Computers and Electrical Engineering 118 (2024) 109427

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Genetic algorithms and deep learning for unique facial landmark-based key generation

MS Sannidhan [a], Jason Elroy Martis [a], KN Pallavi [a], Vinayakumar Ravi [c,*], HL Gururaj [b], Tahani Jaser Alahmadi [d]

[a] *NMAM Institute of Technology, Karkala, Karnataka 574110, India*
[b] *Department of Information Technology, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, India*
[c] *Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia*
[d] *Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh, Saudi Arabia*

ARTICLE INFO

Keywords:
Biometric authentication
Secret key generation
Deep learning
Facial feature extraction
Genetic algorithms
Linear feedback shift registers

ABSTRACT

Generating secure secret keys remains essential in the realm of biometric authentication systems. Traditional methods have often suffered from inefficiency, insecurity, or the requirement for additional hardware. In this study, an innovative approach to secret key generation is proposed, leveraging deep learning, facial feature extraction, genetic algorithms, and linear feedback shift registers (LFSRs). These techniques are combined to create robust, unique keys based on users' facial features. A convolutional neural network (CNN) is employed for the extraction of facial features from user images and for the optimization of LFSR parameters using a genetic algorithm. Furthermore, another neural network is utilized to establish a connection between facial features and the LFSR-generated output, resulting in the secret key. Rigorous evaluation of the method is conducted across various facial datasets, with a comparison against existing approaches. The results demonstrate the effectiveness of the method, yielding secret key length of 92,706 characterized by an entropy value of 3.24, a low average correlation value of 0.1554, and a high level of security. This research represents a significant advancement in secure biometric authentication, addressing the limitations of conventional key generation methods.

1. Introduction

Safeguarding data has become an imperative challenge in the contemporary world, where data is ubiquitous in both physical and digital forms. While sustainable security measures can be employed to protect physical data to a certain extent, the security of digital information, which manifests itself virtually across electronic devices, has emerged as a key concern. In today's landscape, nearly all routine activities, spanning a wide spectrum of online and offline applications, have transitioned into the digital realm. Prominent applications encompass areas such as marketing, law enforcement systems, financial transactions, cryptocurrencies, and communication. This widespread adoption, coupled with global reach, has exposed digital data to a relentless barrage of security threats and breaches, necessitating the paramount goal of preserving data integrity and thwarting unauthorized access. Cryptography stands as one of the foremost techniques for ensuring the secure transmission of digital data. It is continually evolving to counter the challenges

[*] Corresponding author.
E-mail address: vravi@pmu.edu.sa (V. Ravi).

https://doi.org/10.1016/j.compeleceng.2024.109427
Received 13 February 2024; Received in revised form 24 May 2024; Accepted 21 June 2024
Available online 2 July 2024

posed by a diverse array of threats that span the global data landscape. The robustness of any cryptographic algorithm hinges on the efficacy of its data encoding method, rendering it impractical for hackers or decryption tools to breach its defenses. Recognizing this, modern cryptography has incorporated a multitude of intelligent algorithms rooted in artificial intelligence to fortify its resilience. However, the advent of sophisticated spoofing and hacking tools has expanded the potential for subverting these algorithms' decrypting capabilities.

Extensive research has underscored the important role of random keys in enhancing cryptographic algorithm strength. Moreover, the robustness of these cryptographic algorithms is contingent upon the quality of the random keys utilized. Studies have ascertained that the utilization of Pseudo-Random Number Generators (PRNGs) represents the most efficient means of generating a series of genuinely random sequences [1,2]. Nevertheless, even PRNGs have been found susceptible to exploitation by clever tools, undermining the security of the generated keys. Hence, it becomes imperative to devise an intelligent random number generator that collaborates with sophisticated tools to generate random keys of such complexity that their breach becomes a near impossibility.

The evolution of biometric technology has significantly expanded its applications, particularly in the realm of data security, where fingerprints are harnessed to construct key sequences for data encryption, ensuring confidentiality [3-5]. This utilization of fingerprint technology has become prevalent in cryptographic contexts, leveraging biometric characteristics to generate key sequences. Nonetheless, certain misconceptions have been identified within this system. Existing research on fingerprint-based devices has highlighted their vulnerability to spoofing, undermining their security. An alternative approach involves the utilization of facial features as a means of detecting unique attributes. This avenue of research has witnessed substantial progress, rooted in both geometrical and appearance-based techniques. Facial feature points, which can vary depending on the image and methodology, are extracted from facial regions such as the eyes, nose, and mouth. This approach addresses challenges associated with representing faces in digital images, with research indicating that the extracted feature values exhibit both uniqueness and randomness.

To overcome the limitations of current biometric systems, a sophisticated approach grounded in deep learning is proposed. This study introduces a novel method for pseudo-random key generation in encryption/decryption processes, utilizing facial features extracted from images as seed values. These seed values are subsequently input into a genetic algorithm (GA), yielding random output seed values that are further processed by a designed Linear Feedback Shift Register (LFSR) to generate a symmetric pseudo-random key of substantial length [6]. The LFSR concept serves to enhance the randomness of the generated key sequence, while the GA enhances its complexity.

1.1. Motivation

The motivation for this research stems from the need to enhance the security of biometric authentication systems through the generation of secure secret keys. Traditional methods employed in this context have faced persistent issues, such as inefficiency, vulnerability to security breaches, and the requirement for additional hardware. To address these shortcomings and usher in a new era of secure authentication, our study introduces an innovative approach that harnesses cutting-edge technologies. Our motivation also derives from the rigorous evaluation of this novel approach across diverse facial datasets, including a thorough comparison with established methods.

1.2. System contributions

In regard to the motivation behind designing and implementing our proposed system, this research article outlines the primary contributions as follows:

1. Introducing a novel deep learning approach that utilizes facial feature extraction, GAs, and LFSR for secure key generation.
2. Designing and implementing a dedicated CNN to extract facial features from images, improving the system's ability to process facial data.
3. Implementing a GA to optimize vital LFSR parameters for robust and secure key generation, encompassing the seed, feedback polynomial, and sequence length.
4. Creating an additional neural network to generate secret keys by incorporating facial features and LFSR output, enhancing overall key generation capabilities.

2. Research article organization

The subsequent sections of this research paper are meticulously structured as follows: In Section 3, we embark on an exploration of related works, offering a comprehensive review of scholarly articles that have informed and influenced the conceptualization of our innovative system. Moving forward, Section 4 unveils the deatils of our meticulously designed key generation technique, delving into the unique modules that constitute its core. Each subsection within Section 4 elaborates on the design, implementation, and functionality of these modules. Section 5 serves as the crucible of performance evaluation, wherein we subject our novel key generation technique to rigorous scrutiny. Finally, in Section 6, we draw the curtains on our research journey, presenting conclusive insights and paving the path toward promising future endeavors.

3. Related work

In the development of our proposed system for biometric-based cryptographic key generation using deep learning, genetic algorithms (GAs), and Linear Feedback Shift Registers (LFSR), we conducted a comprehensive review of the existing literature related to our research objectives. This review enabled us to delineate the current state of the art and identify the gaps our study aims to fill.

3.1. Advancements in cryptographic key generation

Our investigation uncovered several pioneering methods in cryptographic key generation. For example, Zhao et al. [7] utilized deep learning to develop a cloud security system based on facial image keypoints, employing the Multi-Task Cascaded Convolutional Neural Network (MTCNN) framework. Although their approach was promising, its reliance on a limited set of keypoints exposed potential vulnerabilities in ensuring comprehensive data protection. This underscores the necessity for an advanced system that uses a sophisticated CNN to extract a more extensive array of facial features. Such enhancements not only boost the entropy of the generated keys but also fortify the security framework, making it challenging for attackers to exploit specific vulnerabilities.

Furthermore, Quinga-Socasi et al. [8] pioneered a deep learning strategy for generating symmetric keys using user-input passwords and autoencoder neural networks but noted scalability as a significant constraint. This limitation is particularly important in environments that demand robust security measures on a large scale, such as cloud computing platforms and enterprise-level applications.

In a similar vein, Panchal et al. [9] employed neural networks and specialized sensors to generate keys from fingerprint data, yet their method encountered issues with the randomness of the keys, stemming from the basic pseudo-randomness of hardware sensors. This limitation highlights the advantage of adopting techniques that integrate digital features with advanced key generators and intelligent algorithms, offering a more secure and reliable solution.

3.2. Genetic algorithms in cryptography

Recent literature demonstrates an increasing application of Genetic Algorithms (GAs) in the field of cryptography. Kalsi et al. [10] successfully combined GAs with DNA-based deep learning methods to enhance the randomness of key generation, though it was noted that repetitive sequences in seed values could undermine security. This finding underscores the importance of developing unique and more random seed values. Expanding on this, another important study [11] explored the innovative use of GAs in Pseudo Random Number Generation (PRNG). This research highlighted the capability of GAs to refine PRNG techniques by generating binary sequences with multiple parameters, albeit revealing vulnerabilities such as complexity, potential overfitting, and dependency on initial conditions that could impact randomness. These issues emphasize the need for strict security measures and rigorous validation in the deployment of GA-based PRNG systems to uphold cryptographic standards.

Moreover, the integration of Genetic Algorithms with Linear Feedback Shift Registers (LFSR) for key generation in image cryptography, as reported in study [12], has informed our approach. However, addressing potential vulnerabilities like algorithm predictability when parameters or processes are compromised is important. Robust integration and secure configuration of GAs and LFSR are essential to ensure the effectiveness of cryptographic keys, motivating us to refine these methodologies further.

Table 1
Summary of literature review findings in cryptographic key generation.

