UWS UNIVERSITY OF THE WEST of SCOTLAND

School of Computing, Engineering & Physical Sciences

Computing and Information Systems Journal

Vol 22, No 3, 2018

Edited by Abel Usoro



www.uws.ac.uk

© University of the West of Scotland, 2018

All authors of articles published in this journal are entitled to copy or republish their own work in other journals or conferences. Permission is hereby granted to others for the publication of attributed extracts, quotations and citations of material from this journal. No other mode of publication or copying of any part of this publication is permitted without the explicit permission of the University.

Computing and Information Systems is published normally three times per year, in February, May, and October, by the University of the West of Scotland.

The editorial address is Dr Abel Usoro, School of Computing, Engineering & Physical Sciences, University of the West of Scotland, High Street, Paisley PA1 2BE, UK.
Tel (+44) 141 848 3959; fax (+44) 141 848 3542, e-mail: cis@uws.ac.uk or abel.usoro@uws.ac.uk

Editorial Policy

Computing and Information Systems offers an opportunity for the development of novel approaches, and the reinterpretation and further development of traditional methodologies taking into account the rate of change in computing technology, and its usage and impact in organisations.

Computing and Information Systems welcomes articles and short communications in a range of disciplines:

- Organisational Information Systems
- Computational Intelligence
- E-Business
- Knowledge and Information Management
- Interactive and Strategic Systems
- Engineering
- E-Learning
- Cloud Computing
- Computing Science

The website for Computing and Information Systems is http://cis.uws.ac.uk

An Improved Image Scrambling Algorithm Using $\{2^n -1, 2^n, 2^n +1\}$

Damilare Peter Oyinloye and Kazeem Alagbe Gbolagade
Department of Computer Science
Kwara State University, Ilorin, Nigeria
damilare.oyinloye@kwasu.edu.ng and kazeem.gbolagade@kwasu.edu.ng

ABSTRACT

Purpose: The purpose of this work is to implement an improved image scrambling algorithm based on length three moduli set to achieve a highly reduced size (size on hard disk) of the encrypted or (moduli) image(s) and a more enhanced secured image.

Methodology: This Scheme involves Pixel Scrambling by random number generation and a second layer encryption with Residue Number System (RNS) Forward Conversion with respect to the moduli set $\{2^n-1, 2^n, 2^n +1\}$ for the encryption stage. The decryption process starts with Reverse Conversion with respect to the moduli set $\{2^n-1, 2^n, 2^n +1\}$ to decode the pixel and Pixel Unscrambling by the saved sequence from the random number generator.

Finding/Results: This scheme achieves an enhanced image encryption process and a more efficient decryption process without loses of any important information of the recovered plain image and about 92% compression ratio of the decrypted image. Also this scheme is efficient in internal memory allocation. It produces cipher images that is at least 2/3 the size of the original image. This saves a minimum 25% of internal memory allocation. Thus a 512 x512 plain image that requires about 262144 bytes of memory results in a 768 x 256 encrypted image requiring just 196608 bytes of memory and saving 65536 of internal memory that would have been allocated.

Research Limitation: The scheme is suitable for both colored and grayscale images but it produces a better decrypted image (in terms of picture quality) with grayscale images compared to colored images.

Originality/Value: This research work is valuable because it combines random scrambling algorithm with Residue Number System (a not-so-conventional number system that operates a zero carry system)

Keywords*: Residue Number System, MRC, Information Security, Encryption, RSA, Forward Conversion, Backward Conversion*

Paper type: *Research paper.*

1 Introduction

The security of information is a major concern in every facet of life. The increasing usage of internet and networking technologies has greatly increased the threat to both information and data. Images have found usage in diverse areas such as in medicine, military, science, engineering, art, entertainment, advertising, and education. With the high usage of digital and sophisticated methods for and storing and transfer of images, the issue of protecting the authenticity, integrity and confidentiality of images has become a great concern. Encryption of images is simply coding or protecting the intelligible element of an image which automatically turns the image to something that cannot be easily recognized or decoded by anyone except the recipient of such images. Several image encryption and scrambling methods have been explored to improve the security hidden and encrypted data (Wang, Chang, Liu, Song, and Liu (2015). Image scrambling methods usually scramble the pixel location of digital images in such a way that they become unrecognizable and indistinguishable. (Radu, Cristina, Iustin and Cristina, 2014).

Furthermore, image encryption uses two basic methods: the replacement and scrambling method. The replacement method is changing the value of the pixel in the image while the scrambling method is changing the position of the pixel in image that makes the image chaotic for an intruder. The user who has the key is the only one who can access the content of the original image and can reconstruct the original image from the scrambled or decrypted entity. Scrambling techniques are solely based on permuting coordinates of pixels.

This technique uses the steps below to implement scrambling of an image:

a) build a matrix with the same size of the image to be scrambled (let size of the matrix be s) and every element in it is assigned different natural numbers taking values from 0 to s-1;

b) map the generated matrix to the image matrix each row and column at a time, every element's value in the newly generated matrix is as the coordinate of the corresponding pixel in image matrix;
c) move every pixel to the next position to them; if the mapped pixel coordinate is designated as x, move the pixel into the position with the mapped coordinate (x+1) mod s.

From the methods above, the most important step is to generate a matrix as coordinates shifting path (Rhine and Bhuvan (2015). Rhine and Bhuvan (2015), established that the scrambling method can withstand tougher third party intrusions compared to the replacement method, hence our choice. There are two (2) types of scrambling methods, the one that is based on a 2D matrix transformation and the other based on 2D Arnold transformation (Alhassan and Gbolagade, 2013). Encryption based on 2D Arnold transformation is not suitable for images with different numbers of rows and columns; it increases the data redundancies which automatically increases the disk size of such images and reduces the quality (Alhassan and Gbolagade, 2013).

This scheme will implement a pixel scrambling algorithm) with moduli sets **{2^n-1, 2^n, 2^n +1}** using both forward and reverse conversions) to achieve an enhanced image encryption process and a more efficient decryption process without lost of any inherent information of the recovered plain image.

2 Literature Review

2.1 Residue Number System (RNS)

Residue Number System comprises of a set of moduli that are independent of each other. An integer is represented by the residue of each of the moduli and arithmetic operations are based on residues individually. The advantages of using the RNS over the conversional system include parallelism, modularity, "a carry-free" operation and fault tolerance (Gbolagade and Cotofana, 2008). These features make RNS to be widely acceptable and used in Digital Signal Processing (DSP) applications such as image processing, convolution, digital filtering and fast Fourier transform (Alhassan and Gbolagade, 2013).

Let $\{m_1, m_2, m_{3,\ldots,}m_n\}$ be a set of positive integers all greater than 1. m_i is called a modulus, and then n-tuple set $\{m_1,m_2, m_{3,\ldots,}m_n\}$ is called a moduli set. Consider an integer number Y. For each of the modulus in $\{ m_1, m_2, m_{3,\ldots,}m_n \}$, we have $y_i = Y \bmod m_i$, (denoted as $|Y|m_i$). Thus the number Y in this system is represented as $Y = (y_1, y_2, y_{3,\ldots,}y_n)$,$0 \leq y_i < m_i$. Siewobr, Gbolagade and Cotofana (2014).

For example, take the moduli set {7,8,9}, the number 150 can is depicted in RNS as:

$$y_1 = |Y|_{m1} = |150|_7 = 3$$
$$y_2 = |Y|_{m2} = |150|_8 = 6$$
$$y_3 = |Y|_{m3} = |150|_9 = 6$$

Thus, the RNS representation of 150 is $(3,6,6)_{RHS(7,8,9)}$.

In order to avoid redundancy, moduli set must be pair wise relatively prime. Thus gcd (m_i ,m_j) = 1 for $i \neq j$, where gcd is the greatest common divisor of m_i ,m_j.

Let $M = \prod_{i=1}^{n} m_i$, then the RNS representation is unique for any integer $Y \in [0, M-1]$ M is called the dynamic range. Siewobr and Gbolagade (2012).

2.2 Forward Conversion

Forward conversion is done by a forward converter which decomposes a weighted binary number into a residue represented number with regards to a moduli set. It is the conversion from a conventional representation to a residue representation by dividing the number X by each of the given moduli and then collecting their remainder (Gbolagade and Cotofana, 2008).

For example, given moduli set {3, 4, 5}, 4 is depicted in RNS as:

$$x_i = |X|_{m_i} \ldots \ldots (2.1)$$

$$x_1 = |X|_{m_1} = |4|_3 = 1;\ x_2 = |X|_{m_2} = |4|_4 = 0;\ x_3 = |X|_{m_3} = |4|_5 = 4$$

Therefore, the depiction of 4 in RNS is $(1, 0, 4)_{RNS(3,4,5)}$. (Baagyere, Boateng and Gbolagade, 2011).

2.3 Reverse Conversion

The conversion from residue to a conventional representation (either binary or decimal representation) is known as reverse conversion. The two most widely used techniques of reverse conversion are the Mixed Radix Conversion (MRC) and Chinese Remainder Theorem (CRT) and Gbolagade, 2009; Gbolagade 2009).

2.4 Chinese Remainder Theorem

For a moduli set $\{m_1, m_2, \ldots, m_k\}$ with $\gcd(m_i, m_j) = 1 \text{ for } i \neq j$ and a dynamic

range $M = \prod_{i=1}^{k} m_i$, the residue number $(x_1, x_2, \ldots, x_k)$ can be converted into the decimal number X if the moduli set are co-prime (Salifu and Gbolagade, 2016; Gbolagade, 2009).

$$X = \left| \sum_{i=1}^{\kappa} m_i \left| m_i^{-1} x_i \right|_{mi} \right|_m \quad \ldots\ldots (2.2)$$

$M = \prod_{i=1}^{n} m_i$ and $M_i = \frac{M}{m_i}$ and $M_i^{-1}$is the multiplicative inverse of $M_i$with respect to m_i.

2.5 RNS to Weighted Conversion

A (R/D) Residue to Decimal converter (decoder) is required to convert from a RNS to decimal. The two methods used to convert RNS to weighted systems are the Mixed Radix Conversion (MRC) and the Chinese Remainder Theorem (CRT).

The CRT is employed in this research (Bankas and Gbolagade, 2014).

Given $\{m_1, m_2, m_{3,\ldots,} m_n\}$ which $\gcd(m_i, m_j)=1$ for $i \neq j$ and dynamic range $M = \prod_{i=1}^{n} m_i$, by the CRT, an integer labeled Y whose depiction in RNS is $(y_1, y_2, y_{3\ldots\ldots} y_n)$ can be converted from its residue as

$$Y = \left| \sum_{i=1}^{n} M_i \left| M_i^{-1} y_i \right|_{m_i} \right|_M \quad (2.3)$$

Where $M_i = \frac{M}{m_i}$ and M_i^{-1} is the multiplicative inverse of M_i with respect to m_i.

Considering {7, 8, 9) as a moduli set,

10 = (3,2,1) M= 7 x 8 x 9 = 504

$$M_1 = \frac{504}{7} = 72, \quad M_2 = \frac{504}{8} = 63,$$

$$M_3 = \frac{504}{9} = 56$$

$$M_1^{-1} = 72^{-1}, M_2^{-1} = 63^{-1} = 7,$$

$$M_3^{-1} = 56^{-1} = 5$$

Therefore, by the CRT

$$Y = |(72 \times 5) + (72 \times 6) + (56 \times 5)|_{504}$$

$$Y = |360 + 378 + 280|_{504}$$

$$Y = |1018|_{504}$$

$$Y = 10$$

Hence:

$$Y = |(72 \times |4 \times 3|_7) + (72 \times |7 \times 2|_8) + (56 \times |5 \times 1|_9)|_{504}$$

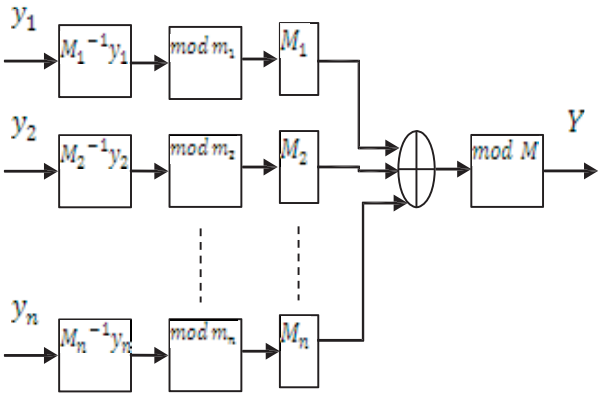


Figure 2.1 (Diagrammatical depiction of CRT)

3 Methodology

The two main parts of the methodology are encryption algorithm and decryption algorithm as explained in the rest of this section.

