# A New Application for Encrypting a Database Model

Alberto Arteta*, Luis Fernando Mingo, Vijaya Kumar Gurumoorthy, and Akshay Harshakumar

***Abstract*—This paper aims at coming up with a technique that secures existing Form Maintaining Encryption with Advanced Encryption Standard (AES), by using Elliptical Curve Encryption (ECC). ECC is more efficient than Rivest–Shamir–Adleman (RSA) in that a 256-bit ECC provides the same level of security as n 3072-bit RSA. The application of this technique is applied to a data repository, part of a Database system. Pattern recognition in the encrypted data is run to test the encryption strength. In many database applications, the combination of Form Maintaining Encryption (FME) with RSAand AES are common. Our paper introduces and studies a hybrid approach of Feature Manipulation Engine (FME) and ECC, given the computational advantages of the latest versus AES.**

***Index Terms*—Asymmetric encryption, data pattern matching, enhanced database encryption**

## I. Introduction

Presently cryptographic techniques are very advanced [1–8]. The Information storage requires an organizational structure, which is provided by database systems [1]. Important and precious information is usually made part of databases. The main role of databases is to provide a method by which users can read, write and modify their data. Therefore, these databases also come with a lot of different types and schemes, in addition to the traditional models of data storage. Databases are used to store complex forms of data in simple relational forms with no redundancy.

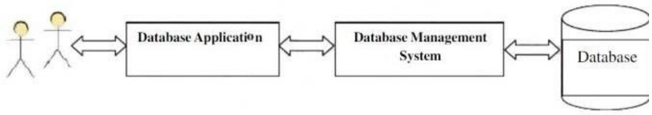


Fig. 1. A database system: Overviews of connections between frontend and backend.

A table storing data usually concentrates on a single information theme. For example, a banking database table usually has information columns regarding customer names, account balances, and so on.

Every organization has a set of sensitive or non-sensitive data that is its biggest asset. One of the biggest challenges faced by organizations today is to keep the data secure, ie. sensitive information. Credit card numbers, social security numbers, etc. are considered highly sensitive. Cases of attacks on information security by malicious entities are on the rise today. Customers and people, in general, are increasingly becoming aware of these risks and are ever more cautious of organizations. This has led them to put in more research and models of security to put their customers' minds at ease. Cryptography has become a huge tool in maintaining the confidentiality and security of personal information and is being used everywhere, such as encryption of files or complete storage media, to do online transfers through the Internet and Virtual Private Network (VPN) technologies. Out of all recent technological measures to ensure the security of data, cryptography is the most successful and is now even being used interchangeably with encryption. Cryptography has four main goals to protect information [9]:

1. Privacy of message
2. Integrity of message
3. Authentication of sender
4. Non-repudiation of sender

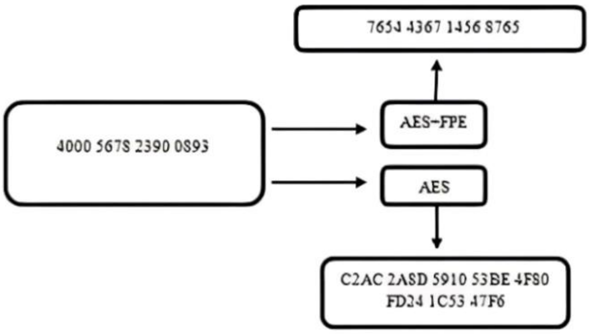


Fig. 2. A credit card number encrypted with FME+AES.

The optimal way to ensure the security of data being stored in databases is to encrypt your database or data in it. But encryption can only protect the security of your data. Unauthorized or malicious entities are still out there to access and declassify your valuable information. This means that it is essential to have proper encryption/decryption schemes to ensure a secure model. Many famous encryption/decryption schemes are being used nowadays like AES, Data Encryption Standard (DES), and Blowfish. One drawback associated with these is that it could lead to various performance issues due to overheads introduced by so many keys:

- The processing time for encrypting data could be large with these schemes.
- Encrypted data may have additional factors that demand more storage than the original.
- Encrypted data may take longer to be processed during database operations and there could be performance overheads associated with this.