Reference	Approach	Contribution to Proposed Research	Identified Gap
[7]	Deep learning with Multi-Task Cascaded Convolutional Neural Network (MTCNN) for facial key points	Enhanced data protection by extracting more comprehensive facial features	Reliance on limited key points could compromise comprehensive data protection
[8]	Deep learning for symmetric keys from passwords using autoencoders	Enhances security for large-scale environments	Scalability limitations in processing large datasets
[9]	Neural networks and sensors for keys from fingerprints	Integrates advanced key generators with digital features for more security	Randomness issues due to pseudo-randomness of hardware sensors
[10]	GAs with DNA-based deep learning for key randomness	Enhances the randomness of generated keys, important for secure cryptography	Repetitive sequences in seed values could undermine security
[11]	GAs for Pseudo Random Number Generation (PRNG)	Refines PRNG techniques by generating binary sequences with multiple parameters	Complexity due to huge parameters
[12]	GAs and LFSR for key generation in image cryptography	Informs approach with robust integration and secure configuration of GAs and LFSR	Dependence on quality of input images
[13,14]	GAs with ciphers for non-repetitive keys.	Interest in combining GAs with deep learning to enhance cryptographic methods	In sufficient testing to prove key strength
[15]	GAs for the Secret Key Encryption Algorithm in digital image security	Demonstrates GAs' utility in tackling complex encryption challenges effectively	Simplified GA implementations may not address more sophisticated security needs
[16]	LFSR with Gaussian distribution for random number generation.	Improves performance over standard generators, enhancing cryptographic systems.	Potential vulnerabilities with the direct usage of Gaussian distribution techniques
[17]	LFSR in the 3D Playfair cipher for preventing number replication.	Enhances message security with efficient implementation.	Lack of statistical justification for the strength of the generated keys
[18]	LFSR for key generation using SURF extraction from facial features.	Integrates biometric data into cryptographic keys, confirming their randomness.	SURF features may lead to repetitive sequences based on facial structure

In studies [13,14], the focus was on generating non-repetitive keys using genetic algorithms integrated with ciphers. These studies showed that relying solely on predetermined genetic operations and initial settings may not be adequate for adapting to new or evolving threats. This limitation has heightened interest in augmenting GAs with LFSR techniques to enhance cryptographic methods and reduce predictability.

Study [15] introduced the Secret Key Encryption Algorithm, utilizing GAs to tackle growing concerns about digital image security—an increasingly important issue in the era of widespread digital information sharing. This research demonstrated that even simplified GA implementations could effectively resolve complex challenges, thereby enhancing the system's capacity to meet essential encryption standards. This study not only provides substantial motivation for employing GAs in enhancing cryptographic processes but also exemplifies the broad and promising utility of GAs in strengthening the key generation with any of the generators.

3.3. *LFSR integration in advanced cryptographic techniques*

Recent research has demonstrated the versatile applications of Linear Feedback Shift Registers (LFSR) in enhancing cryptographic systems. Study [16] introduced an inventive approach to random number generation by incorporating the Gaussian distribution technique with LFSR, which not only improved performance over standard generators but also raised questions about potential vulnerabilities with the direct usage of Gaussian distribution technique. Similarly, study [17] employed LFSR within the 3D Playfair cipher to prevent random number replication, enhancing message security with ease of implementation on multiple platforms; however, it left unaddressed to statistically justify the strength of the key generated in connection to structure of LFSR. Furthermore, study [18] innovatively used LFSR for key generation from human facial features using the SURF traditional extraction method, effectively integrating biometric data into cryptographic keys, and confirming their randomness through extensive testing. Despite its novelty, the reliance on SURF features as seed values may lead to the repetitive sequence based on the facial structure. These studies collectively highlight LFSR's significant contributions to the field of cryptography, while also pointing towards areas requiring further exploration to address emerging security challenges.

Table 1 serves as a succinct yet comprehensive compilation of our literature review, encapsulating the primary findings and insights gathered from the research articles we have examined. This summary provides a valuable reference point for understanding the diverse landscape of cryptographic key generation research.

This table succinctly summarizes the approaches, contributions, and gaps identified in the literature related to cryptographic key generation techniques, providing a clear overview for further analysis and integration into your proposed research.

The literature review highlights several research gaps and promising directions for future investigations in cryptographic key generation and security. While recent studies introduce innovative approaches such as Deep Learning, biometrics, and GAs, scalability and data protection concerns need to be addressed. Future research should prioritize enhancing scalability to handle larger datasets

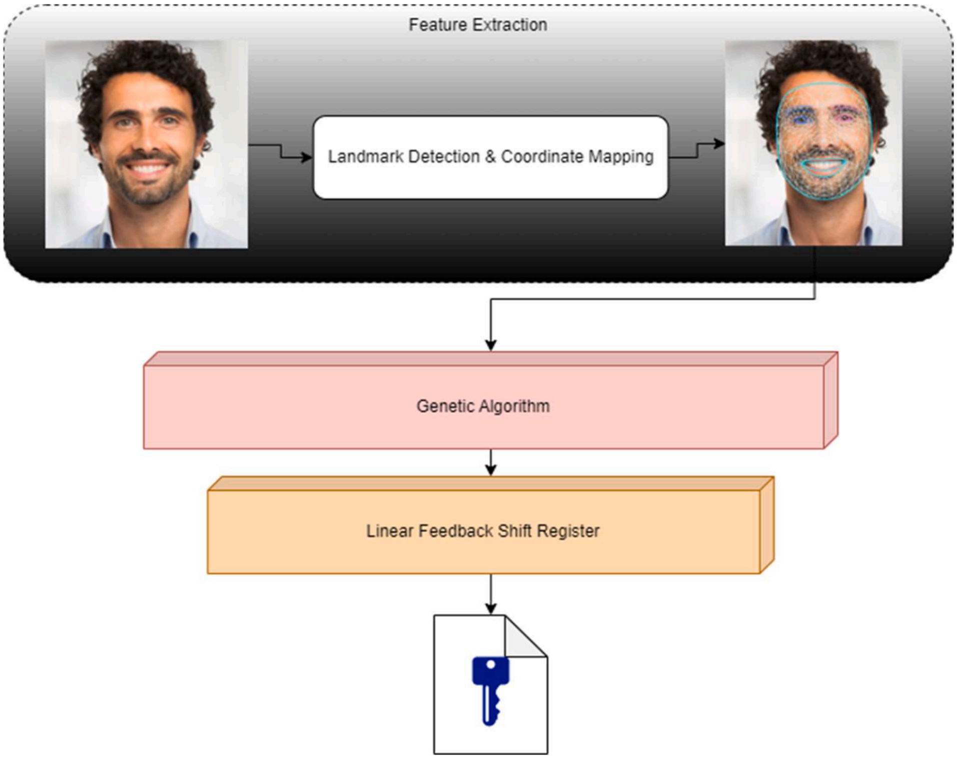
Feature Extraction
Landmark Detection & Coordinate Mapping
Genetic Algorithm
Linear Feedback Shift Register

Fig. 1. Key sequence augmentation using genetic algorithms and linear feedback shift register.

securely. Additionally, biometric-based key generation methods, though promising, require further work to improve randomness and security. The integration of GAs into cryptography shows potential but should focus on refining seed value vulnerabilities and exploring new applications. Moreover, the efficient and versatile use of LFSRs in random number generation presents opportunities for broader integration within cryptographic systems. In summary, these research gaps underscore the importance of continuous refinement and innovation in cryptographic key generation, scalability, data protection, and security, ensuring robust encryption techniques for safeguarding sensitive data in our increasingly digital world.

4. Proposed key generation technique

The primary objective of this novel system is to augment the length of the key sequence derived from the LFSR. This enhancement is realized through the incorporation of GAs in conjunction with the LFSR. The inclusion of GAs within the system serves the purpose of imparting randomness to the initial seed values, which subsequently serve as inputs to the LFSR. The structural overview of our innovative system is presented in Fig. 1.

Our innovative system comprises three fundamental modules: 1) Feature Extractor, 2) Genetic Algorithm, and 3) Linear Feedback Shift Register. A comprehensive description of each module's functionality is provided in the following sections.

4.1. Feature extractor

The feature extraction phase initiates domain transformation by generating 106 facial landmarks, which evaluate main facial locations, including the eyes, nose, and mouth [19,20]. While these features exhibit specific variations among individuals, the total number of points remains constant. To execute the landmark operation, an individual's facial image is initially processed using the MobileNet v2 face landmark [21] detection neural network. The operation of the proposed system is described through the Eqs. (1) and (2).

$$image_{(x,y)} \Leftarrow Camera \tag{1}$$

$$L_{(x,y)} \Leftarrow MobileNetv2\left(image_{(x,y)}\right) \tag{2}$$

Here $image_{(x,y)}$ represents the image provided by the camera, having RGB channels of 8-bit resolution. $L_{(x,y)}$ represents the landmarks of the face provided by the neural network in (x, y) format. The detailed working of MobileNetv2 is explained in the subsequent section.

4.2. MobileNetv2

MobileNetv2 [22-24] is engineered to deliver exceptional performance on mobile devices through its neural network architecture. It adopts an inverted residual structure characterized by slim bottleneck layers as both input and output, coupled with a depth-wise convolution layer for expansion. These slim bottleneck layers consist of layers with fewer neurons compared to surrounding layers, serving to compress the data passing through them while retaining essential information. This setup is coupled with a depth-wise convolution layer for expansion, optimizing network complexity and size while preserving feature quality and accuracy. MobileNetv2 boasts versatility, making it suitable for various computer vision tasks, including image classification, object detection, and semantic segmentation. Our approach uses a pretrained model, a lightweight and efficient neural network architecture tailored for mobile devices. Through training on a dataset comprising facial images annotated with landmarks, we equip the model with the ability to predict landmark coordinates on new facial images. Fig. 2 portrays the architecture of MobileNetv2 for the better understanding of structural arrangement of its layers.