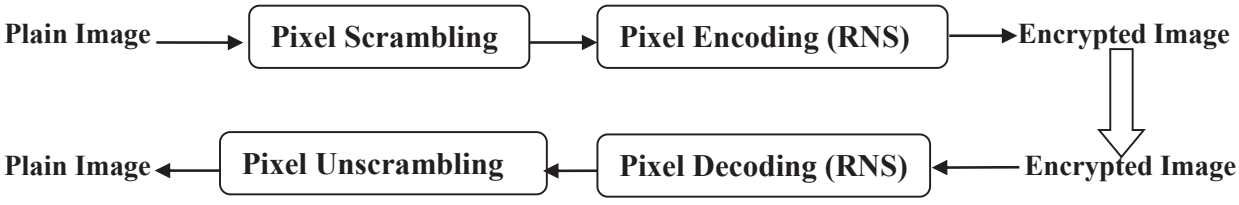


Figure 3.1 (Block diagram of the scheme)

3.1 Encryption Algorithm

(i) Pixel Scrambling (Scrambling Algorithm)

1. Original Image Input
2. Find number of rows and column (in total)
3. Point to the first pixel in the original image.
4. Let counter be 1.
5. Build new position of the current pixel in the cipher image by generating two random numbers (n1, n2) for row and column.
6. If the new position of cipher image as been generated before, then proceed to step 4, otherwise proceed to step 6.
7. Store n1 (value) in the array K1

8. Store n2 (value) in the array K2.
9. Current pixel of the original image will take the position (n1, n2) in the cipher image.
10. If all the pixels of the original image is exhausted, move to step 13 else proceed to step 10
11. Navigate to the next pixel on the original image.
12. Counter increased by 1
13. Proceed to 5
14. Terminate

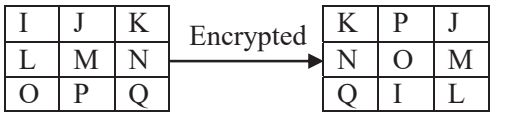


Figure 3.2 Encryption process of a 3 x 3 matrix image

(ii) Pixel Encoding (Forward Conversion)

Taking {9,8,7} as moduli set 2^n-1, 2^n, 2^n+1 (n=3)

m_1=9 m_2=8 m_3=7

128	108	66
151	113	179
139	120	132

Step 1: Generate Moduli images

m_1=9

2	0	3
7	5	8
5	0	4

m_2=8

0	4	2
7	1	3
3	0	4

m_3=7

2	3	3
4	1	4
6	1	6

Step 2: Send Moduli Images

3.2 Decryption Algorithm

(i) Pixel Decoding (Reverse Conversion)

$$Y = \left|\sum_{i=1}^{n} M_i |M_i^{-1} y_i|_{m_i}\right|_M$$

Equation 1

N=3

$$X = |M_1|M_1|_{m1} x_1 + M_2|M_2|m_2 x_2 + M_3|M_3|m_3 x_3|M$$

Equation 2

$$M_i = \frac{M}{m_i}$$

Equation 3

M=504

M_1= 56 $|M_1^{-1}|m1 = 5$

M_2=63 $|M_2^{-1}|m2 = 7$

M_3=72 $|M_3^{-1}|m3 = 4$

X_{11}=|56X5X2 + 63X7X0 + 72X4X2|$_{504}$

X_{11}=128 X_{12}=108 X_{13}=66

X_{21}=151 X_{22}=113 X_{23}=179

X_{31}=139 X_{32}=120 X_{33}=132

(ii) Pixel Un-Scrambling (Unscrambling Algorithm)

1. Encrypted image input.
2. Keys K1 & K2 inputed.
3. Let counter be 1
4. Navigate to the first pixel of the cipher image.
5. n1=K1
6. n2=K2
7. Obtain (n1, n2 values) from the cipher image and let it inputted in the current position of deciphered image.
8. If the counter is not the last position of K1 and K2 proceed to 9 else move to 13
9. Increment counter by 1
10. Navigate to the next pixel of the cipher image
11. Proceed to step 5
12. Terminate

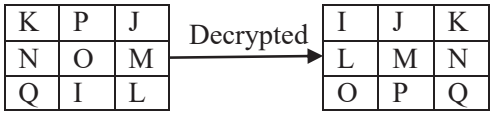


Figure 3.3 (Decryption process of a 3 x 3 matrix image)

4 Result Discussion

4.1 Encryption Analysis

The encryption part of our scheme is based on two different algorithms, the scrambling algorithm and forward conversion (RNS). The scrambling stage has little effect on the size of the image or the pixel value of the images but rather it only changes the position of the pixels which enhances the security of the image.

4.2 Pixel Encoding Analysis

The pixel encoding brought out two distinct results. It drastically reduces the value of pixel and the size of the original image but keeps its coherent attributes. The drastic reduction in the value of pixel automatically speeds up time required for computation which is useful for other image processing processes.

Better put, the size reduction increases the speed of data during transmission over a network, since smaller bits are needed to stand in for the pixels and it also reduces the bandwidth required for transfer. Table 4.1 compares the size of the unencrypted and encrypted images when saved in a Joint Photographers Expert Group format. It is interpreted from the table below that the system obtains a reduction in size of up to 92%.

It is also efficient in memory allocation (Internal). It produces cipher images that is at least 2/3 the size of the original image. This saves a minimum 25% of internal memory allocation. Thus a 512 x512 plain image that requires about 262144 bytes of memory results in a 768 x 256 encrypted image requiring just 196608 bytes of memory and saving 65536 of internal memory that would have been allocated.

Table 4.1: sizes of unencrypted and encrypted images compared.

Image Type	Size(Size on Disk)		
	Unencrypted Image(kb)	Moduli Image (kb)	Compression ratio (%)
Love.jpg grayscale (512x512)	152	20	86.84
Mountain.png (320x301)	163	5	96.93
Easy.bmp (640 x 480)	301	23	92.35

4.3 Time Complexity Analysis

We analyzed the time taken for our scheme to produce an output. The simulations were done a system with 1.86GHz processor and 1024 MB ram. The results obtained are shown in Table 4.2

Table 4.2: Sizes of Original and encrypted images compared

Image details		Encryption algorithm		Decryption algorithm	
Image	Number of elements	Scrambling (seconds)	Encoding (seconds)	Decoding (seconds)	Unscrambling (seconds)
Lena(512 x512)	262,144	0.1100	7.5218	9.4065	0.1010
Checkerboard(256 x256)	196,608	0.2181	4.6122	4.8484	0.2050
Koala(448 x336)	45,184	0.1160	9.8514	11.4281	0.1847

4.4 Memory Requirement Analysis

The memory requirement for both the plain and cipher images are analyzed below. The size of an image before encryption is far higher on memory compared to the size of the cipher image i.e the encrypted image. It is the direct implication of the reduced pixel value implemented at the second encryption stage.

Table 4.3: Memory requirement of plain and encrypted images compared.

Image	Size(disk space)		Memory requirement	
	Plain-image(kb)	Encrypted image (kb)	Plain-image (Bytes)	Encrypted image (Bytes)
Cameraman	64.00	32.00	65536	32768
Peppers	281.00	140.00	589824	14336
Rice	44.00	20.00	65536	20480

4.5 Security Analysis

We analyzed the security strength of the scheme as shown below, the two layer encryption this scheme implements (Random scrambling and Forward Conversion-Residue Number System) correctly repositioned the pixels of the image and while the forward conversion reduced the value of the pixels which makes it difficult to guess the original image from the encrypted and also makes the image smaller in terms of the size on memory.

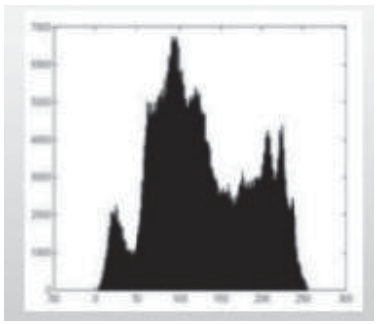


Fig: 4.1 (a) Original Image Histogram

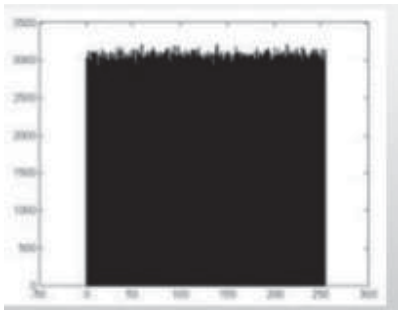


Fig 4.1 (b) Encrypted Image Histogram

Figure 4.1 (a) and (b) are the histograms of the plain image and the encrypted image. It is clear from the images above that you cannot guess or deduce Figure 4.1 (a) (Original image) from Figure 4.1 (b) (Histogram of encrypted image).

The two layer encryption this scheme implements (i.e Random scrambling and Forward Conversion-Residue Number System) has changed the position of the pixels and also the value of the pixels which makes it difficult to guess the original image from the encrypted.

5 Conclusion

This research proposed a simple scrambling algorithm to encrypt and decrypt the images based on random number generation and RNS (Forward and Reverse Conversion). The image is first encrypted by moving the position of each pixel in the image without having to change the pixel value of the image. The original image is read pixel by pixel, row by row and each pixel will assume a new position in the scrambled image. The new position is chosen using random number generation from the random number generator. The key will be generated as a matrix during the encryption process and also the key saves the position of each pixel in the encrypted/scrambled image. The encryption layer transforms the scrambled image to moduli images which automatically adds an extra security layer to our data (Image).

References

Alhassan, S, Gbolagade, K. A. (2013)" Enhancement of the Security of a Digital Image using the Moduli Set 2^n-1, 2^n, 2^n+1**,** *International Journal of Advanced Research in Computer Engineering & Technology, 2*(7): 2223-2229

Baagyere, E. Y., Boateng, K. O. and Gbolagade, K. A. (2011) Bioinformatics: An important Application Area of RNS, *Journal of Engineering and Applied Sciences*, *6*(2): 174 -179.

Bankas E. K. and Gbolagade, K. A. (2014) A New MRC Adder-based Reverse Converter for the Moduli Set {2^n, 2^{2n+1}-1, 2^{2n-2}-1}, *The Computer Journal* (Oxford), 12, 1-7.

Gbolagade, K. A. (2009) A Shorter Algorithm for Efficient Residue to Decimal Conversion, *Advances in Computer Science and Engineering*, *3*(2), 147-156, ISSN 0973-6999.

Gbolagade, K. A., Cotofana, S. D. (2008) A Residue to Binary Converter for the {2n+2,2n+1,2n} Moduli Set, *Asilomar Conference on Signals, Systems, and Computers*, 1785-1789, California, USA.

Radu, B., Cristina D., Iustin, P. and Cristina, F. (2014) "A New Fast Chaos-Based Image Scrambling Algorithm" 10th international conference on Communication, 1-4.

Rhine, R., Bhuvan, N. (2015) Image Scrambling Methods for Image Hiding: A Survey, *International Journal of Computer Science and Network Security*, 15, 86-91.

Salifu, A. and Gbolagade, K. A. (2016) Mixed Radix Conversion based RSA Encryption System, *International Journal of Computer Application* (IJCA), 150, 43-47.

Siewobr H., Gbolagade, K. A., and Cotofana, S. D. (2014) An Efficient Residue-to-Binary Converter for the New Moduli Set {$2^{n/2}$±1,2^{2n+1},2^n+1}.

International Symposium on Integrated Circuits (ISIC 2014), (to appear), Singapore.

Siewobr H, Gbolagade, K. A. (2012) An Area Efficient RNS-to-Binary Converter for the Moduli Set $\{2^n,2^{2n+1}-1,2^n-1\}$, 4th IEEE International Conference on Adaptive Science and Technology, Kumasi, Ghana, 104-107.

Wang. D., Chang C., Liu Y., Song, G., and Liu, L. (2015) Digital Image Scrambling Algorithm Based on Chaotic Sequence and Decomposition and Recombination of Pixel Values, *International Journal of Network Security*, 17, 322-327.

Development of Iris Biometric Template Security Using Steganography

[1]Saheed Y. Kayode, [2]Abdulsalam S. Olaniyi, [3]Arowolo M. Olaolu, [4]Akinbowale N. Babatunde
[1]Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria.
[1]yksaheed@alhikmah.edu.ng
[2,3,4] Department of Computer Science, Kwara State University Malete, Nigeria.
suleiman.abdulsalam@kwasu.edu.ng, olliray2002@yahoo.com and akinbowale.nathaniel@kwasu.edu.ng

ABSTRACT

Purpose: Traditional iris segmentation methods and strategies regularly contain an exhaustive search of a large parameter space, which is sensitive to noise, time-consuming and no longer secured enough. To address these challenges, this paper proposes a secured iris template.

Approach: This paper proposes a technique to secure the iris template using steganography. The experimental analysis was carried out on matrix laboratory (MATLABR2015A) environment. The segmented iris region was normalized to decrease the dimensional inconsistencies between iris region areas with the aid of the usage of Hough transform (HT). The features of the iris were encoded by convolving the normalized iris region with 1D Log-Gabor filters in order to generate a bit-wise biometric template. Then, least significant bit (LSB) was used to secure the iris template. The Hamming distance was chosen as a matching metric, which gives the measure of how many bits disagreed between the templates of the iris.

Findings: The system operated at a very good training time and high level of conditional testing signifying high optimization with recognition accuracy of 92% and error of 1.7%. The proposed system is reliable, secure and efficient with the computational complexity significantly reduced.

Originality/value: The proposed technique provides an efficient approach for securing the iris template.