This is where we need to think of maintaining the form of existing data when we try to encrypt it. This is a major field of research in recent times. It has some major advantages like:

- The security of database systems and their transparency is increased by keeping the form of data together.
- It also is good at applying masks to data to prevent it from being utilized in case of breaches.

Manuscript received October 9, 2022; revised December 5, 2022; accepted February 13, 2023.

Alberto Arteta, Luis Fernando Mingo, Vijaya Kumar Gurumoorthy, and Akshay Harshakumar are with Troy University, USA.

Luis Fernando Mingo is with Polytechnic University of Madrid, Spain.

*Correspondence: aarteta@troy.edu (A.A.)

DOI: 10.7763/IJCTE.2023.V15.1336

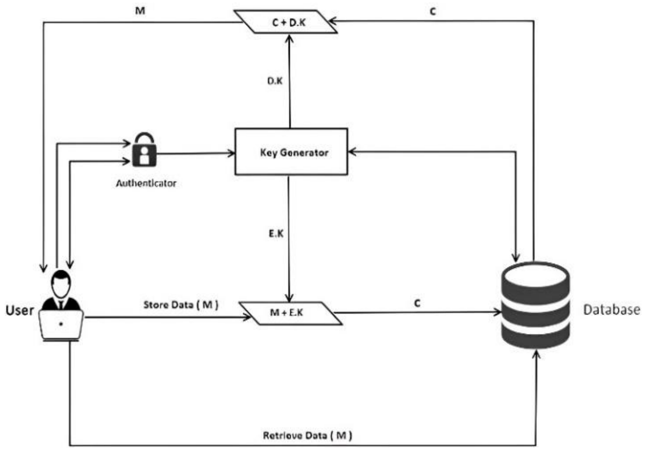


Fig. 3. A database store and retrieve schema.

Form maintaining encryption, as the name suggests, aims at keeping the original form of the data after encryption. This is suitable for Database applications, given the need for the data to be unchanged and available as much as possible. That is one of the reasons why Feature Manipulation Engin (FME) is selected as the desired method. This would mean that the output after encryption would share them for integrity with the input before encryption. For example, we consider data containing Social Security Numbers. The SSNs before and after encryption will have similar forms to it.

## II. Techniques Used

### A. Form Maintaining Encryption

FME can be defined as a symmetric key cipher that encrypts an input message A and converts it into output message B while making sure that A and B have the same form. The two classical definitions of FME are stated below:

1. Basic FME: The problem that FME tackles is defined, i.e., it makes sure that the input and output fall on the same domain. FME can be described as a function shown in Eq. (1).

   $$E: X \times K \rightarrow K$$

   where,
   E: a function that performs permutations and is reversible.
   X: denotes the key space.
   K: denotes the domain to which the input belongs, and conversely the output.
2. The generalized FME: States that the complexity of the message space is a determining factor in the complexity of the output space.

### B. Advance Encryption Standard

AES [9] used asymmetric key, created by Vincent Rijmen and Daemen in 1999, to overcome the disadvantages of DES [10] algorithm. AES utilizes the same key to be used for encryption and the reverse transformation, called decryption [3]. The block and key sizes are first analyzed by the AES algorithm before applying them to the input data. 128, 192, and 256 bits are the key sizes that are employed in the AES algorithm.

There are four main transformations performed by the AES Columns, and Add Round Key. In total, ten rounds are performed. Rounds will repeat the four transformations mentioned above. transformations are dropped, namely, Mix Columns, whichthe four transformations are described as:

1. Sub Byte: This employs an S-box substitution table.
2. Shift Rows: This shifts the rows of the State array by different offsets.
3. Mix Columns: This mixes the data within each column of the State array.
4. Add Round Key: This function is defined by adding a round key to the State.