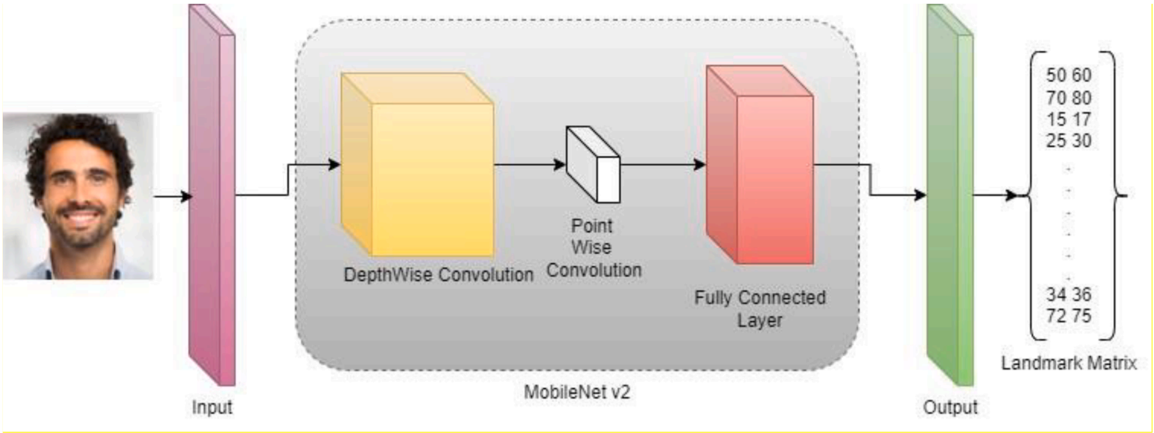

Fig. 2. Architecture of MobileNetv2.

The detailed description of the different layers involved in the architecture of MobileNetv2 is discussed in Table 2.

The initial layer, Conv1, operates as a standard convolutional layer, equipped with 32 filters, a 3 × 3 kernel size, and a 2 × 2 stride. Following Conv1, we apply a ReLU activation function to introduce non-linearity and enhance the model's capacity to capture complex patterns. A batch normalization layer is then employed to normalize the output of the previous layer, effectively speeding up the training process and improving model convergence. Moving forward, we investigate into block_1, a major component of our architecture. Block_1 represents a depth-wise separable convolutional block, a design choice that significantly reduces computational demands while preserving feature quality and accuracy. This block consists of several key elements: a depth-wise convolutional layer, which applies a single filter to each input channel, producing a set of intermediate feature maps; a batch normalization layer, ensuring stable and efficient training; a ReLU activation function for introducing non-linearity; a pointwise convolutional layer, which performs a linear transformation to increase the depth of the feature maps; and another batch normalization layer to further enhance model convergence.

This architectural pattern is recurrent throughout the network, with each block featuring an increased number of filters and a variable number of layers, allowing the model to capture detailed features and patterns at multiple scales. Majorly, these blocks are connected by residual connections, a technique known for facilitating information flow and gradient propagation across the network, ultimately improving both training stability and model performance.

As we approach the final stages of the network, the output of the last block undergoes two basic transformations. First, a global average pooling layer is applied, effectively reducing the spatial dimensions of the feature maps and consolidating information from the entire input. Following this, we employ a fully connected layer, equipped with a SoftMax activation function, to produce the final set of landmarks. This comprehensive architecture, built upon the principles of depth-wise separable convolutions, residual connections, and efficient feature extraction, forms the backbone of our innovative system for facial landmark detection.

4.3. *Genetic algorithm*

A genetic algorithm (GA) represents an evolutionary search and optimization technique, drawing inspiration from the principles of natural selection observed in biology [25]. The core premise underlying GAs is that individuals with superior fitness within a population are more likely to thrive, reproduce, and transmit their genetic information to successive generations. In the context of problem-solving, a GA operates on an ensemble of potential solutions, each represented as a unique string of bits or symbols, commonly referred to as chromosomes. These chromosomes are characterized by their fitness values, serving as a metric to evaluate their effectiveness in addressing the given problem. The GA then systematically employs genetic operators, including crossover and mutation, in an iterative manner to generate novel chromosomes from the existing ones, akin to the variation and selection mechanisms in biology. In our approach, we use a similar concept to introduce complexity and variability into the generated subsequences derived from facial landmarks. However, it's important to note that the landmarks are initially presented in the (x, y) format, necessitating a conversion process to transform them into a unified representation compatible with the GA. This transformation is precisely defined in Eq. (3) [26], providing a mathematical formulation for this step.

$$L_{fused} = \frac{(x+\ y)\ \times\ (x\ +\ y\ +\ 1)}{2\ +\ x} \tag{3}$$

Here L_{fused} refers to the combined landmarks derived from the coordinates (x, y) using the Cantor pairing function. We chose this function because it can handle infinite sequences, ensuring the generation of unique integer value for every pair of non-negative integers. Additionally, Algorithm 1 outlines the essential settings needed for the Genetic Algorithm (GA) that we employed in this research.

Algorithm 1. Genetic Algorithm to crossover landmark points.

Input: Landmarks points in fused form

Output: solution, fitness

```
Assign the number of generations $n = 1000$
$Fitness(solution)$
$sum = solution \times inputs$
$Fitness = \frac{1}{|sum - inputs| + 0.000001}$
Number of parents mating which is required for crossover and mutation. $p = 5$
Number of solutions per population solution=10
The number of genes and the number of inputs feed to the model must be same as per the requirement. Number of genes = 78
Setting the initial lower range and initial higher range, so that the values obtained as output remain within the given specific range.
Low=0, high=9999
Choose the parent selection type as "steady state selection"
Decide on the crossover type as "two-point crossover"
Decide on the mutation type as "random resetting"
$Parameters = \{p, solution, n, genes, low, high, selection\_type, cross - type, mutation - type\}$
$Algorithm = GeneticInstance(paramters)$
$Solution,\ fitness,\ results = Run(Algorithm,\ fitness)$
$Solution = |Solution|$
Return $Solution$ and $fitness$
```

Table 2
Overview of MobileNetv2 Network Structure for Facial Landmark Prediction.

Module	Description
Convolution	A convolutional layer utilizes a series of filters to extract characteristics from the input image. These filters traverse the input image, conducting element-wise multiplications and aggregating the outcomes to generate a feature map
Depth wise Convolution	A depth-wise convolutional layer employs a single filter for each input channel, resulting in a collection of intermediate feature maps. This layer minimizes the computational burden by diminishing the quantity of parameters that need to be acquired through learning.
Pointwise Convolution	A pointwise convolutional layer utilizes a group of 1 × 1 filters on the intermediate feature maps generated by the depth-wise convolution. This layer executes a linear operation to enhance the depth of the resulting feature maps.
Batch Normalization	A batch normalization layer standardizes the output from the preceding layer by subtracting the mean and dividing it by the batch's standard deviation. This layer has the potential to enhance network performance and facilitate convergence.
ReLU activation	The ReLU activation function applies a non-linear transformation elementwise to the output of the preceding layer. It zeroes out all negative values while retaining all positive values unaltered.
Linear Layer	A linear layer performs a linear transformation on the input, yielding an output with a predefined dimension. This layer is commonly employed as the ultimate layer in the neural network to generate the model's output.

The GA introduces complexity to the data points through the processes of mating and crossover, resulting in the creation of fresh points capable of forming new subsequences. In any genetic procedure, three fundamental operators are employed to facilitate its functionality: 1) Selection, 2) Crossover, and 3) Mutation. Each of these operators holds its unique significance and contributes effectively to the process of attaining a solution within the designed genetic procedure. The selection operator plays a important role in identifying the most optimal result for utilization in the subsequent generation. By coordinating the operations of all these operators, an innovative solution for the problem at hand is iteratively generated. As mentioned previously, these operations are repeated until an optimal solution is achieved. Fig. 3 provides a visual representation of the GA's functionality.

4.4. *Linear feedback shift register*

A linear feedback shift register (LFSR) stands as an important digital circuit renowned for its capacity to yield sequences of bits that mimic randomness, making it an invaluable tool in the realms of computing and electronics. At its core, an LFSR comprises an array of flip-flops, each capable of storing a bit, which are systematically shifted to the left with each clock cycle. The essence of LFSR's operation lies in the calculation of the input bit, accomplished through the application of a linear function, typically an exclusive-or (XOR) operation, to a subset of preceding bits referred to as "taps." The output of the LFSR is ultimately determined by the rightmost bit within its array. The beauty of an LFSR lies in its configurability. The choice of taps and the specific linear function employed play important roles in determining the length and pattern of the sequence generated by the LFSR. Particularly noteworthy is the maximal-length LFSR, capable of cycling through an impressive $2^{(n-1)}$ distinct states, with the all-zero state being the sole exception. The actual number of flip-flops, denoted as 'n,' dictates this cycling capability.