Keywords: *Biometric, Iris Recognition System (IRS), Least significant bit (LSB), Hough Transform, 1D Log-Gabor filters.*

Paper type: *Research paper.*

1 Introduction

Biometric techniques have come to be an outstanding option and secured means of authentication and successfully sustaining the emerging and rising ubiquitous computing (Saheed, Hambali, Adeniji and Kadri, 2017). Biometric recognition techniques have unfolds as a promising option for securing valuable information (Hambali and Jimoh, 2015). Instead of requiring identification cards, passwords or keys, biometrics are capable of identifying face, fingerprints, signatures, iris or retinas of an individual for convenient and easy verification (Simranjitkaur and Jasmeen, 2014).

First developed by Dr. John Daugman in the 1990s (Daugman, 1991), who got the idea from Flom and Safir's patented theoretical design, Iris Biometric identification systems have successfully been implemented and can be found in important applications around the world, ranging from offices to log entry in systems, to airports and passport offices (to associate visa details).

A lot of research effort on iris recognition has been done in the last few decades (Daugman, 1991; Iris recognition and identification, n.d, Daugman, 1993).

Boddeti and Kumar (2010) investigated the use of wave front-coded imagery for iris recognition. Daugman (2003) used a multi-scale quadrature approach to extract the texture phase structure information of the iris to produce 2,048-bit (256 bytes) iris code and used Hamming distance for matching. Wildes et al. (1996) presented a Laplacian pyramid for investigating the iris image, by applying normalized correlation and fisher's linear discriminant for matching the iris. This paper proposes a method for iris template security using steganography technique.

The rest of this paper is organized in sections. Section 2 presents the material and methods.

Section 3 is the proposed system. In section 4, the results and discussion is presented; while in section 5, the conclusion is discussed.

2 Material and methods

2.1 Iris Dataset

The system will make use of the CASIA database as a total of 50 subjects will be administered with three (3) images for training and one (1) for testing the effectiveness of the system.

2.1.1 Segmentation

This stage is performed to eliminate the iris template from the eye image. The HT, a standard computer vision algorithm is employed here to get the centre coordinates and radius of the pupil and iris region.

2.1.2 Normalization

This is carried out to remap each pixel within the iris region to a pair of polar coordinates. It is performed based on the Daugman's rubber sheet model. Here, the centre of the pupil is regarded as the radial vector-circle and reference point via the iris region.

2.1.3 Feature extraction and selection

The iris biometric can be examined in several approaches to get its digital template T. In all cases, solutions depend on iris texture filtering, but they can vary significantly. The phase-based approach proposed by Daugman, will be used in this paper to extract relevant and important features.

2.1.4 Training

At this stage the extracted features will be used to train the system so as to formulate a machine based knowledge discovery for the iris pattern.

2.1.5 VC Module

This phase generates a secured iris template.

2.1.6 Least significant bit security

The extracted iris template is saved in the database and the use of built in pixel data, a random number is produced and is saved in the database as the extracted template image's title by the use of a steganography algorithm LSB to encode. This number is now given as input to the VC module that produced two shares, out of which one of the share is save in the database and the other share is save on the user ID card. This process ends the enrollment module.

3 System Architecture

Below is the application flowchart.

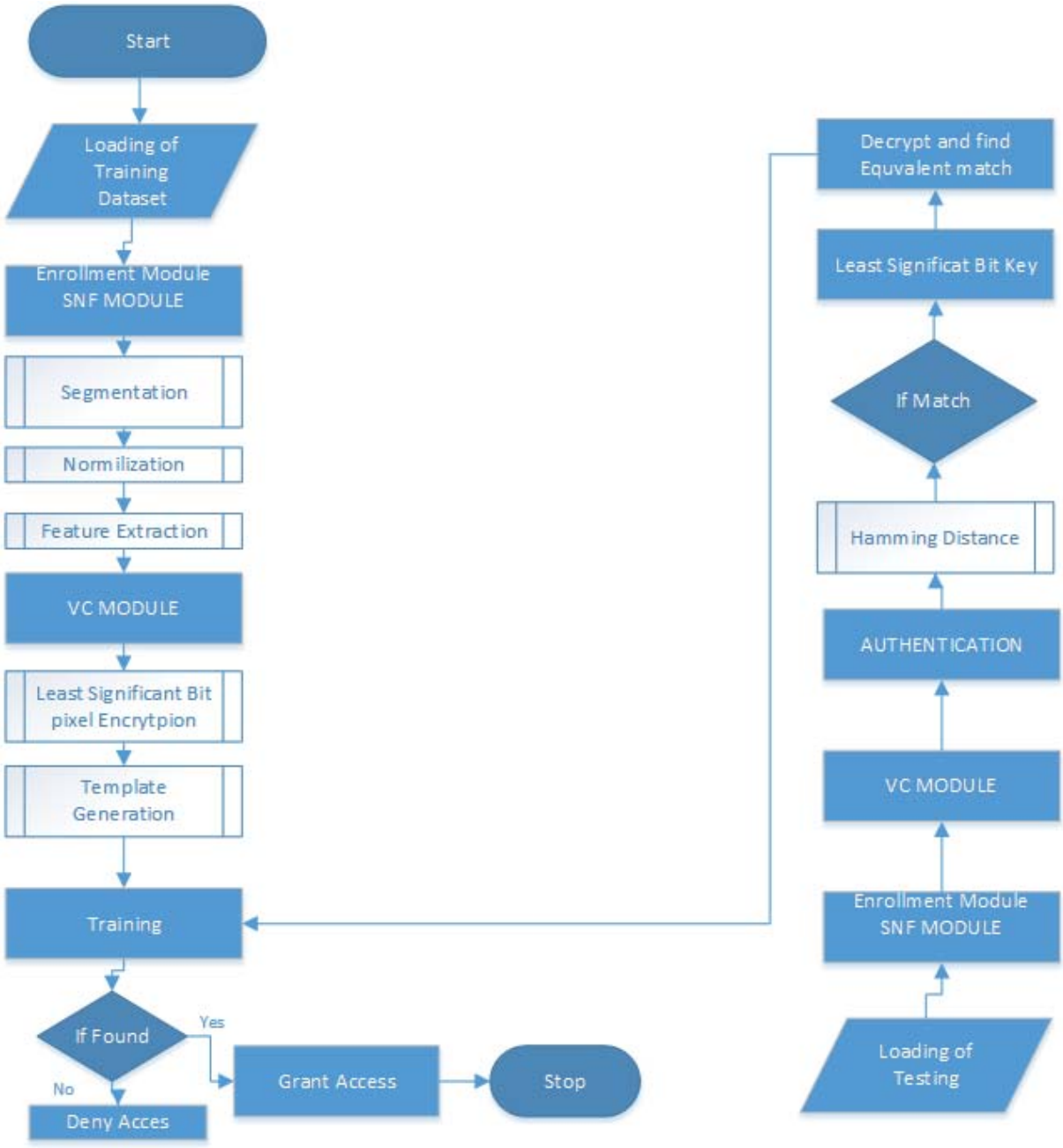


Figure 3.1 System Flow Chart

3.1 Authentication

For matching, hamming distance is the metric chosen for recognition. The result obtained is used as the goodness of match, with smaller values indicating better matches. If two patterns are obtained from the same iris, the hamming distance between them will be close to 0 due to high correlation

3.2 Proposed System

The developed iris recognition system was tested with five databases from the Chinese Academy of Sciences' Institute of Automation (CASIA) Database. Four of the databases contain five samples with a grouping of three for training and two for testing. The fifth database contains two samples to show the response of the system when the selected sample has not been loaded to the system. The system was developed with respect to five Iris samples per class, three of the iris datasets are passed into the training set so as to build an experimental knowledge process for future query. Two subjects of each class was used to test the accuracy of the model as the two subjects are not passed into the training database but set aside for recognition modules.

Table 3.1 Database Parameter Settings

Parameters	Value
Number of training sets	60
Number of Testing Sets	24
Iris Per subject	6
Total Number of Class	10
Training Iris Per Class	6
Testing Iris Per Subject	2
Imposter	4
Total Iris Images	88

For each individual, we make use of six data sets for training, and two data sets which comprises of the left and right iris for testing. The total number of training sets is sixty; testing sets is 24 and it consists of 10 classes. Four of the testing images were imposter that is, the images were not found in the database.

4.0 Results and Discussion

4.1 Feature Extraction

The feature extraction is the generation of matrix that represents each iris image. This stage obtained the following image processing result.

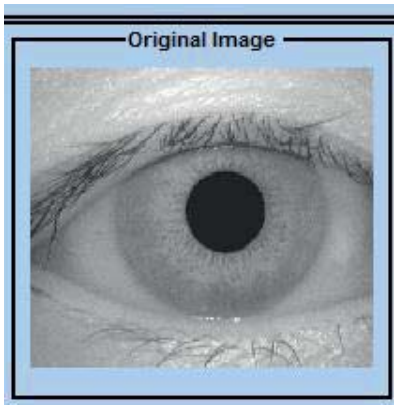


Figure 4.1 Fetching of Original Iris

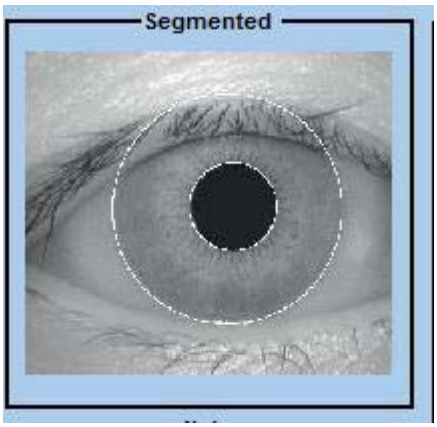


Figure 4.2 Iris Localization

The stage obtained a polarized normalized arrangement as well as reducing noise.

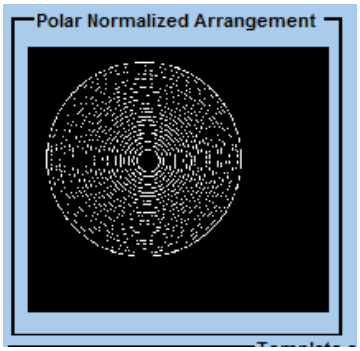


Figure 4.3 Polar Normalized Arrangement

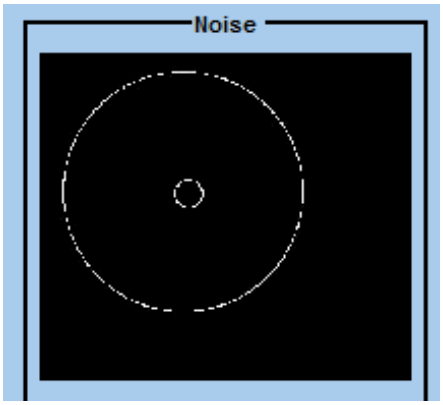


Figure 4.4 Polar Normalized Arrangement



Figure 4.5 Extracted feature template



Figure 4.6 Mask

The template and the mask are the last stage of feature extracted by the doughman integro operator algorithm. The template is encoded to be the cover message and the mask as the message image. The Hamming encoding works on the template and the mask to obtain a unique distance for each of the iris template for the purpose of recognition.

The page shows the different extraction processes, the segmentation, the normalization and the iris template. It shows the processes of feature extraction, the segmentation process, the polar normalization process, the noise removal process and the template of the iris generated.

4.2 LSB Steganography

The Least Significant bit helps to generate a secure template by taking the template image as the cover message and the mask as the message image there by creating a template that encodes the pixel value of the extracted feature. The results of this, is shown in the figure below:

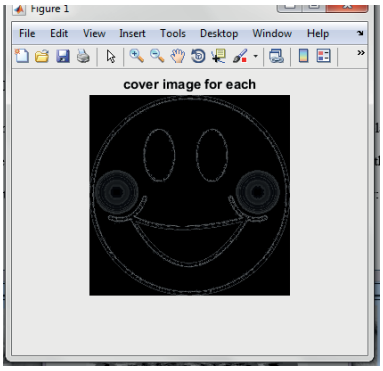


Figure 4.7 Stego Image

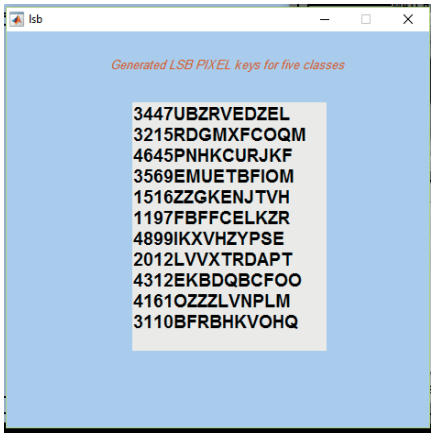


Figure 4.8 LSB Secured Keys

4.3 Trained Samples

The image was trained with the templates generated by the least significant bit cryptography so as to store up a template pattern for a particular iris thereby making it easy to recognize iris from one individual to the other.