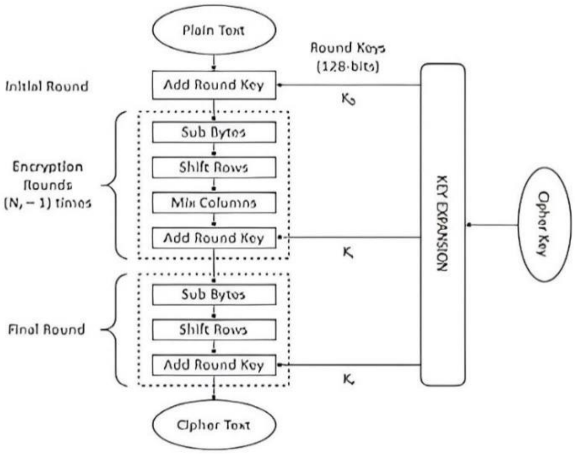


Fig. 4. AES Encryption Flowchart.

### C. Form Maintaining Encryption Techniques

The designing of FME was done in catering to the block cipher by using AES-128-bit encryption algorithm as the base for encrypting the data. The two techniques that help in achieving this are detailed here.

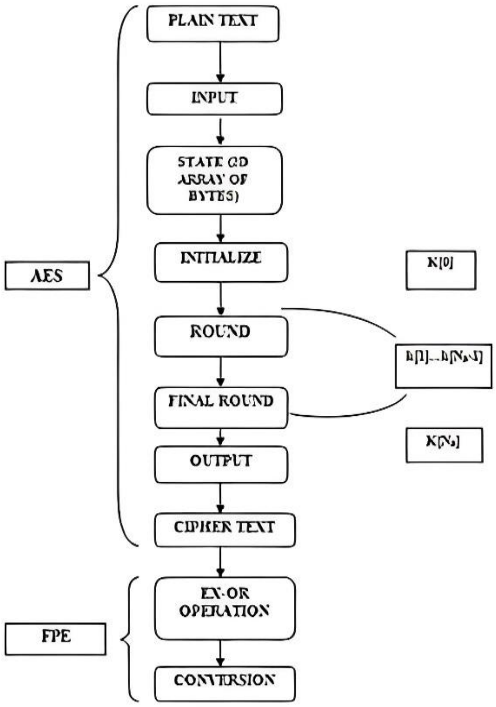


Fig. 5. Flowchart of FME + AES.

1. Exclusive OR Operation: The exclusive OR operation will divide the resultant cipher text bits into bit groups. The operation is performed on each group. The last higher order bytes are eXclusive OR with the lower order bytes. After the completion of this step, we will get a

block. At the end of the last round, the p-digit number is encrypted as q-bit data output.

2. Translation Method: In this method, 5211 codings are applied to hex digits to get precise decimal digits. It is made sure that these decimal digits are within the proper range of 0 to 9. Here, a four-digit hexadecimal number can denote 2 two-digit decimal numbers. After this step, it is ensured that both input and output maintain the same format.

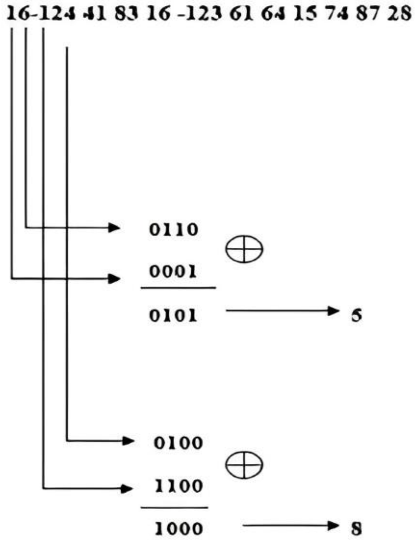
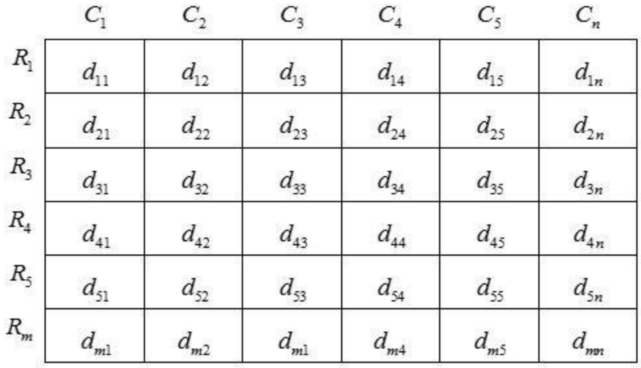


Fig. 6. eXclusive OR performed on AES output.