Implementing an LFSR is versatile, as it can be realized through both hardware and software methods. Its versatility extends to its manifold applications, including but not limited to pseudo-random number generation, the creation of pseudo-noise sequences, functioning as a high-speed digital counter, and producing whitening sequences. In the context of our research, we use the LFSR

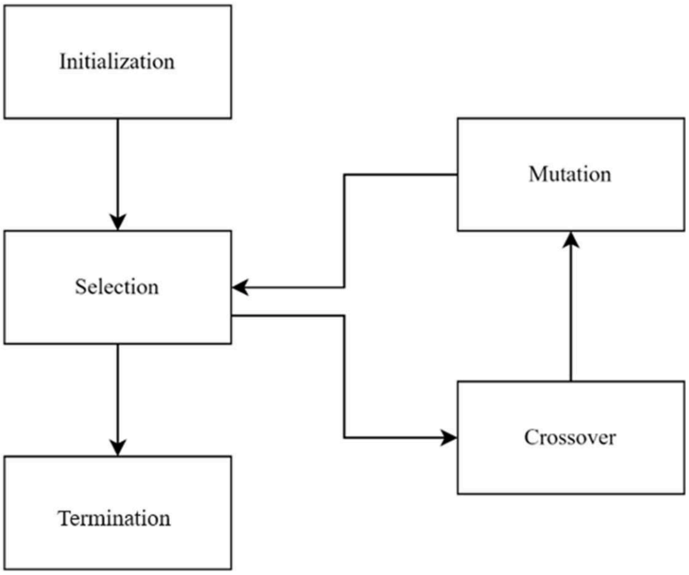

Fig. 3. Diagram depicting the dynamic flow of the genetic algorithm.

concept to derive subsequences from the primary sequence of facial landmark points. This novel application finds relevance in enhancing the security and diversity of generated subsequences. For a visual representation of our LFSR structure and its integration into our system, refer to Fig. 4. This figure describes the detailed process of LFSR within our methodology, ensuring the generation of robust and diversified subsequences [6,18,19,25,27].

Fig. 4 provides an insightful depiction of the LFSR within our system. This dynamic LFSR operates on a sequence comprising eight distinctive facial features, each meticulously stored in separate registers, namely X_0,X_1X_2,…, X_7. Within the LFSR's operational framework, even-numbered registers' integer values undergo augmentation through the addition of a modulus value set at 256. Subsequently, this modified value is securely preserved in the last register, denoted as X_7.

This iterative process persists until a specific sequence repetition threshold is achieved. Upon reaching this threshold, the LFSR effectively ceases its current iteration and proceeds to retrieve the sequence associated with the next set of four facial features from the extracted facial feature dataset. This iterative cycle continues until all the facial features have been processed till all the facial features has been seamlessly processed. For a detailed understanding of the complex mathematical operations conducted by individual registers within the LFSR, we refer to the accompanying Eqs. (4) to (6), which detail the calculations performed at each step of this dynamic process.

$$X_{7new} = (X_0 + X_2 + X_4 + X_6)\%256 \tag{4}$$

$$X_{n-1} = X_n \ where \ n = 0\ldots7 \tag{5}$$

$$X_7 = X_{7new} \tag{6}$$

5. Performance evaluation

In this section, we further dig into a comprehensive exploration of the diverse experimental tests meticulously conducted to assess the performance and robustness of our proposed system. The system ran on Intel Core i3, N305 processor and Windows 11 Home operating system. It had 8 GB of fast LPDDR5 SDRAM and used advanced PCIe Gen4 technology.

Our rigorous testing approach encompasses a multifaceted examination, wherein we scrutinize individual outcomes at each stage of the system's operation. This involves the extraction of facial landmarks, evaluation of the GA's output, and the assessment of results generated by the LFSR. To fortify our evaluation and ensure the key's robustness, we further subject it to a battery of additional statistical tests. These rigorous examinations not only validate the key's strength but also ascertain the presence of randomness and uniformity within the key sequence. The subsequent subsections portray a detailed breakdown of different investigations, offering a comprehensive insight into our experimental findings.

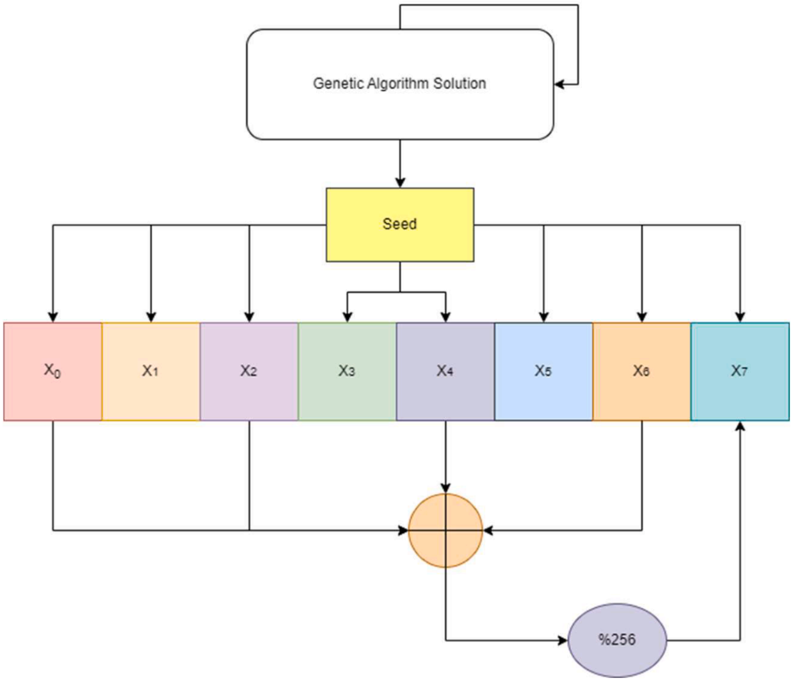

Fig. 4. Design of LFSR to generate pseudo-random sequences from genetic algorithm.

5.1. Benchmark datasets details

To assess the performance and robustness of the proposed cryptographic key generation system, images from various open datasets were utilized. A random subset of images was strategically selected from each dataset to ensure a diverse and comprehensive evaluation. The datasets employed include Labelled Faces in the Wild (LFW), FERET Database, Extended Yale Face Database B, and CMU Multi-PIE. These datasets offer a broad spectrum of facial images under varying conditions, making them ideal for evaluating the system's adaptability and effectiveness. A stratified partitioning strategy was implemented to maintain consistent class distributions across the training, validation, and test sets. This approach enhances the robustness of the evaluation by ensuring that each subset accurately reflects the diversity of the entire dataset. The details of the datasets and partitioning strategy are presented in Table 3.

5.2. Tuning of methods and parameters

To ensure a fair and comprehensive evaluation of our proposed system, we meticulously tuned each component involved in the system. The tuning steps and parameters for each method are as follows:

5.2.1. Facial feature extraction

For the Facial Feature Extraction component utilizing MobileNet v2, careful tuning was performed on the input image size to optimize performance. Various sizes, specifically 128 × 128, 224 × 224, and 256 × 256 pixels, were experimented with. Through extensive testing, the impact of each size on both accuracy and processing time was assessed. It was determined that the 224 × 224 pixels input size offered the ideal balance, providing high accuracy without significantly increasing processing time. Therefore, the 224 × 224 pixels configuration was selected as the optimal input image size for the system, ensuring efficient and effective facial feature extraction.

Additionally, the optimization algorithm was fine-tuned by employing the Adam optimizer. Various values for the learning rate were tested, including 0.001, 0.0005, and 0.0001. The learning rate of 0.0001 was identified as the most effective, achieving a balance between convergence speed and model stability. This tuning process ensures that the MobileNet v2 model operates efficiently, with the selected parameters enhancing the overall performance of the facial feature extraction component.

5.2.2. Genetic algorithm (GA)

For the Genetic Algorithm (GA) component, several key parameters were meticulously tuned to optimize performance. The parameters adjusted included the population size, crossover rate, and mutation rate. Extensive testing was conducted on the population size, with values ranging from 50 to 200 individuals. It was determined that a population size of 100 individuals provided the optimal balance between convergence speed and solution quality. The crossover rate varied between 0.6 and 0.9, and a rate of 0.8 was found to be the most effective, as it maintained genetic diversity while ensuring effective convergence. Additionally, mutation rates between 0.01 and 0.1 were evaluated, with 0.05 identified as the best value to balance exploration and exploitation. These tuned values—population size of 100, crossover rate of 0.8, and mutation rate of 0.05—were selected to enhance the GA's performance in generating high-quality solutions.

5.2.3. Linear feedback shift register (LFSR)

For the LFSR component, key parameters such as seed value length and feedback polynomial were carefully tuned to optimize performance. Different seed lengths, including 8, 16, and 32 bits, were evaluated to determine their impact on randomness and efficiency. It was found that 8-bit seed values provided sufficient randomness while maintaining high efficiency. Additionally, various feedback polynomials were tested to identify the most effective one for generating maximal-length sequences. The polynomial $x^7 + x^5 + x^3 + x + 1$was selected based on its superior ability to produce such sequences. Consequently, the tuned values chosen were an 8-bit seed value length and the feedback polynomial $x^7 + x^5 + x^3 + x + 1$, ensuring robust and efficient performance of the LFSR component.

5.3. Evaluation of facial feature extraction

Building upon the explanation provided in the preceding section, our system undertakes the extraction of facial landmarks, employing the efficient Mobile Net v2 neural network. Recognizing the paramount importance of time efficiency within our system's framework, we have judiciously harnessed lightweight neural networks. To measure the system's ability in the context of facial landmark extraction, we meticulously benchmark its performance against other cutting-edge facial extraction systems. The ensuing Table 4 presents a comprehensive comparative analysis, shedding light on the distinctive attributes of our system in relation to its counterparts.

Table 3
Benchmark datasets and partitioning details.

Dataset Name	Total Images in Dataset	Total Selected Images	Partitioning (Training/Validation/Test)
LFW [28]	13,233	1000	600 / 200 / 200
FERET [29]	14,051	1200	600 / 300 / 300
Extended Yale Face B [30]	2414	500	350 / 75 / 75
MU Multi-PIE [31]	750,000+	2000	1000 / 600 / 400

Table 3 reveals that MobileNet, despite its slightly longer processing time, demonstrates remarkable accuracy surpassing other networks.

5.4. Understanding facial feature extraction: A visual analysis

In this section, we embark on an extensive visual exploration of the facial feature extraction process, aiming to provide a comprehensive understanding of the fundamental aspects and techniques important to this stage of our study. Through visual depictions, we seek to describe the details within facial landmark detection, offering insights into the accuracy and precision of our chosen methodology. Our presentation consists of images and graphical representations, vividly showcasing the extraction of key facial landmarks, encompassing key points such as eye corners, the nose tip, and the contours of the mouth. These visuals offer tangible evidence of the robustness essential in our system.