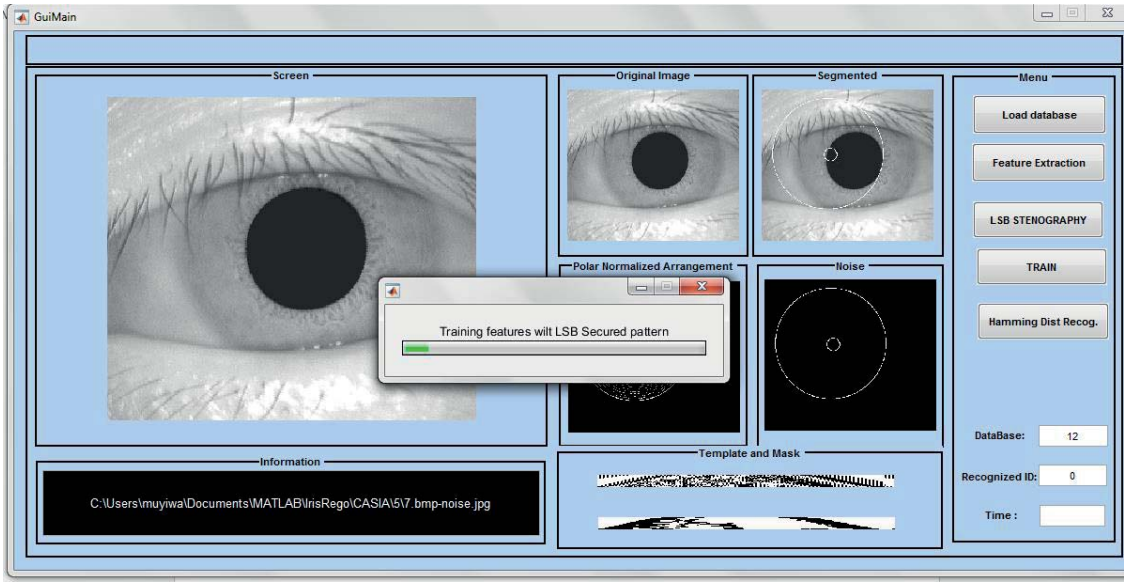


Figure 4.9 Training Sample Interface

4.4 Hamming Distance Recognition

This page shows the result of the testing of the images. The iris image was detected to recognised correctly by correlating the iris probe image with the name of the class and subjecting it to the one located in the database.

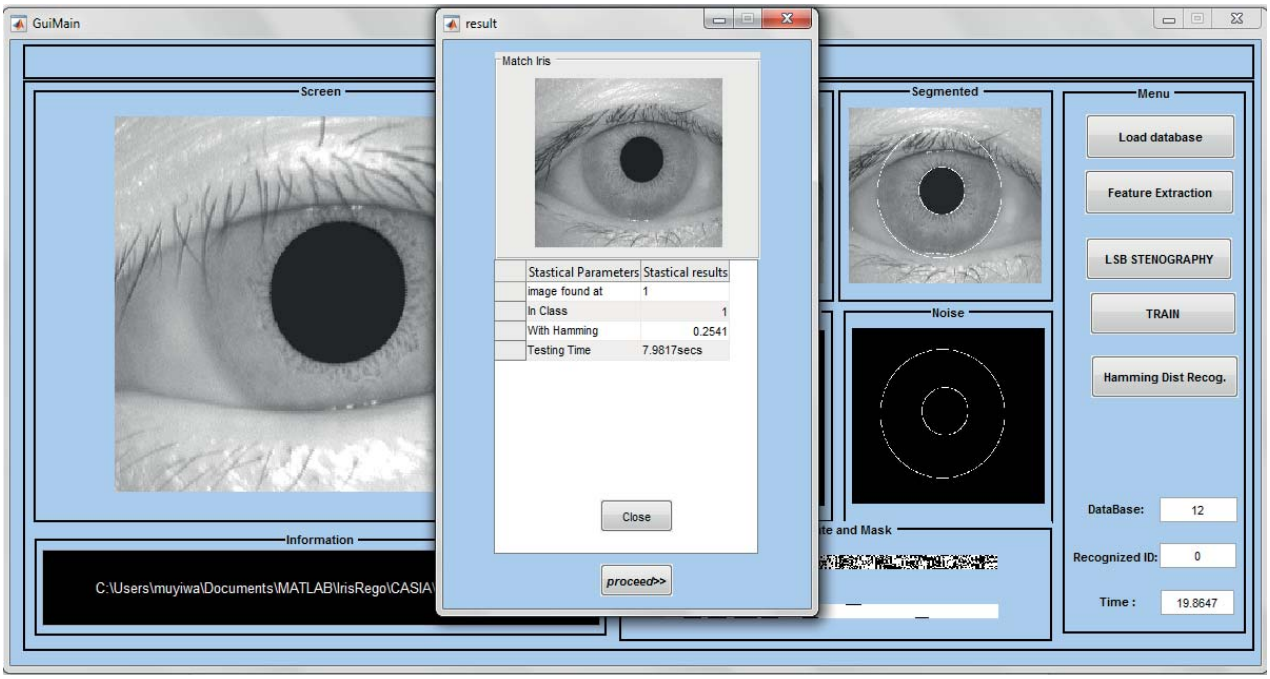


Figure 4.10 Testing Sample Interface

4.5 LSB Authentication Initiation

The least significant bits (LSB) key will be provided here for next stage to be accessed.

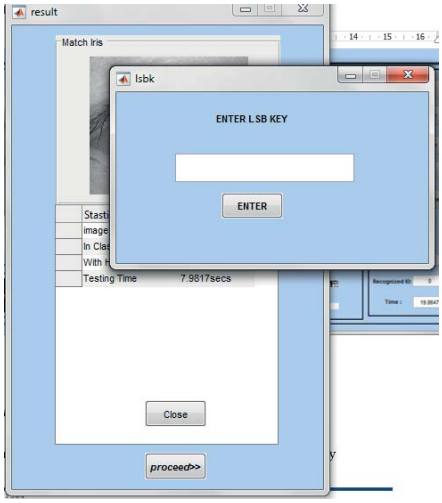


Figure 4.11 LSB Initiation Authentication

4.6 LSB Authentication

The figure below shows when the least significant bit (LSB) key is triggered to search the iris encoded database to find its match.

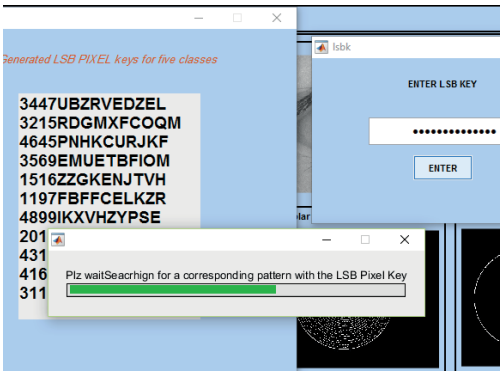


Figure 4.12 LSB Authentication

The figure below shows when the incorrect key is inserted

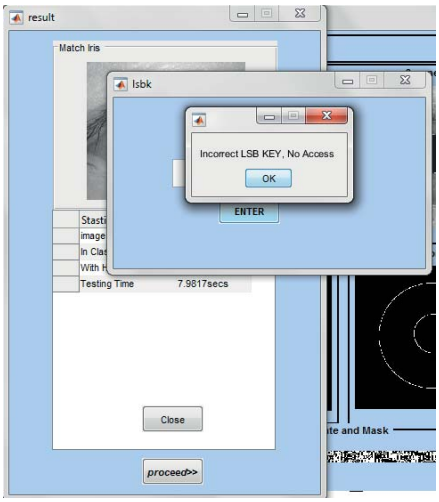


Figure 4.13 Blocked Authentication

The figure below shows when the incorrect key is inserted

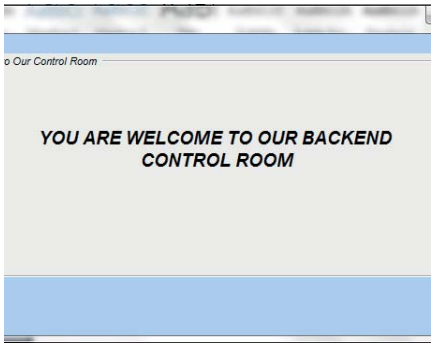


Figure 4.14 Access Grant

4.7 Presentation of recognition results

The recognised results are presented on Tables 4.1 and 4.2 as well as Figure 4.6.

Table 4.1 Hamming Result Evaluation

Serial number	Class	Index of image recognized	Hamming distance	Testing time	Matched
1	1	1	0.2541	5.5567secs	Matched
2	1	3	0.2279	5.4991secs	Matched
3	2	2	0.1347	5.5254secs	Matched
4	2	2	0.1620	5.2367secs	Matched
5	3	1	0.2541	5.3485secs	Matched
6	3	3	0.2279	5.4863secs	Matched
7	4	2	0.1347	6.4831secs	Matched
8	4	2	0.1620	6.388secs	Matched
9	5	2	0.1620	6.3616secs	Matched
10	5	1	0.2541	4.9278secs	Matched
11	6	1	0.2541	4.5416secs	Matched
12	6	1	0.2541	5.4471secs	Matched
13	7	3	0.2279	4.676secs	Matched
14	7	2	0.1347	5.2055secs	Matched
15	8	2	0.1620	5.4618secs	Matched
16	8	1	0.2541	5.4533secs	Matched
17	10	3	0.2279	5.4833secs	Mismatched
18	9	2	0.1347	5.4022secs	Matched
19	10	2	0.1620	5.0313secs	Matched
20	1	2	0.1620	5.1632secs	Matched

Table 4.2 Imposter Result

Serial number	Class	Index of image recognized	Hamming distance	Testing time
21	No Class Found	No Image Found	No Value Found	5.1822secs
22	No Class Found	No Image Found	No Value Found	5.8061secs
23	3	0.1347	0.1347	5.0359secs
24	No Class Found	No Image Found	No Value Found	6.2378secs

The system operated at a very good training time and high level of conditional testing, signifying high optimization as shown in figure 4.6.

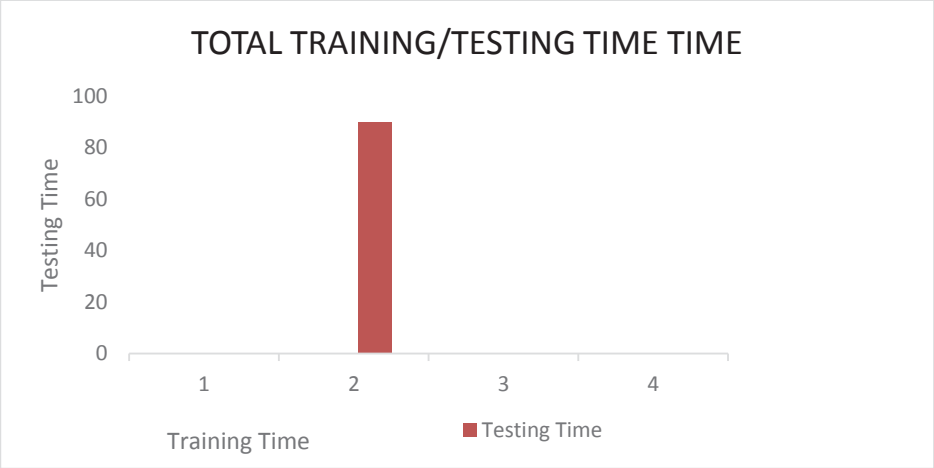


Figure 4.6: Training and testing time

5 Conclusion

This paper proposed an iris biometric system security based on steganography. The LSB algorithm was chosen to secure the iris template. The proposed system operated at a very good training time. The 1 D log-Gabor filter was adopted for feature extraction and Hamming distance was chosen and used as a matching metric. The proposed system is efficient, reliable and provides a better security.

References

Boddeti, V. N. and B. V. K. Vijaya Kumar (2010) Extended-Depth-of-Field Iris Recognition Using Unrestored Wavefront-Coded Imagery *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 40*(3), 495-508. http://www.sinobiometrics.com/resources.htm (Accessed on 15 January 2017).

Daugman, J. (1993) High Confidence Visual Recognition of Persons by a Test of Statistical Independence *IEEE Trans. Pattern Analysis and Machine Intelligence*, *15*(11), 1148-1161.

Daugman, J. (2003) Demodulation by Complex-Valued Wavelets for Stochastic Pattern Recognition, *Int'l J. Wavelets, Multiresolution and Information Processing*, 1(1), 1-17.

Daugman, J. (2004) "How Iris Recognition Works", *IEEE Transactions on Circuits and Systems for Video Technology, 14*(1), 21-30.

Daugman, John (1991) Biometric Personal Identification System Based on Iris Analysis,

https://patents.google.com/patent/US5291560A/en (Accessed on 17 January 2017).

Hambali, M. A. and Jimoh, R. G. (2015) Performance Evaluation of Principal Component Analysis and Independent Component Analysis Algorithms for Facial Recognition, *Journal of Advances in Scientific Research& Its Application (JASRA)*, 47–62.

Iris Recognition and Identification, Retrieved 30th July 2017 from http://www.iris-recognition.org/, 2017 (accessed on 17 January 2017).

Saheed, Y. K., Hambali, M. A., Adeniji, I. A. and Kadri, A. F. (2017) Fingerprint Based Approach for Examination Clearance in Higher Institutions *FUOYE Journal of Engineering and Technology,* 2(1), March, 47-50.

Simranjit, Kaurand Jasmeen T. (2014) Survey of Different Approaches in Biometric Iris Recognition System, *International Journal of Advanced Research in Computer Science and Software Engineerin,* 4(7), 768-771.