## III. Proposed Method

In this paper, we propose a technique that uses Elliptic Curve Encryption in addition to the existing form, maintaining Encryption with AES, to improve the security of the existing model described above. ECC has the added benefit that a 256-bit ECC encryption is as secure as a 3072-bit RSA encryption. To make our method clearer, our method is a comparison of AES vs ECC when combining with FME, and it is not a process of data extraction.



| | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ | $C_n$ |
|---|---|---|---|---|---|---|
| $R_1$ | $d_{11}$ | $d_{12}$ | $d_{13}$ | $d_{14}$ | $d_{15}$ | $d_{1n}$ |
| $R_2$ | $d_{21}$ | $d_{22}$ | $d_{23}$ | $d_{24}$ | $d_{25}$ | $d_{2n}$ |
| $R_3$ | $d_{31}$ | $d_{32}$ | $d_{33}$ | $d_{34}$ | $d_{35}$ | $d_{3n}$ |
| $R_4$ | $d_{41}$ | $d_{42}$ | $d_{43}$ | $d_{44}$ | $d_{45}$ | $d_{4n}$ |
| $R_5$ | $d_{51}$ | $d_{52}$ | $d_{53}$ | $d_{54}$ | $d_{55}$ | $d_{5n}$ |
| $R_m$ | $d_{m1}$ | $d_{m2}$ | $d_{m1}$ | $d_{m4}$ | $d_{m5}$ | $d_{mn}$ |

Fig. 7. Database structure.

In the existing model, the main point of vulnerability is the exposure of encryption keys in AES. This is solved in this proposed method by employing the following steps:

Encryption with AES:

$$e_{ij} = E\left(d_{ij}, k_{ij}\right), i = 1,2, \dots, m, j = 1,2, \dots, n$$

Encryption with ECC:

$$\varepsilon_{ij} = \varepsilon\left(k_{ij}, PK_{ij}\right), i = 1,2, \dots, m, j = 1,2, \dots, n$$

Decryption with ECC:

$$d_{ij} = D\left(e_{ij}, k_{ij}\right), i = 1,2, \dots, m, j = 1,2, \dots, n$$

The samples were taken from a data repository of Troy University, reflecting information about Academic programs. The sentences were 50 char long. There were 2 samples of ciphertext generated by each method, C1 generated by AES. and C2 generated by the private key of ECC

## IV. Cryptoanalysis

This method has been tested with one of the standard cryptoanalysis techniques such as statistical pattern matching. Finding the patterns was the role an ANN played. The input values were mapped to encrypted blocks. For every encrypted block of data, a mapping to an identifier was used. The block size was 2 bytes. A transformation to chars in Unicode was used as follows.

Encrypted block i (EB) => XY, where XY were Unicode char, for all i. This format fitted the ANN topology, the inputs to the ANN were in the form XY. The same method to assign a mapping was used for ECC and AES. The configuration of the ANN is as follows. We chose these settings (Input layer, hidden layer, and output layer) to best fit our encrypted mapped blocks.

### A. Axo-axonic Connections

ANN models use weighted activation transfer functions. Connection types such as axo-somatic, axo-axonic, and axon-synaptic [11–13] are widely used. This paper is focused on the second kind of axo-axonic connections types. The principle consists of propagating the action of neuron N3 as synapse S12, see Fig. 8.

To model the previous connection type, two neural networks were required [14–16]. This architecture was named Enhanced Neural Networks ENN [17–21].