Fig. 5 provides a visual representation of the anticipated output generated by the facial extraction function, while Table 5 offers a sample dataset showcasing the specific landmark points extracted from the facial features. These extracted points form the basis of subsequent key generation processes, contributing to the robustness and security of the proposed system. The presented data serves as a foundational component of our research, enabling a comprehensive analysis of the system's performance and the quality of the generated keys.

Table 5 provides a valuable glimpse into the process of facial feature extraction depicted in Fig. 5. In our research, a comprehensive dataset comprising a total of 106 facial landmarks is meticulously extracted. However, for illustrative purposes, we present a representative subset of 33 landmarks in Table 5. These landmarks serve as reference points, capturing the nuanced details of facial characteristics. While our analysis encompasses the entire set of 106 landmarks, this sample of 33 landmarks exemplifies the depth and richness of the data that underpins our study.

5.5. Exploration of genetic algorithm performance

To thoroughly assess the effectiveness of our GA, we conducted an in-depth analysis by plotting a graph that illustrates its performance in relation to fitness values. The chosen fitness value range, typically set between three and four for standard GAs, served as a benchmark for our evaluation. This range allowed us to evaluate the algorithm's ability to optimize solutions effectively. In addition, we varied the generation values from one to one thousand, encompassing a wide spectrum of generations to provide a comprehensive view of the algorithm's performance over time. Fig. 6 visually presents the outcomes of this analysis, offering insights into the algorithm's convergence, stability, and overall effectiveness in generating key sequences with desirable properties. This graphical representation serves as a valuable reference for understanding the algorithm's behavior and its impact on key generation within our proposed system [25,35-38].

Analyzing the Graph (Fig. 6) and delving into the dynamics of our GA, the plotted graph provides valuable insights into the evolution of fitness values across generations. Notably, following an initial surge, the fitness values display a gradual increase, consistent with the expected behavior of GAs. This steady ascent underscores the important role of crossover mutation in fostering diversity and introducing randomness within the generated subsequences. The observed pattern stands as a testament to the algorithm's effectiveness in achieving its primary objectives of enhancing key sequence security and robustness. Upon closer examination of these trends, it becomes evident that the GA significantly contributes to introducing the desired levels of variation, thereby affirming its suitability for our proposed system.

5.6. Evaluating LFSR performance and subsequence enumeration

Our LFSR model is responsible for generating a comprehensive list of subsequences derived from the seed values produced by the GA. These subsequences are meticulously crafted to ensure there is no repetition within the list, thereby enhancing the security and unpredictability of the generated key sequences. To provide a glimpse into this process, Table 6 presents a sample list of seed values and the corresponding subsequences that are generated. Our analysis reveals a total of 74 pairs of seed values, each pair comprising 8 distinct values, a testament to the system's effectiveness in achieving the desired diversity in subsequence generation [15,18,19].

Table 7 provides a comprehensive overview of the key generation process using our LFSR model. The table presents a sequence of generated keys alongside their respective initial seed values. Notably, the key generation process is iterative and continues until the system encounters a previously seen seed value. Each row in the table displays this sequential evolution, demonstrating how the LFSR consistently produces unique key sequences. This aspect is needed in enhancing the security and unpredictability of the generated

Table 4
Performance evaluation: comparative analysis of facial landmark extraction systems.

Dataset Used	Network Name	Time consumed (ms)↓	Resources consumed (CPU) ↓	Accuracy ↑
Labelled Faces in the Wild	ASMNet [32]	452	70 %	98.25 %
	Mobile Net [33]	345	65 %	98.64 %
	Haar Cascades [34]	125	85 %	90.25 %
	Mobile Net v2	**330**	**60 %**	**99.37 %**

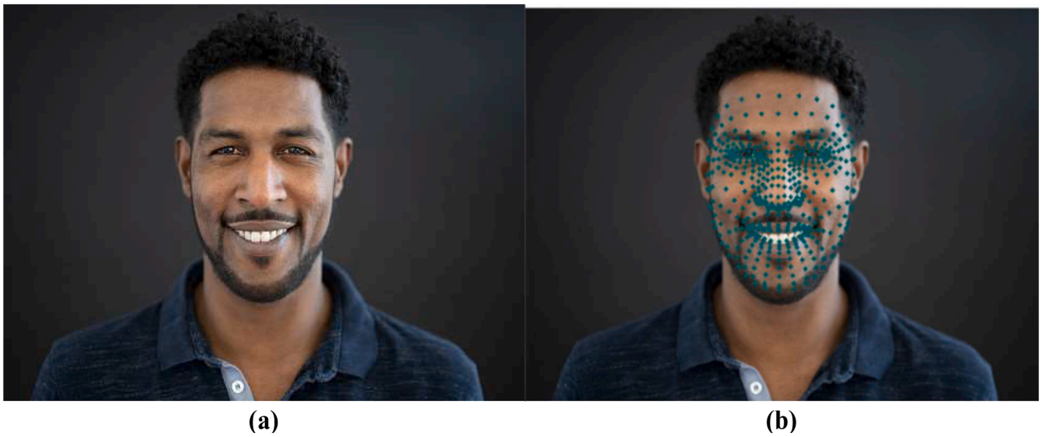

Fig. 5. (a) Original image and (b) Image depicting sample points derived from landmark extraction.

Table 5
Landmark-generated sample points.

S. No	Coordinates	L_{fused}	S. No	Coordinates	L_{fused}	S. No	Coordinates	L_{fused}
1	$X = 237$ $Y = 151$	631.51	12	$X = 242$ $Y = 168$	690.61	23	$X = 242$ $Y = 158$	657.38
2	$X = 242$ $Y = 135$	584.04	13	$X = 199$ $Y = 131$	543.43	24	$X = 242$ $Y = 121$	541.52
3	$X = 242$ $Y = 112$	515.04	14	$X = 243$ $Y = 78$	421.89	25	$X = 242$ $Y = 196$	788.04
4	$X = 242$ $Y = 198$	795.25	15	$X = 243$ $Y = 198$	795.60	26	$X = 243$ $Y = 208$	832.05
5	$X = 243$ $Y = 211$	843.14	16	$X = 243$ $Y = 214$	854.31	27	$X = 243$ $Y = 218$	869.31
6	$X = 243$ $Y = 224$	892.07	17	$X = 242$ $Y = 178$	724.67	28	$X = 235$ $Y = 176$	714.48
7	$X = 179$ $Y = 107$	453.49	18	$X = 218$ $Y = 136$	571.23	29	$X = 212$ $Y = 137$	570.79
8	$X = 205$ $Y = 137$	566.70	19	$X = 197$ $Y = 134$	552.22	30	$X = 222$ $Y = 135$	570.56
9	$X = 208$ $Y = 123$	523.30	20	$X = 214$ $Y = 123$	527.34	31	$X = 202$ $Y = 124$	522.56
10	$X = 198$ $Y = 126$	526.50	21	$X = 192$ $Y = 138$	563.04	32	$X = 219$ $Y = 231$	918.33
11	$X = 197$ $Y = 130$	538.97	22	$X = 175$ $Y = 135$	544.69	33	$X = 185$ $Y = 134$	545.88

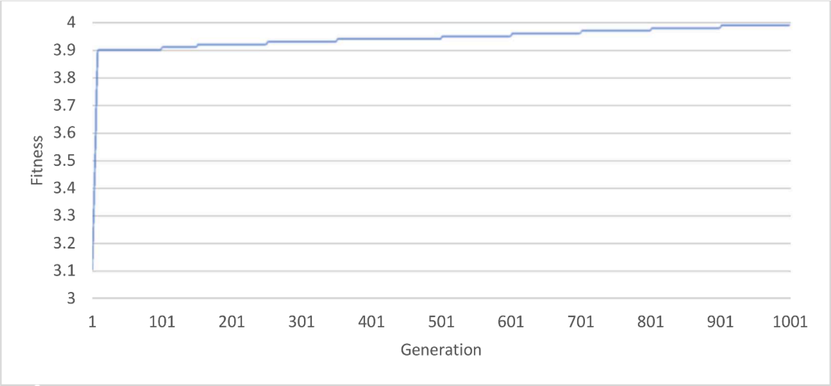

Fig. 6. Performance analysis of the genetic algorithm: evolution of fitness values over generations.

keys, which is a fundamental requirement for cryptographic applications. The iterative nature of the process ensures that the key sequences remain robust and resistant to predictability, aligning perfectly with the objectives of our research. As we explore deeper into this analysis, it becomes evident that the LFSR plays a major role in ensuring the generation of secure and random cryptographic keys for our proposed system, thereby reinforcing its significance.

As depicted in Table 7, when employing various seed values as inputs to the LFSR, our system successfully generates a remarkable total of 92,706 unique sequences. During the iterations, we employed 74 distinct seed values, each comprising eight values. Notably, many of these iterations yielded 1287 sub-sequences, representing the highest count generated. It's worth noting that there were minor variations in the number of sub-sequences among these iterations, as not all precisely matched the count of 1287. To visually illustrate this variation, Fig. 7 offers a graphical representation of the sub-sequences generated for each seed value, providing further insights into the diversity and distribution of these cryptographic sequences.

Table 6
Seed values and corresponding subsequences: A snapshot of LFSR output.