Wildes, R., Asmuth, J., Green, G., Hsu, S., Kolczynski, R., Matey, J. and McBride, S. (1996). A Machine-Vision System for Iris Recognition, *Machine Vision and Applications,* 9, 1-8.

Skin Diseases Predictive Model Using Individual Base and Ensemble Base Approach

Tinuke Omolewa Oladele [1], Taye Oladele Aro[2], Adebisi Samuel Segun[3]
[1, 2, 3] Department of Computer Science, University of Ilorin, Ilorin, Nigeria.
tinuoladele@gmail.com, taiwo774@gmail.com, adebisimuel@gmail.com

ABSTRACT

Purpose: This paper aims at predicting skin diseases using Neural Network, Decision Tree and Ensemble method based on majority voting.

Design/Methodology/Approach: The study employed three selected machine learning algorithms (classifiers) to predict skin diseases in WEKA data mining tool environment. Dermatological dataset was used to evaluate the predictive system. The performance evaluation was done based on accuracy, precision, true positive, false positive and recall. Comparative analysis was conducted on the three data mining techniques.

Findings: Each individual classifier and the ensemble method was put through training and testing phase using N-fold cross validation technique (N value was set to 10) and J48 gave accuracy of 93.9891% while Multi-layer Perceptron gave 96.9945%. The accuracy of 95.3552% was obtained in ensemble method using majority voting. Precision and recall rate of 0.94 for J48, 0.97 for precision and recall values for MLP and 0.95 for precision / recall values for ensemble method.

Practical Implications: The paper provides a better skin disease predictive model with much higher accuracy and generalization that avoids both over-fitting and under-fitting by combining two machine learning algorithms via the majority voting method. The developed skin disease predictive system will help the medical personnel to effectively predict skin disease accurately.

Originality/Value: Several works have been proposed by researchers in data mining for diagnosis or prediction of skin diseases, but this study's major concern is on the performance evaluation in term of over-fitting, under-fitting and generalization of individual base classifier and ensemble based approach in predicting skin diseases.

Keywords: *Prediction, Decision Tree, Neural Network, Majority Voting Technique*

1 Introduction

The skin of a human represents one of the predominant organs of the body (Sudha, Aramudhan, & Kannan, 2017) that protects against infection, heat, injury, and any part of the body in which ultraviolet (UV) radiation can cause damage (Parvin & Jafar, 2017). Several types of reactions are becoming more pronounced as symptoms in skin disease prediction (Amarathunga, Ellawala, Abeysekara, & Amalraj, 2015). The skin diseases can be very risky, mostly when it is not given attention at an early stage. The technique of data mining includes an approach of getting knowledge that is hidden from huge data (Adeyemo & Adeyeye, 2015).

The approaches available in data mining contribute to the transformation of large data into relevant data for knowledge discovery, prediction and intelligent decision making in healthcare systems (Aro, Muhammed, Ayoade, & Oladipo, 2017). The field of medicine is significant for the application of data mining and analysis technique as it concerns human life which is the most important thing in the society (Singh, Naveen, & Samota, 2013).

Numerous applications have been developed through data mining include data analysis for effective medical policy making, prevention and detection of various diseases, errors reduction in hospitals and falsified insurance detection (Kaur & Bawa, 2015). The analytical predictive system is a branch of data mining that automatically produce model for classification using training dataset and employ the same model for further prediction of classes of datasets that are not classified (Varun, Vijay & Gayathri, 2012). Data mining using predictive approach handles learning algorithms to help clinicians in monitoring tasks, diagnostics and therapeutics. In modeling through predictive method, the collection of data is achieved first, then a model will be calculated statistically, predictions are made and the model is validated as further data become available. Clinical data mining is based on research strategic to reclaim, study and deduce

both qualitative and quantitative information available from medical datasets or records.

Many classification techniques in data mining such as Naive Bayes, KNN, Neural Network, Logistic Regression, Support Vector Machine (SVM) and decision tree have been used to develop predictive and diagnostic systems (Ahmad,, Qamar, & Rizvi, 2015). Among these techniques, the decision tree has been discovered as one of the most powerful classification techniques used in learning process and decision support system (Seema, Rathi & Mamta, 2012).

A predictive technique like decision tree can be used in classification, clustering and predictive task. This technique applies a divide-conquer approach to split the problem search space into subsets. The decision tree's classification is an easy to understand approach, which is especially used when there is a need to know the trained knowledge model structure.

This paper selected J48 algorithm under the decision tree while Multi-layered perceptron was picked from the Neural Network. These learning algorithms have complementary advantages and disadvantages. Decision Tree's knowledge representation is easily understood by humans, unlike the case of Neural Networks. Decision Trees have the problem of dealing with noise during data training phase, which is not obtainable in Neural Networks. Decision Trees learn fast and Neural Networks learn relatively slowly. Therefore, with these outstanding features combining Decision Trees and Neural Network using an ensemble-based approach should produce better predictive system for skin diseases.

2 Related Work

Kadhim, (2017) used technique in data mining to classify skin diseases. For mining and processing of image data phase, the study employed decision tree. Digital image principles and decision tree were applied to detect skin diseases using some attributes identified in a digital image of a skin. The main phases in the system are: pre-processing, features extraction and classification by decision tree. The classification phase was enhanced in order to make system more accurate.

Karthik, et al (2017) presented a predictive system for dermatological condition using Naïve Bayesian classification approach. The system focused only skin diseases of 6 categories. The Naïve Bayes result was based on the parameter settings produced by probability occurrence of a particular kind of dermatological disease along with the percentage. The dermatology dataset was downloaded from the UCI repository site was used for testing of the developed skin disease predictive model.

Sudha et al. (2017) discussed the relationship among input and response attributes for improving disease diagnosis in medical area. The Response Surface Methodology (RSM) was used to develop a relationship between input attributes of skin disease and predicts the psoriasis patients with the help of independent and dependent variables. The performance of RSM model revealed that the developed empirical relationship has the greatest conformity with test results. The Analysis of Variance (ANOVA) was performed to mathematical analyze the outcome. The final experimental result showed that the developed empirical model is suitable for skin disease.

Manjusha, Sankaranaayana and Seena (2015) studied dermatological diseases of different types with like symptoms which may even prove deadly if not properly attended at exact time. The study used medical dataset of 230 instances with 22 attributes. Analysis was conducted on data gathered from the southern part of Kerala, India. For better prediction, calculation of accuracy of mining algorithm is important. Two data mining classification algorithms: Naive Bayes (NB) and J48 were used for data analysis. WEKA Open Source data mining tool was used to carry out analysis and also to reveal the chances of different dermatological diseases and also to find out the probability of occurrence of each disease.

3 Methodology

The proposed predictive model considered the individual-based and ensemble-based methods for prediction of skin diseases. The J48 algorithm was selected under the decision tree category while Multi-layered perceptron was chosen from the Neural Network category. This experiment was carried out step-by-step with focus on applying the selected classification algorithms and their ensemble on the dermatology dataset. The procedures included environment setup, data preprocessing, choosing the data mining software, running the simulation and evaluation of the performance of the classifiers.

The preprocessing stage involves putting together of clinical and histopathological attributes of patients and collating all of them together in a single database and formatting the

database into a usable dataset for the purpose of this study. The WEKA (Wekaito Environment for Knowledge Analysis) data mining tool was employed for conducting this study. The J48 and Multi-layered perceptron algorithm was applied on the dataset as well as their ensemble too. The selected classifiers were combined through a technique referred to as ensemble by majority vote. This study concern is on 6 major skin diseases as its research subjects.

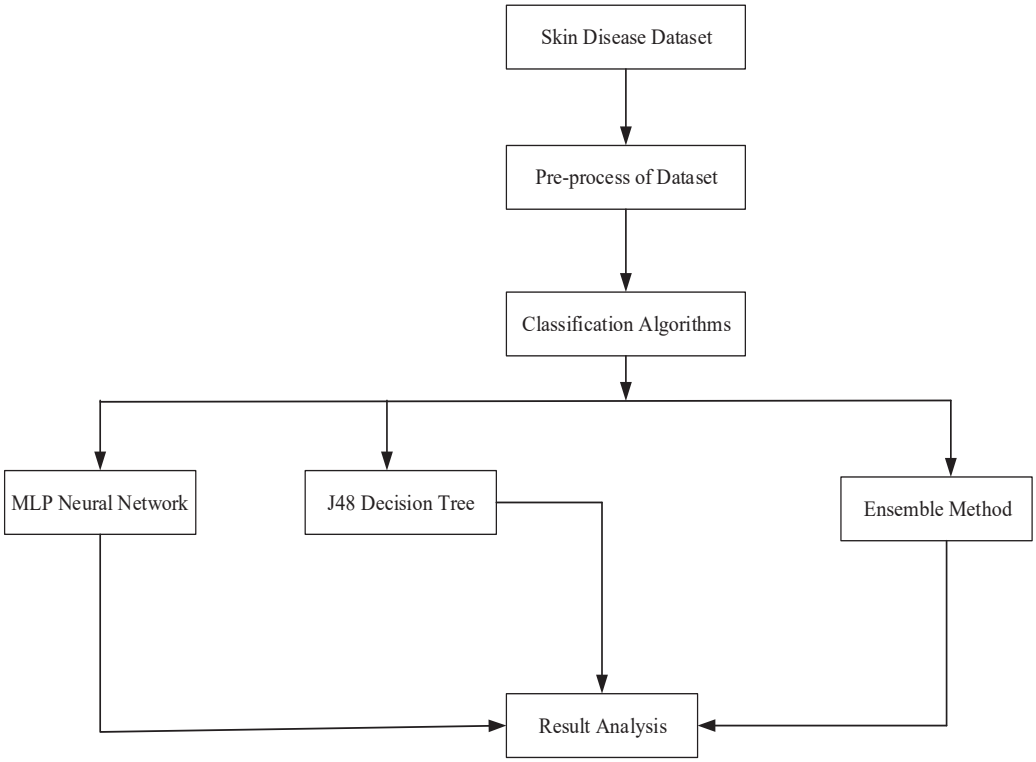


Figure 1: Proposed Skin Diseases Predicted Model

3.1 Skin Disease Dataset

This study was developed on the basis of the information gotten from the publicly available online skin diseases dataset from Dermatology Database of Gazi University, School of Medicine. After filtering and correcting missing values, 366 skin disease data were obtained which was used for prediction of six skin diseases which showed similar symptoms.

Among all, 112 pieces of data were psoriasis, 61 seboreic dermatitis, 72 lichen planus, 49 pityriasis rosea, 52 cronic dermatitis and 20 pityriasis rubra pilaris. It was also discovered that the database contains 34 attributes, 33 of which were linear valued and one of them was nominal. They all shared the clinical features of erythema and scaling, with very little differences. The diseases in this group were psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis, and pityriasis rubra pilaris as shown in Table 1

Table 1: Skin Disease Database

Class Code	Class	Number of Instances
1	psoriasis	112
2	seboreic dermatitis	61
3	lichen planus	72
4	rosea	49
5	cronic dermatitis	52
6	pityriasis rubra pilaris	20

3.2 Individual Base Classifier for Skin Diseases Predictive System

The skin disease dataset will serve as input to each base classifier. Multilayer perceptron neural network and J48 decision tree machine learning algorithm and their individual performance will be recorded respectively and the output measured and evaluated.

3.3 *Ensemble Base Classifier for Skin Diseases Predictive System*

The ensemble base skin disease predictive model was developed by combining the J48 and Multi-layered perceptron algorithms together via the Majority Voting method. The skin disease dataset also served as the input to this developed predictive model which was trained and tested using the 10-fold cross validation technique and the results of the model performances were measured and evaluated accordingly.