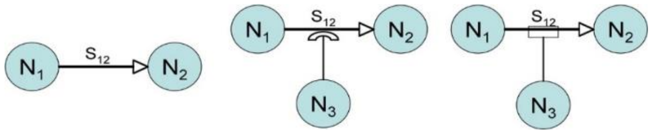


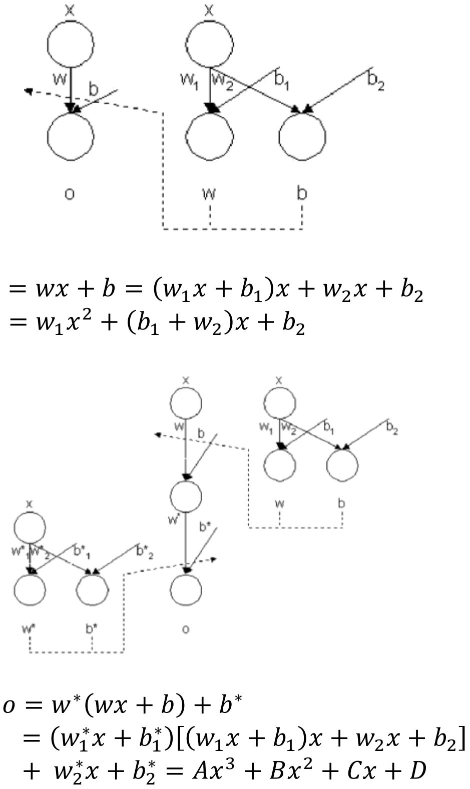
Fig. 8. Neural networks connections: axo-dendritic, axo-synaptic, and axo-axonic (from left to right).

### B. Taylor Approximation with Axo-axonic Networks

The formula to be used in the ANN for the objective function relates to the error in Taylor approximations..

$$|e(x)| \leq \frac{1}{(n+1)!} M|x-a|^{n+1} \tag{1}$$

Enhanced Neural Networks behave as n-degree polynomial approximators depending on the number of hidden layers in the architecture. To obtain such behavior, all activation functions of the net must be lineal function $f(x) = ax + b$.



$$o = wx + b = (w_1x + b_1)x + w_2x + b_2$$
$$= w_1x^2 + (b_1 + w_2)x + b_2$$

$$o = w^*(wx + b) + b^*$$
$$= (w_1^*x + b_1^*)[(w_1x + b_1)x + w_2x + b_2]$$
$$+ \ w_2^*x + b_2^* = Ax^3 + Bx^2 + Cx + D$$

Fig. 9. ENN architectures and output expressions.

As shown in Fig. 9 and output equations, the number of hidden layers can be increased to increase the degree of the output polynomial, that is, the number n of hidden layers control, in some sense, the degree n+2 of the output polynomial of the net. Table I shows how the degree of the output polynomial increases according to the number of hidden layers in the net.

TABLE I: NUMBER OF HIDDEN LAYERS VS. DEGREE OF OUTPUT POLYNOMIAL

| Hidden layers | Degree P (x) | Output Polynomial |
|---|---|---|
| **0** | **2** | $\boldsymbol{o = w_1x^2 + (b_1 + w_2)x + b_2}$ |
| 1 | 3 | $\boldsymbol{o = (w_1^*x + b_1^*)\begin{bmatrix} w_1x^2 + \\ (w_2 + b_1)x \\ +b_2 \end{bmatrix} + w_2^*x + b_2^*}$ |
| n | n+2 | $\boldsymbol{o = \sum_{i=0}^{n+2} a_ix^i}$ |

The only condition that the learning algorithm must verify is that weights must be adjusted to values related to the successive derivates of function $f(x)$ that the pattern set represents. Usually, such function is unknown therefore if the network converges with a low mean squared error then all weights of the net have converged to the derivates of function $f(x)$ (the pattern set unknown function), and such weights will gather some information about the function and its derivates that the pattern set represents. As an example, function $f(x) = \sin(x)\cos(x)$ can be approximated using Eq. (2), with a given point $a = 0$. Such an equation can be reduced to $\tilde{f}(x) = x - \frac{4}{6}x^3$, using a polynomial $P(x)$ degree 3. This is a mathematical approach, but what happens if such a function is the pattern set to an enhanced neural network mentioned before?

$$\hat{f}(x) = \sum_{i=0}^{n} \frac{f^i(a)(x-a)^i}{i!} + \frac{f^{n+1}(\xi)(x-a)^{n+1}}{(n+1)!} \quad (2)$$

A one-hidden-layer neural network was used to obtain a 3-degree polynomial as the output expression. Fig. 10 shows such architecture, after the training stage, the final configuration is shown. The output equation of the net is $o = x - \frac{4}{6}x^3$, equivalent equation with $\tilde{f}(x)$.