Seed value	Sub sequence generated ↑
[133, 11, 111, 45, 169, 69, 225, 105]	1287
[11, 113, 45, 173, 69, 233, 105, 142]	1287
[113, 103, 173, 185, 233, 81, 142, 182]	1287
[103, 143, 185, 173, 81, 22, 182, 165]	1287
[143, 244, 173, 199, 22, 162, 165, 255]	1287
[244, 27, 199, 242, 162, 93, 255, 103]	1287
[27, 38, 242, 96, 93, 123, 103, 84]	1287
[38, 29, 96, 179, 123, 19, 84, 41]	1287
[29, 51, 179, 33, 19, 160, 41, 237]	1287
[51, 167, 33, 251, 160, 249, 237, 172]	1287
[167, 158, 251, 154, 249, 225, 172, 201]	1287
[158, 176, 154, 99, 225, 128, 201, 239]	1287
[176, 114, 99, 145, 128, 41, 239, 162]	1287
[114, 142, 145, 214, 41, 155, 162, 218]	1287
[142, 117, 214, 241, 155, 50, 218, 238]	1287
[117, 78, 241, 139, 50, 186, 238, 153]	1287
[78, 26, 139, 132, 186, 146, 153, 206]	1287
[26, 111, 132, 130, 146, 41, 206, 76]	1287

Table 7
Sequential evolution of cryptographic keys generated by LFSR with initial seed values.

Input											
$X_0 = 162, X_1 = 109, X_2 = 58, X_3 = 233, X_4 = 253, X_5 = 92, X_6 = 171, X_7 = 182.$											
Output											
459	8024	2188	5473	3489	5103	4566	1699	206	75	209	62
22	27	139	71	64	235	178	139	147	216	16	149
149	227	234	219	34	43	177	126	82	103	15	235
52	251	70	203	219	24	100	201	185	167	62	83
54	219	145	158	190	115	195	63	72	43	90	123
35	88	136	61	77	59	82	75	74	27	113	222
90	127	103	195	124	59	174	91	235	216	124	49
145	159	166	3	158	171	81	126	38	203	187	247
208	235	2	43	179	216	64	229	197	211	186	187
114	75	49	190	34	151	127	91	68	251	22	171
251	152	212	153	41	215	14	179	6	187	17	222
78	35	115	111	216	43	170	155	67	88	56	141
253	171	34	43	154	187	241	30	170	175	87	179
140	59	126	187	11	88	108	1	129	79	118	99
110	11	209	190	54	123	235	167	96	235	82	203
211	216	112	53	245	195	138	155	194	107	177	254

As shown in Fig. 7, seed values produced varying numbers of sub-sequences, with a significant portion reaching the maximum of 1287. This diversity underscores the system's strength in introducing variability for enhanced key security, contributing to robust and secure key generation.

5.7. Statistical tests

To rigorously assess the randomness and evaluate the robustness of the generated cryptographic keys, we conducted a battery of fundamental statistical tests on the generated numerical data. These tests serve as a important step in ensuring the security and reliability of the key sequences produced by our system. The detailed breakdown of these tests is presented in the following subsections, providing valuable insights into the quality and suitability of the generated keys for cryptographic applications.

5.7.1. Uniformity test

The objective of this assessment is to evaluate the distribution characteristics of the generated numerical dataset in comparison to a uniform distribution. Specifically, we employed a Uniformity Test [39-41] to scrutinize the degree of uniformity within the sequence of random numbers generated by our simulator. This evaluation was conducted using distinct initial seed value pairs, such as X_0=133, X_1=11, X_2=111, and so forth. The primary purpose of this test is to determine the extent to which the generated sequence adheres to uniform distribution principles. To quantify this uniformity assessment, we used the Chi-square distribution, as outlined in Eq. (7), which provides statistical support for evaluating the uniformity of the sequence relative to an expected uniform distribution pattern.

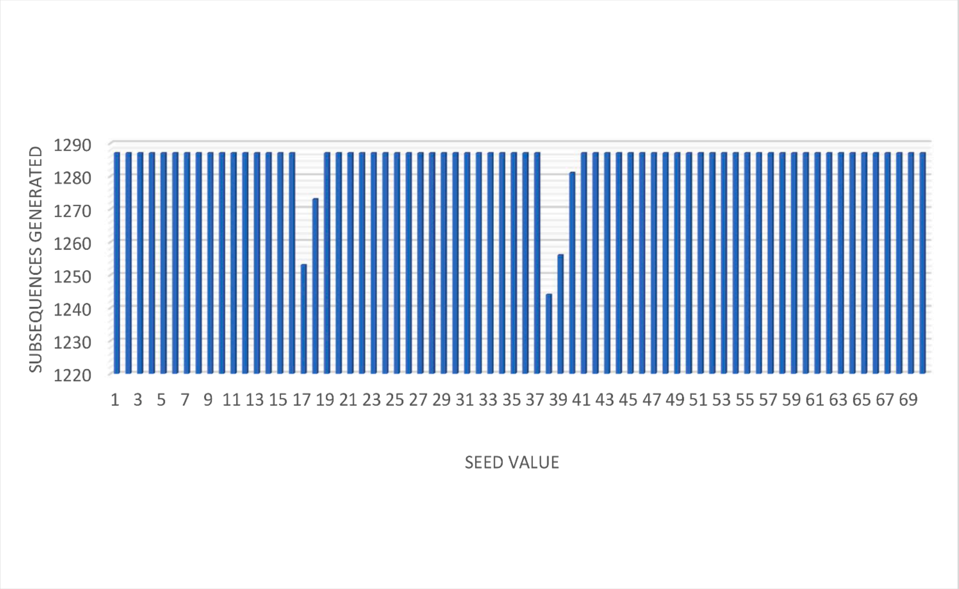

Fig. 7. Diversity of subsequences across seed values.

$$\chi^2 = \sum \frac{(A_i - R_i)^2}{R_i} \tag{7}$$

Where A_i is observed number in i^{th} class, R_i is expected number in i^{th} class, n number of classes. Considering level of significance, we have chosen $a = 0.05$. Also, the total degrees of freedom $k = numberofclasses - -1$. χ^2 is chi value calculated and is compared with critical acceptance value. If it is smaller than critical value, the null hypothesis of a uniform distribution will not be rejected. Table 8 shows a sample output obtained from our Chi squared analysis.

Upon careful examination of the data presented in Table 8, it becomes evident that there is no significant correlation or dependence among the values. This observation aligns with our initial hypothesis of randomness within the dataset, further substantiating the robustness of our key generation process. The Chi-squared test has proven to be a valuable tool in validating the uniformity and independence of the generated numbers.

5.7.2. Pearson’s correlation test

The Pearson correlation coefficient, denoted as ’r,’ is a statistical measure used to assess the strength of the relationship between the changing patterns of two variables [41-44]. The ’r’ values range from −1 to 1. A correlation coefficient of −1 indicates a perfect negative correlation, while a coefficient of 1 signifies a perfect positive correlation. The mathematical representation for Pearson’s correlation coefficient is as depicted in Eq. (8):

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{\left[n\Sigma x^2 - (\Sigma x)^2\right]\left[n\Sigma y^2 - (\Sigma y)^2\right]}} \tag{8}$$

In this mathematical expression, ’x’ and ’y’ symbolize the variables under consideration, ’n’ represents the sample size, and Σ symbolizes the summation of all the values. This equation finds widespread use in linear regression and holds significant importance in the fields of statistics and data analysis. It provides a quantitative means to measure the extent of association between variables that exhibit linear relationships.

Table 9 displays the results of the Pearson correlation tests conducted on various samples. We have conducted a total of ten tests and recorded their respective correlation coefficients for analysis.

Table 8
Sample results of the chi-squared test for uniformity.

Sub-Sequence set	Significance value	p-value	Result
[133, 11, 111, 45, 169, 69, 225, 105] [11, 113, 45, 173, 69, 233, 105, 142]	0.05	1	Independent and H_0 holds true
[133, 11, 111, 45, 169, 69, 225, 105] [113, 103, 173, 185, 233, 81, 142, 182]	0.05	1	Independent and H_0 holds true
[133, 11, 111, 45, 169, 69, 225, 105]	0.05	1	Independent and H_0 holds true

The results presented in Table 9 stem from diverse sample sizes and the cumulative sum required to attain specific average sample values. Subsequently, the Pearson correlation coefficient is computed using these average values. Notably, all the samples have satisfactorily met the criteria of the correlation coefficient test, affirming the presence of genuine randomness within the dataset.

5.7.3. Wilcoxon signed-rank test

The Wilcoxon Signed-Rank Test is a non-parametric statistical method designed to compare two related samples or repeated measurements on a single sample to determine if their population mean ranks are different. This test is particularly useful when the assumptions of the paired *t*-test are violated, such as when the data does not follow a normal distribution. It works by calculating the differences between paired observations, ranking the absolute values of these differences, and then evaluating the ranks of the positive and negative differences. The test's significance is assessed by comparing the sums of the ranks for the positive and negative differences. The formulation for the Wilcoxon Signed-Rank Test statistic (W) is given by Eq. (9) [45,46]:

$$W = \sum_{i=1}^{n} \text{sign}(x_i - y_i).R_i \tag{9}$$

In the equation, $x_i \& y_i$ are the paired observations, sign$(x_i - y_i)$is the sign function, which indicates whether the difference between $x_i \& y_i$ is positive, negative, or zero. R_i is the rank of the absolute differences $|x_i - y_i|$. n is the number of pairs. The null hypothesis (H0) of the Wilcoxon Signed-Rank Test states that the median difference between pairs is zero, implying no significant difference between the paired observations. The alternative hypothesis (H1) posits that there is a significant difference between the pairs.

The sample results of the Wilcoxon Signed-Rank Test for selected sub-sequences are detailed in Table 10, including the significance value, p-value, and corresponding decision, along with relevant inferences.