4. RESULTS AND DISCUSSION

4. 1 *Experimental Result of Multilayer Perceptron Algorithm for Classification*

This phase discusses the results of multilayer perceptron neural network for classification of skin disease as shown in Figure 2, Table 2 and Table 3 respectively.

```
Class 6
    Input
    Node 5


Time taken to build model: 83.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         355               96.9945 %
Incorrectly Classified Instances        11                3.0055 %
Kappa statistic                          0.9624
Mean absolute error                      0.0411
Root mean squared error                  0.0927
Relative absolute error                 15.4167 %
Root relative squared error             25.4006 %
Total Number of Instances              366

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0.004      0.991     1         0.996      1         1
                 0.918     0.016      0.918     0.918     0.918      0.996     2
                 0.986     0          1         0.986     0.993      1         3
                 0.898     0.016      0.898     0.898     0.898      0.996     4
                 1         0          1         1         1          1         5
                 1         0          1         1         1          1         6
Weighted Avg.    0.97      0.006      0.97      0.97      0.97       0.999

=== Confusion Matrix ===

   a   b   c   d   e   f   <-- classified as
 112   0   0   0   0   0 |   a = 1
   1  56   0   4   0   0 |   b = 2
   0   0  71   1   0   0 |   c = 3
   0   5   0  44   0   0 |   d = 4
   0   0   0   0  52   0 |   e = 5
   0   0   0   0   0  20 |   f = 6
```

Figure 2: Result J48 for Skin Disease Prediction

Table 2: Performance Evaluation of MLP Neural Network

Parameters	Dermatology Data set
Correctly Classified Instances (%)	96.9945
Incorrectly Classified Instances (%)	3.0055
Kappa Statistics	0.9624
Mean Absolute Error	0.0411
Root Mean Squared Error	0.0927
Relative Absolute Error	15.4167
Root Relative Squared Error	25.4006

Table 3: Performance Evaluation of MLP Neural Network

Parameters	Dermatology Data set
True Positive (TP RATE)	0.970
False Positive (FP RATE)	0.006
Precision	0.970
Recall	0.970
F-Measure	0.970
ROC Area	0.999

4.2 Experimental Results of J48 Decision Tree for Classification

This section mentions the results of J48 Decision Tree algorithm for classification of skin disease as shown in Figure 3, Table 4 and Table 5 respectively.

```
Number of Leaves  :      30

Size of the tree :       40

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         344               93.9891 %
Incorrectly Classified Instances        22                6.0109 %
Kappa statistic                          0.9246
Mean absolute error                      0.0264
Root mean squared error                  0.1365
Relative absolute error                  9.9147 %
Root relative squared error             37.397  %
Total Number of Instances              366

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.964     0.024       0.947     0.964       0.956      0.983     1
                 0.885     0.023       0.885     0.885       0.885      0.958     2
                 0.972     0.01        0.959     0.972       0.966      0.983     3
                 0.878     0.009       0.935     0.878       0.905      0.967     4
                 0.981     0.006       0.962     0.981       0.971      0.989     5
                 0.9       0.003       0.947     0.9         0.923      0.973     6
Weighted Avg.    0.94      0.015       0.94      0.94        0.94       0.977

=== Confusion Matrix ===

   a   b   c   d   e   f   <-- classified as
 108   1   2   0   1   0 |   a = 1
   2  54   1   3   0   1 |   b = 2
   0   1  70   0   1   0 |   c = 3
   2   4   0  43   0   0 |   d = 4
   1   0   0   0  51   0 |   e = 5
   1   1   0   0   0  18 |   f = 6
```

Figure 3: Result J48 for Skin Disease Prediction

Table 4: Performance Evaluation of J48 Decision Tree

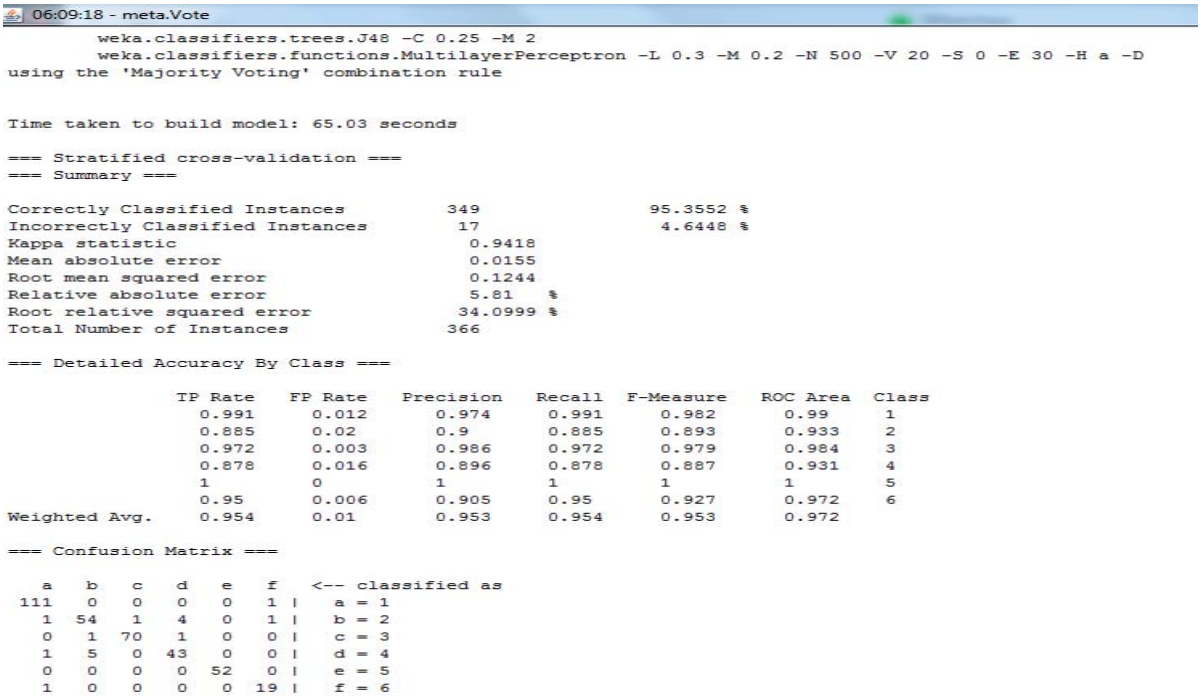
Parameters	Dermatology Data set
Correctly Classified Instances (%)	93.9891
Incorrectly Classified Instances (%)	6.0109
Kappa Statistics	0.9246
Mean Absolute Error	0.0264
Root Mean Squared Error	0.1365
Relative Absolute Error	9.9147
Root Relative Squared Error	37.397

Table 5: Performance Evaluation of J48 Decision Tree

Parameters	Dermatology Data set
True Positive (TP RATE)	0.940
False Positive (FP RATE)	0.015
Precision	0.940
Recall	0.940
F-Measure	0.940
ROC Area	0.977

4.3 Experimental Results of Ensemble-Based For Classification

The results of the combination of Multilayer Perceptron Neural Network and J48 Decision Tree Algorithm ensemble approach for classification of skin disease are shown in Figure 3, Table 4 and Table 5 respectively



```
06:09:18 - meta.Vote
        weka.classifiers.trees.J48 -C 0.25 -M 2
        weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 20 -S 0 -E 30 -H a -D
using the 'Majority Voting' combination rule


Time taken to build model: 65.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         349               95.3552 %
Incorrectly Classified Instances        17                4.6448 %
Kappa statistic                          0.9418
Mean absolute error                      0.0155
Root mean squared error                  0.1244
Relative absolute error                  5.81   %
Root relative squared error             34.0999 %
Total Number of Instances              366

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.991     0.012       0.974     0.991     0.982      0.99     1
                 0.885     0.02        0.9       0.885     0.893      0.933    2
                 0.972     0.003       0.986     0.972     0.979      0.984    3
                 0.878     0.016       0.896     0.878     0.887      0.931    4
                 1         0           1         1         1          1        5
                 0.95      0.006       0.905     0.95      0.927      0.972    6
Weighted Avg.    0.954     0.01        0.953     0.954     0.953      0.972

=== Confusion Matrix ===

   a   b   c   d   e   f   <-- classified as
 111   0   0   0   0   1 |   a = 1
   1  54   1   4   0   1 |   b = 2
   0   1  70   1   0   0 |   c = 3
   1   5   0  43   0   0 |   d = 4
   0   0   0   0  52   0 |   e = 5
   1   0   0   0   0  19 |   f = 6
```

Figure 4: Result Majority Voting by Ensemble-based for Skin Disease Prediction

Table 6: Performance evaluation of Ensemble method.

Parameters	Dermatology Data Set
Correctly Classified Instances (%)	95.3552
Incorrectly Classified Instances (%)	4.64480
Kappa Statistics	0.94180
Mean Absolute Error	0.01550
Root Mean Squared Error	0.12440
Relative Absolute Error	5.81000
Root Relative Squared Error	34.0999

Table 7: Performance measurement of Ensemble method.

Parameters	Dermatology Data set
True Positive (TP RATE)	0.954
False Positive (FP RATE)	0.010

Precision	0.953
Recall	0.954
F-Measure	0.953
ROC Area	0.972

4.4 Experimental Results of Comparative Analysis of Developed Skin Disease Predictive System

The results of each phase of the predictive system of skin diseases were properly analyzed in this section. This is shown in Table 7 and Figure 8.

Table 8: Comparative Analysis of Performance of Machine Learning Algorithms and Ensemble-based

Classifier	Dermatology Data set				
	Accuracy (%)	TP Rate	FP Rate	Precision	ROC Area
MLP	96.9945	0.97	0.006	0.970	0.999
J48	93.9891	0.94	0.150	0.940	0.977
Ensemble Method	95.3552	0.954	0.010	0.953	0.972

The results of comparative analysis from Table 8, and graphical representation in Figure 7, shows that the ensemble approach of skin diseases predictive model gives the most moderate result out of the three skin disease predictive models developed as it is carefully positioned to avoid over-fitting which might likely occur in the MLP Neural Network predictive model and also under-fitting in the J48 Decision Tree predictive model, thereby standing as a more reliable model for skin disease predictive system.

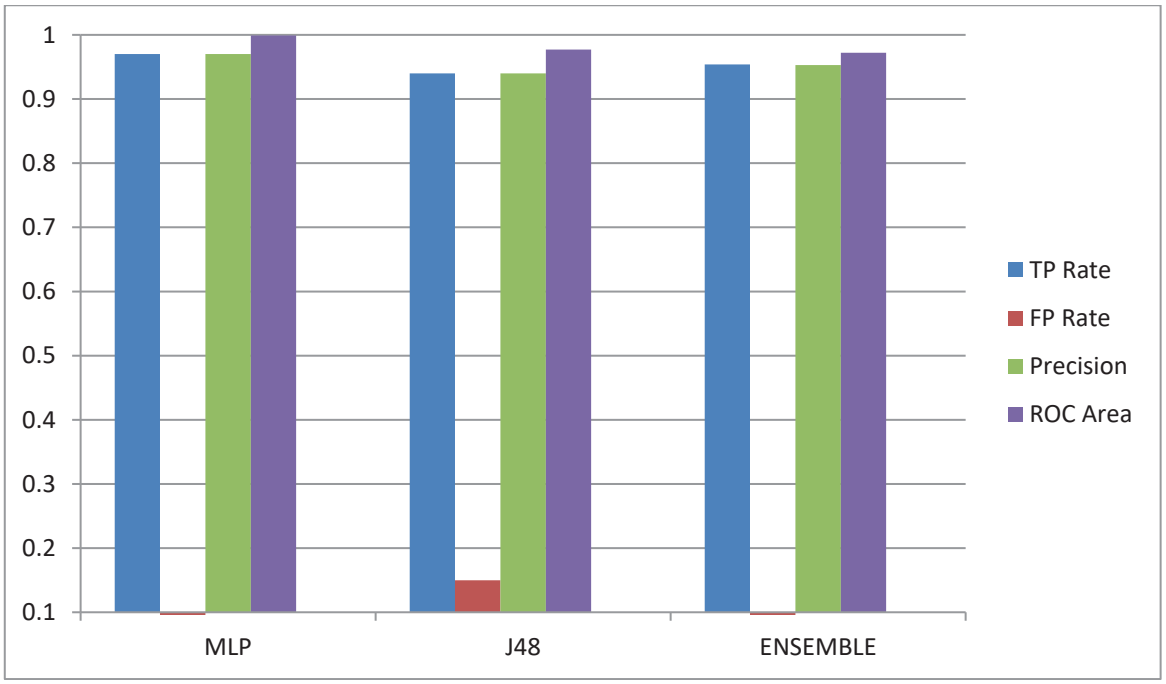


Figure 7: Comparative Analysis of Predictive Skin Diseases Models

5 Summary and Discussion

The prediction of skin disease model using three data mining classification techniques and ensemble-based approach gave different results for each machine learning algorithm (classifier) used. It was observed that the accuracy of 96.9945% was obtained in multilayer perceptron neural network (MLP), while 93.9891% was recorded in J48 decision tree. The misclassification rate of 3.0055% was obtained in MLP, while the J48 presented misclassification of 6.0109%. The MLP got 0.97 True Positive Rate (TP RATE), while J48 got 0.94 True Positive (TP RATE). For the False Positive Rate (FP RATE), MLP recorded 0.006, while J48 obtained 0.015 for False Positive Rate (FP RATE). For precision, MLP gave 0,97, while the J48 gave 0.94. The MLP gave recall rate of 0.97, while the J48 recorded 0.94.

Finally, the overall results from the skin diseases predicted model using ensemble approach based on majority voting gave outstanding results compared with individual based method of multilayer neural network and J48 decision tree algorithm.

It was observed that the ensemble method gave accuracy of 95.355% which is the most the modest result out of the predictive models developed as it is carefully positioned to avoid over-fitting which might occur in the MLP Neural Network predictive model and also under-fitting in the J48 Decision Tree predictive model, thereby

standing as a more reliable model for skin diseases.

6 Conclusion

In this paper, skin diseases predictive model was developed using individual-based and ensemble data mining methods. The best prediction model was obtained by combining J48 and MLP using the Majority voting technique. It was concluded that the developed model helps the doctors to clear their confusion when making decision and predicting diseases with similar symptoms.

At the end of the analysis, it was discovered that the majority voting technique was able to help build the model in less than 2 minutes (65.03secs). It also produced a more accurate and correct result of 95.35% as against the individual-based model. This proves that the combination of Decision Tree and Neural Networks by ensemble method is of great potential that can be successful used in medical diagnosis of skin diseases.

References

Adeyemo, O. O. & Adeyeye, T. O. (2015). Comparative Study of ID3 / C4 . 5 Decision tree and Multilayer Perceptron Algorithms for the Prediction of Typhoid Fever. *African Journal of Computing & ICT*, *8*(1), 103–112.