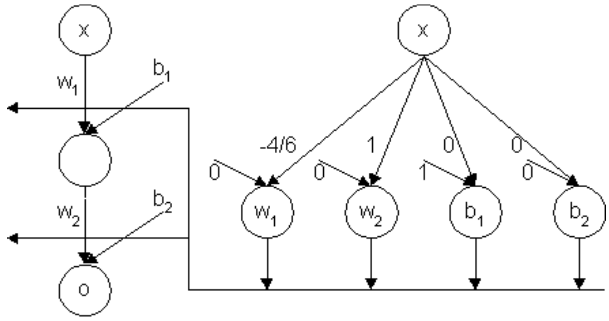


Fig. 10. Approximation of $f(x) = \sin(x)\cos(x)$ with a one hidden layer.

The approximation error using the net in Fig. 10 can be computed using Eq. (1), and therefore MSE $\leq |e(x)|$. The such approximation is not the only one nor the best one, but it can be computed theoretically to provide the net some initial weights to speed up the learning process and to obtain a better approximation that the initial one with a lower error ratio.

## V. RESULTS OF CRYPROANALYSIS

With this configuration, the ANN runs for 50 epochs.

The number of patterns found slightly varied depending on the algorithm used. Below Table II is the table that links the patterns found. We only used 1 hidden layer.

TABLE II: NUMBER OF PATTERNS FOUND IN BLOCKS OF ENCRYPTED TEXT

| | Blocks | Patterns Matched |
|---|---|---|
| AES | 255 | 12 |
| ECC | 255 | 5 |

Table II is simple and suggests that the Artificial Neural Network that inputs mapped ECC blocks potentially find fewer statistical patterns between them than encrypted blocks in AES. Parameters used for the ANN are described above in Section IV, using Taylor's approximation for fitting the objective function. After inputting the blocks in the ANN input layer, after 50 epochs, the highest difference in the found patterns is shown above.

## VI. CONCLUSION

ECC encryption system is an asymmetric key encryption algorithm for public-key cryptography. It simply generates a public and private key and allows two parties to communicate securely. There is one major advantage however that ECC offers over RSA. 256 bits key in ECC offers about the same security as 3072 bits key using RSA. In this paper, we create a more secure way to store and retrieve database information that has been secured using Form Maintaining Encryption.

This configuration, although simplistic, provides enough information to draw some conclusions as the pattern found suggests some correlation with potential vulnerabilities. These results can be seen as one more indicator when choosing the right algorithm to use in database applications along with FME This ensures that FME can now be used without fear of losing encryption keys to AES or unwanted interference from malicious third parties. We conclude that FME will take advantage of ECC encrypting vs AES.

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

Alberto Arteta: Manuscript review, rephrasing, clarifying and reconnecting the sections. Akshay Harshakumar: Original idea and preliminary proposal. Luis F. Mingo: Calculation of patterns results in the Neural network. Vijaya Kumar Gurumoorthy: Format, edition and technical proofreading. All authors have approved the final version.

## References

[1] N. Aljuaid, A. Gutub, and E. Khan "Enhancing PC data security via combining RSA cryptography and video based steganography," *Journal of Information Security and Cybercrimes Research (JISCR)*, vol. 1, no. 1, pp. 5–13, 2018.

[2] A Alsaidi, K. Al-Lehaibi, H. Zahrani, and M. Al-Ghamdi, "Compression multi-level crypto stego security of texts utilizing colored email forwarding," *Journal of Computer Science & Computational Mathematics (JCSCM)*, vol. 8, no. 3, pp. 33–42, 2018.