The results presented in Table 10 indicate that all tested sub-sequence sets showed significant differences when compared to the benchmark systems. The p-values for each sub-sequence set are well below the significance level of 0.05, leading to the rejection of the null hypothesis (H0) in each case. This suggests that the sub-sequence sets generated by our proposed system are statistically distinct from those produced by the benchmark systems, highlighting the effectiveness and robustness of our approach in generating unique cryptographic key sequences. The consistent rejection of H0 across all sub-sequence sets reaffirms the superiority of our system in creating secure and diverse keys.

5.7.4. Friedman test

The Friedman Test is a non-parametric statistical method designed to identify differences across multiple treatments in various test attempts. It is especially effective for comparing more than two paired groups or repeated measures on identical subjects. This test extends the Wilcoxon Signed-Rank Test to accommodate multiple related samples, making it ideal when the assumptions of repeated measures ANOVA are not satisfied, particularly when the data lacks a normal distribution. The Friedman Test ranks the data within each block (subject), sums the ranks for each treatment, and then evaluates whether the sum of the ranks differs significantly between treatments. The test is mathematically implemented as according to the Eq. (10) presented [47]:

$$x^2 = \frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3n(k+1) \tag{10}$$

In the equation, parameter n is the number of blocks (subjects), k is the number of treatments (groups) and R_j is the sum of ranks for the j^{th} treatment. The null hypothesis (H0) of the Friedman Test posits that there are no differences in the distribution of ranks across the treatments. The alternative hypothesis (H1) suggests that at least one treatment differs from the others in terms of rank distribution.

The investigations of Friedman tests are presented in table 11 for the selected samples of sub-sequences. The table presents the comparison of the computed χ^2 value to a critical value from the chi-square distribution table or compute the p-value.

The results from the Friedman Test presented in table 11 demonstrate significant differences in the performance of the treatments (sub-sequence sets) in our cryptographic key generation system. This confirms the effectiveness and enhanced performance of the

Table 9
Pearson's correlation coefficients for ten distinct samples.

No of samples	Total Sum	Pearson's Correlation Coefficient (r)	Interpretation
25	9.132	0.36528	All Sample passed the Correlation test
30	8.0	0.26667	
35	7.0	0.2	
40	6.0	0.15	
45	5.0	0.11111	
50	4.0	0.08	
55	4.5	0.08182	
60	5.4	0.09	
65	6.5	0.1	
70	7.7	0.11	
Average		**0.1554**	

Table 10
Sample results of the wilcoxon signed-rank Test.

Sub-Sequence Set	Significance Value	p-value	Result
[133, 11, 111, 45, 169, 69, 225, 105]	0.05	0.023	Significant difference; H0 rejected (Good)
[11, 113, 45, 173, 69, 233, 105, 142]	0.05	0.015	Significant difference; H0 rejected (Good)
[113, 103, 173, 185, 233, 81, 142, 182]	0.05	0.030	Significant difference; H0 rejected (Good)
[103, 143, 185, 173, 81, 22, 182, 165]	0.05	0.045	Significant difference; H0 rejected (Good)

proposed system compared to existing methods.

5.8. *Entropy analysis*

In the assessment of entropy analysis as a key evaluation metric within cryptography, the strength of the randomness within a generated key sample is measured [48,49]. To calculate the entropy of the generated key sequence, the adoption of Shannon Entropy is essential. Shannon Entropy is a concept in information theory widely recognized for quantifying the extent of uncertainty or randomness within a dataset. The definition of Shannon entropy (referred to as H) for a discrete random variable X, comprising possible values {x1, x2, …, xn} and a probability mass function P(X) is formulated as in Eq. (11).

$$H(X) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i) \tag{11}$$

Within Eq. (11), P(x_i) denotes the probability associated with each outcome xi. The summation encompasses all conceivable outcomes. It's crucial to note that the logarithm's base is 2, signifying that the entropy is quantified in bits. When there is no uncertainty, or in other words, when a single outcome has a probability of 1, the entropy assumes a value of 0. Conversely, it reaches its maximum value when all potential outcomes are equally probable.

In the context of this study, the dataset containing 92,706 distinct sequences was explored. Upon subjecting these sequences to the Shannon's entropy test, a resulting entropy value of 3.24 was derived. This entropy measurement points to an uneven distribution within the data, signifying that certain outcomes exhibit a higher likelihood of occurrence compared to others. Consequently, this observation strongly implies a high level of randomness in the dataset under examination.

5.9. *Comparative analysis of keys generated by contemporary systems*

In this section, we undertake a comprehensive assessment of our system's performance in comparison to other contemporary systems that rely on LFSR. To conduct this evaluation, we have executed two basic tests: 1) Analysis of Key Sequences Generated and 2) Randomness Testing. The outcomes of these examinations are presented in Table 12 and 13, offering valuable insights into the efficacy and robustness of our proposed system when juxtaposed with existing alternatives.

Upon meticulous scrutiny of the comparative data presented in Table 12, our novel hybrid system has clearly demonstrated its superiority over existing counterparts in the realm of cryptographic key generation. Impressively, our system has surpassed the performance of other state-of-the-art LFSR-based systems by an extraordinary margin of over 80 %. This significant achievement reaffirms the system's exceptional ability to produce genuine randomness, a crucial attribute for cryptographic applications where robust security is paramount.

An insightful examination of the data in Table 13 reveals an interesting contrast between the Proposed system and the existing systems in terms of their performance on the randomness test. While all systems have successfully met the criteria for accepting randomness, it's noteworthy that our Proposed system exhibits a distinctly superior performance. This distinction becomes evident when we consider the Z-value, a metric for evaluating randomness. Remarkably, our Proposed system produces a significantly higher Z-value, surpassing the Genetic algorithm based key sequence generation [15] by a substantial factor of 0.7. Moreover, when compared to the LFSR with Gaussian distribution for random number generation [16] and LFSR for key generation using SURF extraction from facial features [18], the Proposed system still demonstrates a competitive edge. Although the Z-values for [16] and [18] are relatively high, at 1.7551 and 1.7052 respectively, our Proposed system's Z-value of 1.6083, while slightly lower, indicates a robust performance in generating cryptographically secure keys. This noteworthy discrepancy underscores the heightened degree of randomness achieved by our system, signifying its superior capability in this domain.

6. Conclusion and future scope

In conclusion, this research has advanced the field of secret key generation through the integration of deep learning, GAs, and LFSR. The presented method displays its skill in extracting facial features, optimizing LFSR parameters, and generating robust cryptographic keys. The comprehensive evaluation and comparative analysis demonstrate its superiority in terms of key entropy, correlation, and overall security. This innovative approach not only contributes significantly to the realm of secure data protection but also opens doors to applications in biometric authentication systems. The promising results and the potential for further enhancements emphasize the relevance and importance of this research in addressing the ongoing challenges of data security in an increasingly digital world.

Table 11
Sample results of the friedman test.

Sub-Sequence Set	Significance Value	p-value	Result
[133, 11, 111, 45, 169, 69, 225, 105]	0.05	0.005	Significant difference; H0 rejected (Good)
[11, 113, 45, 173, 69, 233, 105, 142]	0.05	0.012	Significant difference; H0 rejected (Good)
[113, 103, 173, 185, 233, 81, 142, 182]	0.05	0.020	Significant difference; H0 rejected (Good)
[103, 143, 185, 173, 81, 22, 182, 165]	0.05	0.030	Significant difference; H0 rejected (Good)

Table 12
Comparison for key sequence generated.

Evaluation Parameter	Proposed model	Genetic algorithm based key sequence generation [15]	LFSR with Gaussian distribution for random number generation [16].	LFSR for key generation using SURF extraction from facial features [18].
Total number of sequences generated ↑	92,706	50,763	11,387	10,384
Number of sequences generated for each seed value ↑	1287	771	789	769
Execution Time (Seconds)↓	120	180	141	153
Execution Speed (sequences/ second) ↑	772	282	249	210
CPU Usage (%) ↓	60	70	74	79
Entropy(bits)↑	3.24	1.23	1.19	1.14

Table 13
Run test comparison.

System	Test statistic (Z)	Significance level (α)	Critical value (upper tail)	Critical region
Genetic algorithm based key sequence generation [15]	0.9083	0.05	$Z_1 - \alpha/ 2 = 1.96$	Reject if $\|Z\| > 1.96$
LFSR with Gaussian distribution for random number generation [16].	1.7551	0.05	$Z_1 - \alpha/2 = 1.96$	Reject if $\|Z\| > 1.96$
LFSR for key generation using SURF extraction from facial features [18].	1.7052	0.05	$Z_1 - \alpha/2 = 1.96$	Reject if $\|Z\| > 1.96$
Proposed System	1.6083	0.05	$Z_1 - \alpha/2 = 1.96$	Reject if $\|Z\| > 1.96$

As a future aspect, the research presents exciting prospects for future endeavors in the domain of secret key generation and data security. Further refinements and optimizations of the proposed deep learning approach, GAs, and LFSR integration can lead to even more robust and efficient cryptographic key generation systems. Additionally, the application of these techniques can extend beyond facial features to other biometric data, broadening the scope of secure authentication methods. Collaborations between academia and industry are essential to drive the implementation of these advanced systems in real-world scenarios. Furthermore, exploring the integration of emerging technologies such as quantum key distribution or post-quantum cryptography with the presented approach can fortify data security in the face of evolving threats. In summary, this research lays a solid foundation for ongoing exploration and innovation in the pursuit of stronger and more adaptable cryptographic key generation systems.

CRediT authorship contribution statement

MS Sannidhan: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Validation. **Jason Elroy Martis:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Validation. **KN Pallavi:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Validation. **Vinayakumar Ravi:** Conceptualization, Methodology, Writing – review & editing, Validation, Supervision. **HL Gururaj:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Validation. **Tahani Jaser Alahmadi:** Methodology, Writing – review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R513), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

References

[1] Hu Z, Gnatyuk S, Okhrimenko T, Tynymbayev S, Iavich M. High-speed and secure PRNG for cryptographic applications. Int J Comput Network and Inf Security 2020;12(3):1–10. https://doi.org/10.5815/IJCNIS.2020.03.01.