Ahmad, P., Qamar, S. & Rizvi, S. Q. (2015). Techniques of Data Mining In Healthcare : A Review. *International Journal of Computer Applications*, *120*(15), 38–50.

Amarathunga, A. A. L. C., Ellawala, E. P. W. C., Abeysekara, G. N. & Amalraj, C. R. J. (2015). Expert System For Diagnosis Of Skin Diseases. *International Journal of Scientific & Technology Research*, *4*(1), 174–178.

Aro, T. O., Muhammed, B. J., Ayoade, O. B. & Oladipo, I. D. (2017). A Review on Data Mining Techniques for Heart Disease Prediction. *Anale. Seria Informatica*, *XV*(1), 99–103.

Kadhim, Q. K. (2017). Classification of Human Skin Diseases using Data Mining. *International Journal of Advanced Engineering Research and Science*, *4*(1), 159–163.

Karthik, S., Rahul, B. S., Vibhudhi, M. & Tej, T. (2017). Prediction of Dermatological Condition Using Naive Bayesian Classification. *International Journal of Pharmacy & Technology*, *9*(1), 28988–28994.

Kaur, S. & Bawa, R. K. (2015). Future Trends of Data Mining in Predicting the Various Diseases in Medical Healthcare System. *International Journal of Energy, Information and Communications*, *6*(4), 17–34.

Manjusha, K K, Sankaranaayana, K. & Seena, P. (2015). Data Mining in Dermatological Diagnosis : A Method for Severity Prediction. *International Journal of Computer Applications*, *117*(11), 11–14.

Parvin, S. R. & Jafar, O. A. M. (2017a). Prediction of Skin Diseases using Data Mining Techniques. *Internationa Journal of Advanced Research in Computer and Communication Engineering*, *6*(7),313–318. http://doi.org/10.17148/IJARCCE.2017.6754

Seema, Rathi, M. & Mamta. (2012). Decision Tree: Data Mining Techniques. *International Journal of Latest Trends in Engineering and Technology (IJLTET)*, *1*(3), 150–155.

Singh, D., Naveen, H. & Samota, J. (2013). Analysis of Data Mining Classification with Decision Tree Technique. *Global Journal of Computer Science and Technology*, *13*(13), 1–6.

Sudha, J., Aramudhan, M. & Kannan, S. (2017). Development of a mathematical model for skin disease prediction using response surface methodology. *Biomedical Research*, 355–359.

Varun K. M., Vijay, S. V. & Gayathri, D. B. (2012). Hepatitis Prediction Model based on Data Mining Algorithm and Optimal Feature Selection to Improve Predictive Accuracy. *International Journal of Computer Applications*, *51*(19), 13–16.

RSA Cryptosystem Encryption Based on Three Moduli Set With Common Factor {2n+2, 2n+1,2n}

[1]Saheed Y. Kayode and [2]Gbolagade K. Alagbe
[1]Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria
[2]Department of Computer Science, Kwara State University, Malete, Nigeria
[1] yksaheed@alhikmah.edu.ng, [2]kazeem.gbolagade@kwasu.edu.ng

ABSTRACT
Purpose: This research proposes a method to implement Rivest Shamir Adleman (RSA) cryptosystem encryption and decryption operations in a quick way.
Approach: In this paper, we provide two optimizations to make the operation easy. Then, we introducea Residue Number System (RNS) based on three moduli set with common factor{2n + 2, 2n+1, 2n} for the encryption of the message.
Findings: The RNS promises the proposed algorithm to be highly parallelizable, well adapted to parallel architecture and well suited to hardware implementations. This moduli set provides faster speed and requires little hardware. The key length is also increased as the moduli set are used as part of the private key. Theoretically, our proposal achieved about 40% scale down of arithmetic operations whilst also improving the security of the RSA cryptosystem.
Originality/value: The proposed scheme outperforms two related state-of-art schemes in terms of security, encryption and decryption speed of RSA.
Keywords: *RSA Cryptosystem, modular arithmetic, Residue number system, moduli set, encryption.*
Paper type: *Research paper.*

1 Introduction

The recent paradigm shift in Information Technology cannot be overemphasized. Security has come to be an essential component in many application implementations such as Virtual Private Network (VPN) and electronic commerce. There are many methods for securing information and data; such as cryptography and intrusion detection system.Cryptography is important whenever sensitive information is to be communicated in a network. Computer security has an important role in securing information (Damrudi and Norafida, 2013).

According to Saheed and Gbolagade (2017a), cryptography presents methods for preserving information secret, for determining that information has not been tampered with, and for deciding who authored pieces of information. RSA Algorithm is one of the most important public cryptosystems used for high secure transmission of data.The security of asymmetrickey encryption such as RSA depends on the problem of integer factorization (Menezes, Vanstone and Oorschot, 1996).

The cryptography algorithms can be categorized as symmetric key cryptography and asymmetric key cryptography. In symmetric key cryptography, the same key is used for encryption and decryption operations, whereas two keys are employed for encryption and decryption operations of messages in asymmetric key cryptography. Asymmetric key cryptography provides more security compared to symmetric key cryptography (Abu, Deepthi and Puskar, 2015). The problem with the RSA cryptosystem is that it requires a lot of computational power in order to provide high level of security (Saraiva, 2009). The main bottleneck of public key algorithms is that they are slower compared to symmetric-key simply because of their foundation in modular arithmetic. Hence, how to make a faster and more efficient implementation of public key algorithms such as RSA is a challenge to researchers in cryptography field.

Recently, Saheed and Gbolagade (2017a) proposed an improved RSA algorithm based on RNS. In their paper, two optimizations were investigated to make the RSA cryptosystem operation easy.Abu et al. (2015) presented a parallel algorithm to handle the RSA decryption by investigating the effect of Pthread and computing a unified device architecture (CUDA)

on decryption operation in RSA cryptosystem. Mohammed, Heba and Jawad (2016) presented an acceleration of the RSA Processes based on Parallel Decomposition and Chinese Remainder theorem. They proposed variant decompositions to gain extra speed. Saxena and Kapoor (2014) proposed the new Parallel RSA algorithm based on repeated square and multiply method. Also, a few researchers have worked on parallel RSA such as Fan et al. (2010), Li et al., (2010) and Qing et al. (2010). Fewer multiplications and less delay in a new parallel exponentiation algorithm for RSA were presented by Sepahvandi et al. (2009).

This paper proposes a new encryption approach for RSA cryptosystem based on three moduli set {2n+2,2n+1,2n} with a common factor. This paper is organized in sections.In sections 2 and 3 the RSA cryptosystem and RNS are discussed. Section 4 highlights the proposed RSA cryptosystem and hardware performance comparison, and section 5 concludes the paper.

2 RSA Cryptosystem

The RSA cryptosystem was invented by Ron Rivest, Adi Shamir, and Leonard Adleman in 1977 and was publishedfirst in the August 1978 issueof Scientific American.The RSA algorithm can be used for public key encryption and digital signatures. The RSA cryptosystem is generally used for ensuring authenticity of digital data and providing privacy. Recently, RSA has been deployed in many businesses and commercial systems. It is used in web server applications and to secure web traffic in browsers. It is used to provide authenticity and privacy to emails. It is at the heart of electronic credit card payment systems and also used to secure login sessions (Saheed and Gbolagade, 2017b).

3 Residue Number System

In RNS, a set of moduli which are all relatively prime and independent of one another are given. In this approach, integer numbers are represented by the remainders also known as residues of each modulus and operations of the arithmetic are individually based on those residues (Gbolagade, 2010; Luu,2004). The foundation for an RNS representation is a set of relatively prime integers $\{m_1,m_2,m_3, :::;m_k\}$ such that gcd $(m_i, m_j) = 1$ for $i \neq j$, where the greatest common divisor gcd is of m_i and m_j . For such a system, $M = \prod_{i=1} kmi$ is the dynamic range and any integer $X \in [0; M-1]$ and can be represented uniquely as $X= (x_1, x_2, x_{3...}; x_k)$, where $x_i = |x|_{mi}, 0 \leq xi < mi$. The quest for RNS is the capability to afford carry-free addition as this results in high-speed arithmetic units.

The carry-propagation problem of conventional binary number system representation was then the main bottleneck and challenge of fast arithmetic operation and became the key justification and motivation driving and making researchers to venture into this alternative number system known as RNS for which the residue arithmetic operation in each of the modulus channels is carry free and independent (Chang, Molahosseini, Zarandi and Tay, 2015).The significant property of RSD number representation systems is the ability to perform borrow-free subtraction and carry-free addition (Parhami, 1990).

4 Proposed RSA Cryptosystem

The exponentiation is optimized as explained in this section. To compute m^3mod n, we compute m^2 mod n and one modular squaring, again m^3 mod n and a modular multiplication with m.Then, we convert the encrypted message into residue number system using three moduli sets{2n + 2, 2n+1, 2n} with a common factor. The decryption operation is done in a similar way: first computes c^2 mod n, then c^3 mod n, and c^6 mod n, and c^7mod n by alternating modular multiplication and squaring.

Key Pair

Public key: n= 55, e = 3

Private key: n = 55, d = 7

Key Pair Generation

Primes: p = 5, q =11

modulus: n = pq = 55

Public exponent: e = 3

Private exponent: $d = 3^{-1} \bmod 20 = 7$

Table 1: RSA Encryption and Decryption using moduli set {2n+2,2n+1,2n}

Message	Encryption	Ciphertext(C_1)	Ciphertext (C_2)	Decryption M = c^7mod n	Decryption (C_1)	Decryption (C_2)	Message
M	m^2mod n	m^3 mod n	{2n+2,2n+1,2n}	C^2 mod n	c^3mod n	c^6mod n	c^7mod n
0	0	0	{0,0,0}	0	0	0	0
1	1	1	{1,1,1}	1	1	1	1
2	4	8	{2,2,2}	9	17	14	2
3	9	27	{3,3,3}	14	48	49	3
4	16	9	{4,4,0}	26	14	31	4
5	25	15	{5,0,1}	5	20	15	5
6	36	51	{0,1,2}	16	46	26	6
7	49	13	{1,2,3}	4	52	9	7
8	9	17	{2,3,0}	14	18	49	8
9	26	14	{3,4,1}	31	49	36	9

The main operation of modular exponentiation is multiplication and we have tried to reduce the number of multiplications in our approach and also introduced a residue number system based on three moduli sets{2n+2,2n+1,2n}with a common factor which was proposed by Gbolagade(2010) for further transformation of the cipher-text so as to create more confusion to cryptanalyst. The reduction in the size of the power as showed in table 1 makes the computation of the proposed approach to be faster and reduce computational time.

Table 2: Hardware requirements performance comparison

Metrics	(Ahmad, Wang and Swamy, 1999)	(Gbolagade and Cotofana, 2008)	Propose scheme
Area	four(4) adders two(2) multipliers	Three (3) adders Four(4)multipliers	Four (4) adders Two(2)multipliers
Delay	three (3) additions two(2)multiplications	Two (2) additions Two(2)multiplications	Three(3)additions One(1)multiplication
Modular operations	$M_1/2\ m_3$	M_3	$M_3/2$

Figures 3.1 and 3.2 illustrate the hardware realization of the converter by Ahmad, Wang and Swamy (1999) and Gbolagade and Cotofana (2008) respectively. As can be observed from table 2, this proposal requires less delay and area, and operates on small operands of magnitude. Specifically, this proposed scheme requires four (4) 2:1 adders and two (2) multipliers; figure 3.2 requires one (1) 3:1 adder, two (2) 2:1 adders, and four (4) multiplexers while figure 3.3 requires one (1) 3:1 adder, three (3) 2:1 adders and two (2) multiplexers. In view of critical delay, our proposal requires three (3) additions and one (1) multiplication with mod-n operations. The converter in Gbolagade and Cotofana (2008)requires two (2) additions and two (2) multiplications with mod-2n operations whereas the converter by Ahmad, Wang and Swamy (1999) requires three (3) additions, two (2) multiplications with mod $m_1/2m_3$ operations. Thus, our scheme requires less area with the same delay. Also, due to the reason that the proposed scheme operates on smaller operands of magnitude, less complex multipliers and adders are required which results potentially in even faster implementation.