[3] N. Alanizy, A. Alanizy, N. Baghoza, M. Al-Ghamdi, and A. Gutub, "3-Layer PC text security via combining compression, aes cryptography 2LSB image steganography," *Journal of Research in Engineering and Applied Sciences (JREAS)*, vol. 3, no. 4, pp. 118–124, 2018.

[4] M. Alotaibi, D. Al-hendi, B. Al-Roithy, and M. Al-Ghamdi, "Secure mobile computing authentication utilizing hash, cryptography and steganography combination," *Journal of Information Security and Cybercrimes Research (JISCR)*, vol. 2, no. 1, pp. 9–20, 2019.

[5] M. Alkhudaydi and A. Gutub, "Securing data via cryptography and arabic text steganography," *SN Computer Science*, vol. 2, no. 46, 2021.

[6] A. Gutub and N. Al-Juaid, "Multi-bits stego-system for hiding text in multimedia images based on user security priority," *Journal of Computer Hardware Engineering*, vol. 1, no. 2, pp. 1–9, 2018.

[7] F. Al-Shaarani and A. Gutub, "Securing matrix counting-based secret-sharing involving crypto steganography," *Journal of King Saud University — Computer and Information Sciences*, vol. 34, no. 9, pp. 6909–6924, 2022.

[8] N. Kheshaifaty and A. Gutub, "Engineering graphical captcha and AES crypto hash functions for secure online authentication," *Journal of Engineering Research*, 2021.

[9] E. S. B Hureib and A. Gutub, "Enhancing medical data security via combining elliptic curve cryptography with 1-LSB and 2-LSB image steganography," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 20, no. 12, pp. 232–241, 2020.

[10] Yasmeen, "NoSQL database engines for big data management," *International Journal of Trend in Scientific Research and Development (IJTSRD) International Open Access Journal*, vol. 2, issue 6, Sep.–Oct. 2018.

[11] D. Pritchett, "BASE: An acid alternative," *ACM Queue*, vol. 6, no. 3, pp. 48–55, 2008.

[12] A. R. Pathak, B. Padmavathi, "Survey of confidentiality and integrity in outsourced databases," *International Journal of Scientific Engineering and Technology*, vol. 2, no. 3, pp. 122–128, 2013.

[13] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in *Proc. the 35th Annual Symposium on Foundations of Computer Science*, 1994.

[14] L. F. Mingo, F. Arroyo, C. Luengo, and J. Castellanos, "Enhanced neural networks and medical imaging," in *Proc. the 8th International Conference on Computer Analysis of Images and Patterns*, 1999.

[15] L. F. Mingo, V. Giménez, and J. Castellanos, "Interpolation of boolean functions with enhanced neural networks," in *Proc. the Second Conference on Computer Science and Information Technologies*, 1999.

[16] L. F. Mingo, J. Castellanos, and V. Giménez, "A new kind of neural networks and its learning algorithm," in *Proc. the Information Processing and Management of Uncertainty in Knowledge-Based Systems Conference, IPMU'98*, 1998, pp. 1913–1914.

[17] H. E. Smith and M. Brightwell, "Using datatype-preserving encryption to enhance data warehouse security," in *Proc. the 20th National Information Systems Security Conference*, 1997, p. 141.

[18] M. Bellare, T. Ristenpart, P. Rogaway, and T. Stegers, "Format-preserving encryption," *Lecture Notes in Computer Science*, vol. 45, no. 5, pp. 295–312, 2009.

[19] M. Luby and C. Rackoff, "How to construct pseudorandom permutations from pseudo-random functions," *Siam Journal on Computing*, vol. 17, no. 2, pp. 373–386, 1988.

[20] J. Delacour, *Apprentissage et Memoire: Une Approache Neurobiologique*, Masson Ed. September, 1987.

[21] L. F. Mingo, F. Arroyo, C. Luengo, and J. Castellanos, "Learning HyperSurfaces with neural networks," in *Proc. the 11th Scandinavian Conference on Image Analysis*, 1999, pp. 731–737.