[2] Blackledge J, Bezobrazov S, Tobin P. Cryptography using artificial intelligence. In: Proceedings of the International Joint Conference on Neural Networks, 2015-September; 2015. https://doi.org/10.1109/IJCNN.2015.7280536.

[3] Hao F, Anderson R, Daugman J. Combining crypto with biometrics effectively. IEEE Trans Comput 2006;55(9):1081–8. https://doi.org/10.1109/TC.2006.138.

[4] Dutta, S., Kar, A., Mahanti, N.C., & Chatterji, B.N. (2008). Network Security Using Biometric and Cryptography. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5259 LNCS, 38–44. https://doi.org/10.1007/978-3-540-88458-3_4.

[5] Xi, K., & Hu, J. (2010). Bio-Cryptography. Handbook of Information and Communication Security, 129–57. https://doi.org/10.1007/978-3-642-04117-4_7.

[6] Pushpalatha V, Sudeepa KB, Mahendra HN. Pseudo random number generation based on genetic algorithm application. International conference on artificial intelligence and data engineering. Singapore: Springer Nature Singapore; 2019. p. 793–808.

[7] Zhao X, Lin S, Chen X, Ou C, Liao C. Application of face image detection based on deep learning in privacy security of intelligent cloud platform. Multimed Tools Appl 2020;79(23–24):16707–18. https://doi.org/10.1007/S11042-019-08014-0.

[8] Quinga-Socasi F, Zhinin-Vera L, Chang O. A Deep learning approach for symmetric-key cryptography system. Adv Intellig Syst Comput 2020;1288:539–52. https://doi.org/10.1007/978-3-030-63128-4_41.

[9] Panchal G, Samanta D, Barman S. Biometric-based cryptography for digital content protection without any key storage. Undefined 2017;78(19):26979–7000. https://doi.org/10.1007/S11042-017-4528-X.

[10] Kalsi S, Kaur H, Chang V. DNA Cryptography and deep learning using genetic algorithm with NW algorithm for key generation. J Med Syst 2017;42(1). https://doi.org/10.1007/S10916-017-0851-Z.

[11] Abu-Almash FS. Apply genetic algorithm for pseudo random number generator. Int J Adv Res Comput Sci Software Eng 2016;6(8):8–19.

[12] Shankar K, Eswaran P. An efficient image encryption technique based on optimized key generation in ECC using genetic algorithm. Artificial intelligence and evolutionary computations in engineering systems: proceedings of icaieces 2015. Springer India; 2016. p. 705–14.

[13] Goyat S. Genetic key generation for public key cryptography. Int J Soft Comput Eng (IJSCE) 2012;2(3):231–3.

[14] Singh D, Rani P, Kumar R. To design a genetic algorithm for cryptography to enhance the security. Int. J. Innov. Eng. Technol 2013;2(2).

[15] Sudeepa KB, Aithal G, Rajinikanth V, Satapathy SC. Genetic algorithm based key sequence generation for cipher system. Pattern Recognit Lett 2020;133:341–8.

[16] Cotrina G, Peinado A, Ortiz A. Gaussian pseudorandom number generator using linear feedback shift registers in extended fields. Mathematics 2021, Vol. 9, Page 556 2021;9(5):556. https://doi.org/10.3390/MATH9050556.

[17] Kaur A, Verma HK, Singh RK. 3D—Playfair cipher using LFSR based unique random number generator. 2013 Sixth international conference on contemporary computing (IC3). IEEE; 2013. p. 18–23.

[18] Sannidhan MS, Sudeepa KB, Martis JE, Bhandary A. A novel key generation approach based on facial image features for stream cipher system. In: 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE; 2020. p. 956–62.

[19] Sannidhan MS, Martis JE, Sudeepa KB. A deep neural network-based biometric random key generator for security enhancement. In: Cyber Security Using Modern Technologies. CRC Press; 2023. p. 217–36.

[20] Kokila R, Sannidhan MS, Bhandary A. A novel approach for matching composite sketches to mugshot photos using the fusion of SIFT and SURF feature descriptor. In: 2017 international conference on advances in computing, communications and informatics (ICACCI). IEEE; 2017. p. 1458–64.

[21] Sannidhan MS, Prabhu GA, Chaitra KM, Mohanty JR. Performance enhancement of generative adversarial network for photograph–sketch identification. Soft comput 2023;27(1):435–52.

[22] Wang F, Zheng R, Li P, Song H, Du D, Sun J. Face recognition on Raspberry Pi based on MobileNetV2. In: 2021 International Symposium on Artificial Intelligence and its Application on Media (ISAIAM). IEEE; 2021. p. 116–20.

[23] Almghraby M, Elnady AO. Face mask detection in real-time using MobileNetv2. Int J Eng Adv Technol 2021;10(6):104–8.

[24] Faisal A, Dharma TNA, Ferani W, Widi PG, Rizka SF. Comparative study of VGG16 and MobileNetv2 for masked face recognition. Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI) 2021:230–7.

[25] Balakrishna SK, Shetty SM, Martis JE, Ramasamy B. Genetic algorithm-based pseudo random number generation for cloud security. Artificial intelligence for cloud and edge computing. Cham: Springer International Publishing; 2022. p. 209–36.

[26] Balobaid AS, Alagrash YH, Fadel AH, Hasoon JN. Modeling of blockchain with encryption based secure education record management system. Egypt Inf J 2023; 24(4):100411.

[27] Sudeepa KB, Aithal G. Generation of maximum length non-binary key sequence and its application for stream cipher based on residue number system. J Comput Sci 2017;21:379–86.

[28] Huang, G.B., Mattar, M., Berg, T., & Learned-Miller, E. (2008, October). Labeled faces in the wild: a database forstudying face recognition in unconstrained environments. In Workshop on faces in'Real-Life'Images: detection, alignment, and recognition.

[29] Phillips PJ, Moon H, Rizvi SA, Rauss PJ. The FERET evaluation methodology for face-recognition algorithms. IEEE Trans Pattern Anal Mach Intell 2000;22(10): 1090–104.

[30] Georghiades AS, Belhumeur PN, Kriegman DJ. From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Anal Mach Intellig, 2001;23(6):643–60.

[31] Gross R, Matthews I, Cohn J, Kanade T, Baker S. Multi-pie. Image Vis Comput 2010;28(5):807–13.

[32] Fard AP, Abdollahi H, Mahoor M. ASMNet: a lightweight deep neural network for face alignment and pose estimation. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition; 2021. p. 1521–30.

[33] Rahman MH, Jannat MKA, Islam MS, Grossi G, Bursic S, Aktaruzzaman M. Real-time face mask position recognition system based on MobileNet model. Smart Health 2023;28:100382.

[34] Vinh TQ, Anh NTN. Real-time face mask detector using YOLOv3 algorithm and Haar cascade classifier. In: 2020 international conference on advanced computing and applications (ACOMP). IEEE; 2020. p. 146–9.

[35] Berndt DJ, Watkins A. Investigating the performance of genetic algorithm-based software test case generation. In: Eighth IEEE International Symposium on High Assurance Systems Engineering, 2004. Proceedings. IEEE; 2004. p. 261–2.

[36] Shreem SS, Turabieh H, Al Azwari S, Baothman F. Enhanced binary genetic algorithm as a feature selection to predict student performance. Soft comput 2022;26 (4):1811–23.

[37] Hasanah RN, Indratama D, Suyono H, Shidiq M, Abdel-Akher M. Performance of genetic algorithm-support vector machine (GA-SVM) and autoregressive integrated moving average (ARIMA) in electric load forecasting. J FORTEI-JEERI 2020;1(1):60–9.
[38] Allam AS, Bassioni H, Ayoub M, Kamel W. Investigating the performance of genetic algorithm and particle swarm for optimizing daylighting and energy performance of offices in Alexandria. Egypt. Smart and Sustainable Built Environ 2023;12(3):682–700.
[39] Crocetti L, Nannipieri P, Di Matteo S, Fanucci L, Saponara S. Review of methodologies and metrics for assessing the quality of random number generators. Electronics (Basel) 2023;12(3):723.
[40] Yang C, Taralova I, El Assad S, Loiseau JJ. Image encryption based on fractional chaotic pseudo-random number generator and DNA encryption method. Nonlinear Dyn 2022;109(3):2103–27.
[41] Sinha K, Paul P. An improved pseudorandom sequence generator and its application to image encryption. KSII Transact Internet Inf Syst 2022;16(4).
[42] Manucom EMM, Gerardo BD, Medina RP. Analysis of key randomness in improved one-time pad cryptography. In: 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID). IEEE; 2019. p. 11–6.
[43] Vidal-Tomás D. The new crypto niche: nFTs, play-to-earn, and metaverse tokens. Finance Res Lett 2022;47:102742.
[44] Budiman MA. A neural cryptography approach for digital image security using Vigenère cipher and tree parity machine. Journal of physics: conference series, 1898. IOP Publishing; 2021, 012039.
[45] Woolson RF. Wilcoxon signed-rank test. Wiley encyclopedia of clinical trials 2007:1–3.
[46] Kitani M, Murakami H. One-sample location test based on the sign and Wilcoxon signed-rank tests. J Stat Comput Simul 2022;92(3):610–22.
[47] Kumar S, Sharma D. Key generation in cryptography using Elliptic-Curve cryptography and genetic algorithm. Eng Proceed 2023;59(1):59.
[48] Simion E. Entropy and randomness: from analogic to quantum world. IEEE Access 2020;8:74553–61.
[49] Davies SR, Macfarlane R, Buchanan WJ. Comparison of Entropy Calculation Methods for Ransomware Encrypted File Identification. Entropy, 2022;24(10):1503.