, 14(3), 319-344.
[18] Marwan. M, Kartit. A and Ouahmane. H, (2018), "Security Enhancement in Healthcare Cloud using Machine Learning", Procedia - ScienceDirect, 127, 388-397.
[19] Parameswari. C.D and Shankar. K, (2019), "Medical Image Security - the State-of-the-Art", International Journal of Recent Technology and Engineering (IJRTE), 8(4S2), 542-545.
[20] Al-Issa. Y, Ottom. M.A and Tamrawi. A, (2019), "eHealth Cloud Security Challenges: A Survey", Hindawi -Journal of Healthcare Engineering, 7516035, 1-15.
[21] Marwan. M, AlShahwan. F, Sifou, F, Kartit. A and Ouahmane. H, (2019), "Improving the Security of Cloud-based Medical Image Storage", Engineering Letters, 27(1), 1-19.
[22] Abdul-Majeed A.O, (2014), "Chaotic scheme for image encryption based on Arnold Cat's map", International Journal of Computer Science and Information Security, 12(3), 26-33.
[23] Hariyanto. E and Rahim. R, (2013), "Arnold's Cat Map Algorithm in Digital Image Encryption", International Journal of Science and Research (IJSR), 6(14), 1363-1365.
[24] Chowdhary. C.L, Patel. P.V, Kathrotia. K.J, et al., (2020), "Analytical Study of Hybrid Techniques for Image Encryption and Decryption", Sensors, 20(5162), 1-18.
[25] Ye. R and Zhao. H, (2012), "An Efficient Chaos-based Image Encryption Scheme Using Affine Modular Maps", I. J. Computer Network and Information Security, 2012(7): 41-50.
[26] Khalil. M.I, (2017), "Medical Image Steganography: Study of Medical Image Quality Degradation when Embedding Data in the Frequency Domain", I. J. Computer Network and Information Security, 2017(2): 22-28.
[27] Bhogal. R.S, Li. B, Gale. A and Chen. Y, (2018), "Medical Image Encryption using Chaotic Map Improved Advanced Encryption Standard", I.J. Information Technology and Computer Science, 2018(8): 1-10.
[28] Mehrnahad. Z and Latif.A-M, (2019), "A Novel Image Encryption Scheme Based on Reversible Cellular Automata and Chaos", I.J. Information Technology and Computer Science, 2019(11): 15-23.
[29] Augustyn. D. R, Wycislik. L and Sojka. M, (2021), "The Cloud-Enabled Architecture of the Clinical Data Repository in Poland", Sustainability, 13(14050): 1-17.
[30] Datta. D, Mittal. D, Mathew. N. P and Sairabanu. J, (2020), "Comparison of performance of parallel computation of CPU cores on CNN model", In Proc. 2020 Int. Conf. on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, Piscataway: IEEE, 1-8.
[31] Liang. Z, Qin. Q, Zhou. C, Wang. N, Xu. Y and Zhou. W, (2021), "Medical image encryption algorithm based on a new five-dimensional three-leaf chaotic system and genetic operation", PLoS ONE, 16(11-e0260014): 1-32.
[32] Sarosh. P, Parah. S. A and Bhat. G. M, (2021), "An efficient image encryption scheme for healthcare applications", Multimedia Tools and Applications, 81:7253-7270.
[33] Sara. U, Akter. M and Uddin. M.S, (2019), "Image quality assessment through FSIM, SSIM, MSE and PSNR – A Comparative study", Journal of Computer and Communications, 7(3), 8-18.
[34] Deshpande, R.G., Ragha, L.L. and Sharma, S.K, (2018) "Video Quality Assessment through PSNR Estimation for Different Compression Standards", Indonesian Journal of Electrical Engineering and Computer Science, 11, 918-924.
[35] Markandey. A, Moghe. S, Bhute. Y and Honale. S, (2014), "An image encryption mechanism for data security in clouds", In: 2014 IEEE Global Humanitarian Technology Conference - South Asia Satellite (GHTC-SAS), pp: 227-231.
[36] Massod. F, Driss. M, Boulila. W, Ahmad. J, Rehman. S. U, et al., (2021), "A Lightweight Chaos‑Based Medical Image Encryption Scheme Using Random Shuffling and XOR Operations", Wireless Personal Communications, Springer Publications, 10(07), 1-28.
[37] Hamad. S. H, Khalifa. A, Elhadad. A. A and Rida. S.Z., (2014), "A modified Playfair cipher for encrypting digital images", Journal of Computers and Communication Engineering, 3(2): 1-9.

Authors' Profiles

Chandra Shekhar Tiwari has received Bachelor in Engg and Master of Engg in Computer Science & Engg.. Currently, he scholar in department of Computer Science of Engg from Birla institute of Technology, Mesra, Ranchi (Jharkhand), India.

Dr.Vijay Kumar Jha, Associate Professor, at BIT Mesra (Ranchi) Computer Science and Engg department. He has completed Ph.D.Engg, M.Sc Engg, B.E. His research area Big data Analysis, Data mining and Network security.

How to cite this paper: Chandra Shekhar Tiwari, Vijay Kumar Jha, "Enhancing Security of Medical Image Data in the Cloud Using Machine Learning Technique", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.14, No.4, pp. 13-31, 2022. DOI:10.5815/ijigsp.2022.04.02