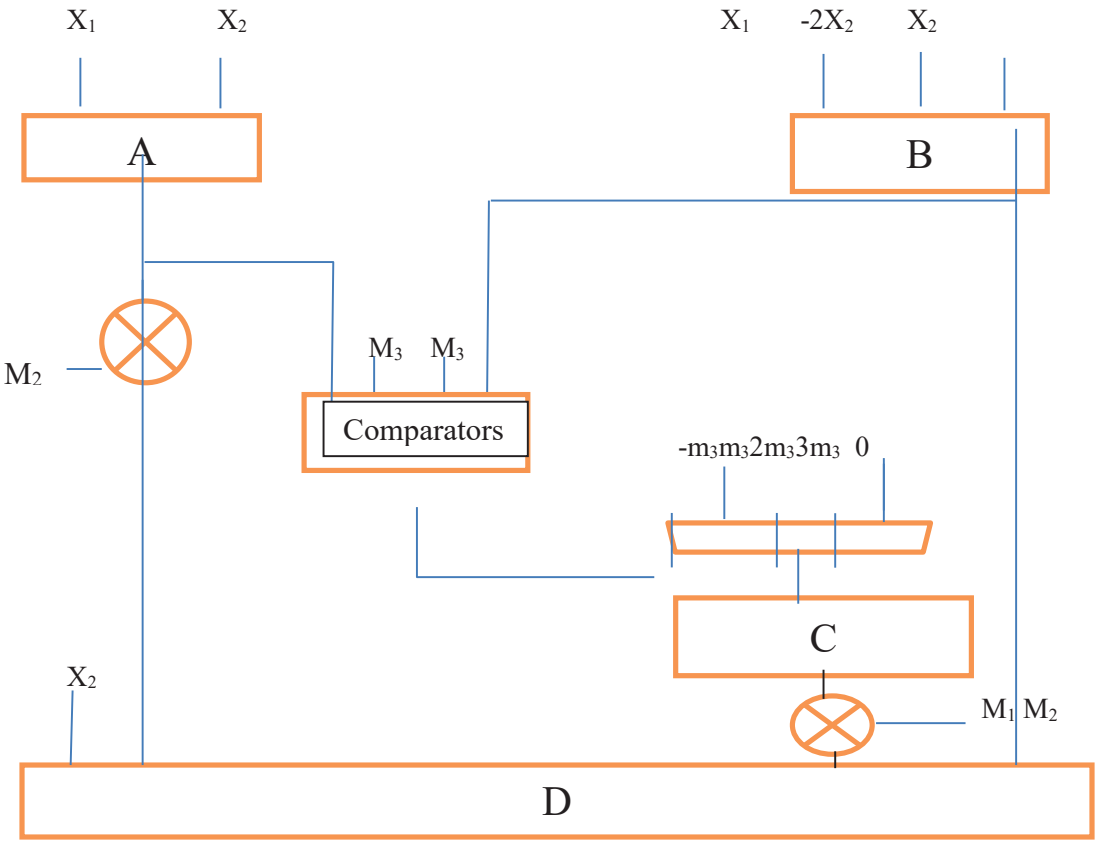


Figure 3.1. The hardware realization of proposed scheme

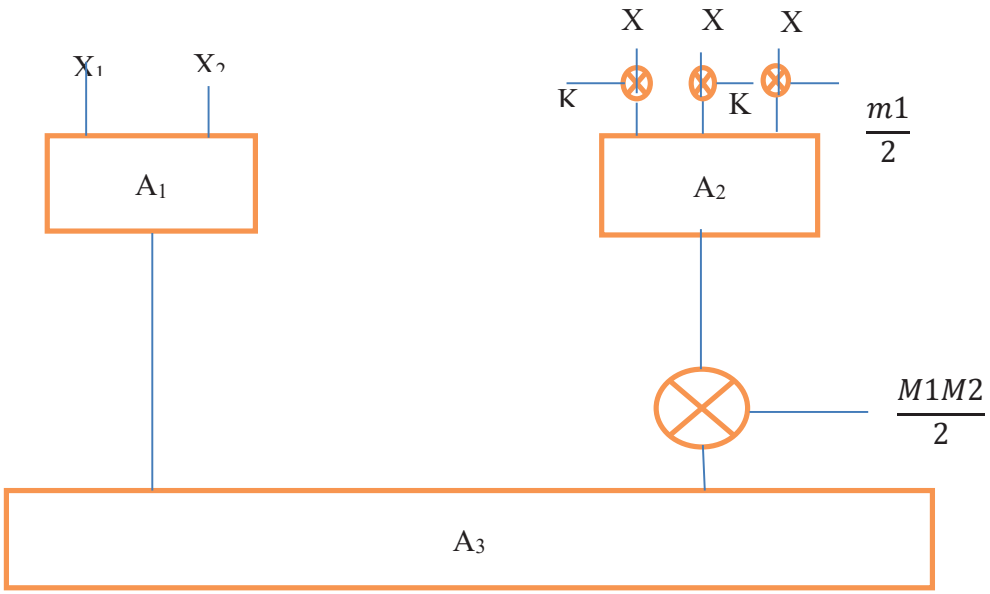


Figure 3.2 Converter Delay Path
Source: Gbolagade and Cotofana (2008)

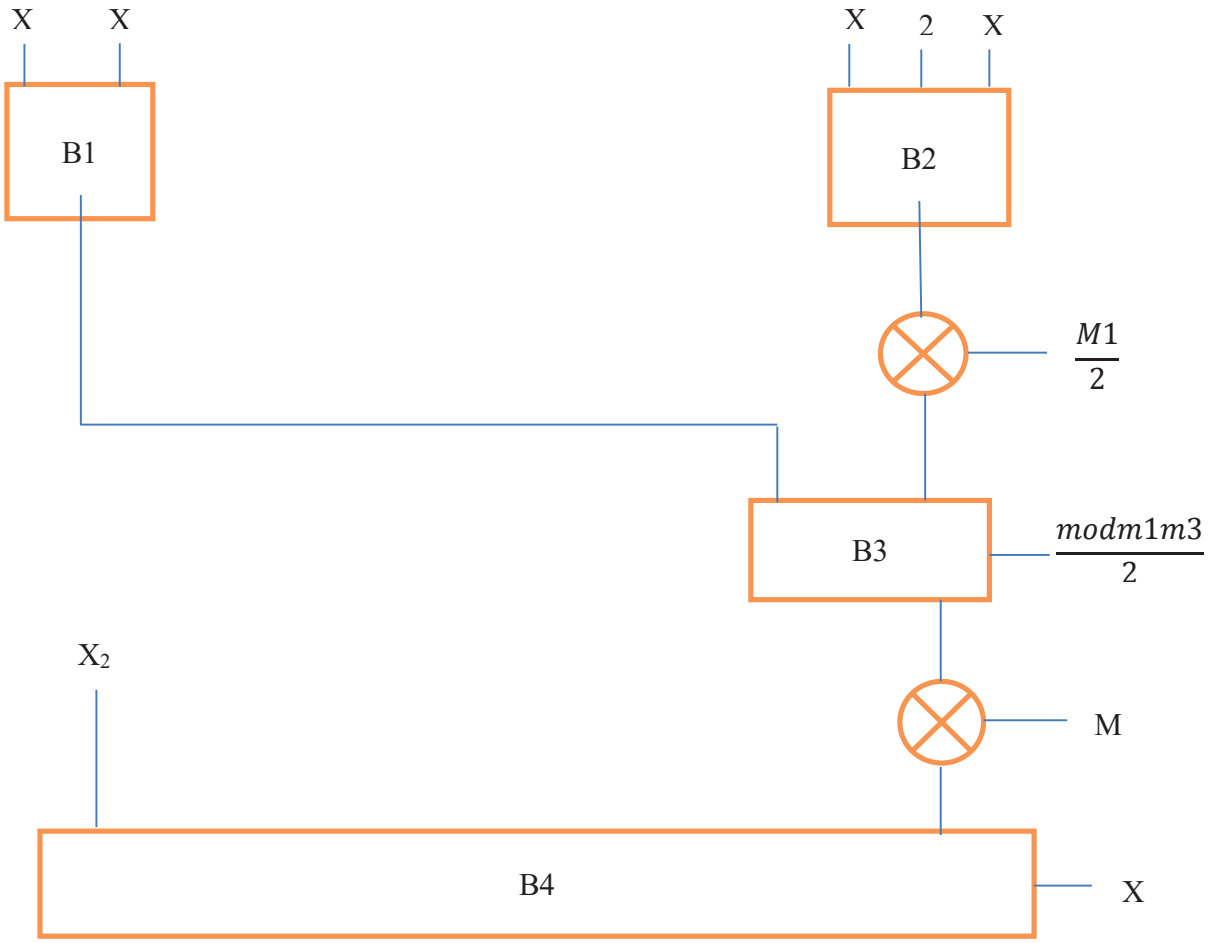


Figure 3.3. Converter Data Path
Source: Ahmad, Wang and Swamy (1999)

5 Conclusion

In this paper, we suggest a better way to enhance the speed of RSA algorithm by splitting and reducing the power operation. Theoretically, our proposed approach optimizes the exponentiation of the RSA algorithm compared to the traditional RSA cryptosystem.Our scheme requires less delay with the same area or less. Also, the total number of arithmetic operations is reasonably reduced by our proposed scheme. In the future work, we intend to look at the implementation of RSA cryptosystem based on Montgomery exponentiation with RNS.

References

Abu, Asaduzzaman, Deepthi Gummadi, & Puskar Waichal (2015) A Promising ParallelAlgorithm to Manage the RSA Decryption Complexity *In: Proceeding of the IEEE Southeast Con 2015*, April 9-12, 2015 – Fort Lauderdale, Florida.

Ahmad, M. O., Wang, Y. & Swamy, M. N. S. (1999). Residue to binarynumber converters for three moduli set.*IEEE Trans. on Circuits and Systems-II,46*(7), 180-183, Feb.

Chip-Hong, Chang Amir Sabbagh Molahosseini, Azadeh Alsadat Emrani Zarandi, & Thian Fatt Tay (2015) Residue Number Systems: A New Paradigm to Datapath Optimization for Low-Power and High Performance Digital Signal Processing Applications *IEEE circuits and systems magazine.* 26-44, 1531-636X/15.2015IEEE.

Damrudi, Masumeh & Norafida, Ithnin (2013) Parallel RSA encryption based on tree architecture. *Journal of the Chinese Institute of Engineers, 36*(5), 658-666.

Fan, W., Chen, X. & Li, X., (2010) Parallelization of RSA algorithm basedon compute unified device architecture.In: 9th international conference on gridand cooperativecomputing (GCC), 1–5November 2010, Nanjing, China.Washington, DC: *IEEE Computer Society*, 174–178.

Gbolagade, K. A. & S. D. Cotofana (2008) Residue number system operandsto decimalconversion for 3-moduli sets.*In Proceedings of 51st IEEE MidwestSymposium on Circuits and Systems,*791-794, Knoxville, USA, August, 2008.

Gbolagade, K. A. (2010) Effective Reverse Conversion in Residue Number System Processors. ISBN 978-90-72298-10-2.PhD Thesis. Printed in The Netherlands.

Li, Y., Liu, Q. & Li, T., (2010) Designand implementation of an improved RSA algorithm. In: International conference on e-health networking, digital ecosystems and technologies (EDT), 17–18 April 2010, Shenzhen, China. Piscataway, *NJ:IEEE Computer Society,* 390–393.

Luu, Mi (2004) *Arithmetic and logic in Computer systems*. New Jersey: John Wiley and Sons.

Menezes, A. J., Vanstone, S.A. & Oorschot, P. C. V. (1996) *Handbook of Applied Cryptography*. Florida: CRC Press.

Mohammed, Issam Younis, Heba, Mohammed Fadhil, Jawad, Zainab Nadhim (2016) Acceleration of the RSA Processes based on Parallel Decomposition and Chinese Remainder Theorem, *International Journal of Application or Innovation in Engineering & Management, 5*(1), January, 12-23.

Parhami, B. (1990) Generalized signed-digit number system: a unifying framework for redundant number representation. *IEEE Transactions on Computers, 39*(1), 89-98.

Qing, L., Yunfei, L. & Lin, H. (2010). On the design and implementation of anefficient RSAvariant. *In: 3rd international conferenceon advanced computer theory and engineering (ICACTE),* 20–22 August 2010, China. Piscataway, NJ: IEEE Computer Society *3*,533–536.

Saheed Y. K. & Gbolagade K.A. (2017b) Efficient RSA Cryptosystem Decryption Based on Chinese Remainder Theorem and Strong Prime. *Anale. Serial Informatică. XV*(2), 43-47.

Saheed Y. K., & Gbolagade K. A**.** (2017a). An Improved RSA algorithm based on Residue Number System. *In: Proceedings of the 13th Nigeria Computer Society International Conference, 28*, 86-91.

Saxena, S. & Kapoor, B. (2014) An efficient Parallel Algorithm for Secured Data Communications using RSA Public Key Cryptography Method. *IACC,* 4th IEEE International Advance Computing Conference.

Sepahvandi, S., Hosseinzadey, M. & Jalali, Amin (2009) An improved exponentiation algorithm for RSA cryptosystem. *In: International conference on research challenges incomputer science, 2009*. ICRCCS ’09, 28–29.

An Improved Image Scrambling Algorithm Using {2n -1, 2n, 2n +1}
Damilare Peter Oyinloye and Kazeem Alagbe Gbolagade 1

Development of Iris Biometric Template Security Using Steganography
Saheed Y. Kayode, Abdulsalam S. Olaniyi, Arowolo M. Olaolu and Akinbowale N. Babatunde 8

Skin Diseases Predictive Model Using Individual Base and Ensemble Base Approach
Tinuke Omolewa Oladele, Taye Oladele Aro and Adebisi Samuel Segun 18

RSA Cryptosystem Encryption Based on Three Moduli Set With Common Factor {2n+2, 2n+1,2n}
Saheed Y. Kayode and Gbolagade K. Alagbe 27

UWS UNIVERSITY OF THE WEST of SCOTLAND

Printed by Printing Services, University of the West of Scotland
University of the West of Scotland is a registered Scottish charity. Charity number SC002520