SPECTRUM OF ENGINEERING SCIENCES
ISSN (E) 3007-3138 (P) 3007-312X VOLUME 2, ISSUE 4, 2024

DATA ENCRYPTION THROUGH BOOLEAN VALUED INFORMATION SYSTEM USING DATA AGGREGATES AND GENETIC ALGORITHM (ENCRYPT BISGA)

Muhammad Sadiq Khan[1], Attaullah[*2], Feroz Ali[3], Adila Tasnim[4], Nizamud Din[5], Muhammad Uzair Khan[*6]

[1,2,3,4,5]*Department of Computer Science, University of Chitral, Chitral, 17200, KP, Pakistan*
[*6]*Department of Mathematics University of Buner Swari, Buner KP, Pakistan*

[*6]uzair@ubuner.edu.pk

DOI: https://doi.org/10.5281/zenodo.23393470

Keywords
Boolean valued information system; Soft set; Fuzzy data; Genetic algorithm, Data encryption; Decision-making

Article History
Received: 01 October 2024
Accepted: 15 November 2024
Published: 30 November 2024

Copyright @Author
Corresponding Author: *
Muhammad Uzair Khan, Attaullah

Abstract
Boolean valued information systems (BIS) illustrates an advanced structure grounded in soft set theory, devoted to handling the complexities of decision-making considering uncertainty. The application of genetic algorithms (GA), commonly referred to as BISGA in academic research, is significant in reconstructing BIS from its aggregates. This paper represents an innovative exploration of BISGA in the realm of data encryption, presenting a novel approach for securing English words. The suggested approach leverages the advantages of BISGA for the encryption and decryption of data, ensuring both confidentiality and integrity. This article presents a comprehensive assessment of the BISGA-based encryption technique, supplemented by a thorough illustration and initial findings. We demonstrate the essential algorithms of the technique, detailing the procedures for the encryption and decryption of English words. The innovative application of BISGA in data encryption demonstrates its potential as a robust and efficient security framework. Moreover, this research contributes to the existing body of knowledge by exploring the feasibility of BISGA in data encryption, paving the way for future investigations into its applications in various domains. The findings of this study will produce considerable implications for the advancement of secure data transmission systems.

1. INTRODUCTION

In some real-life problems, we can't apply Cantor's traditional "hard" set theory, e.g., when there is uncertainty in the membership of elements. To deal with these problems different techniques were presented, e.g., Fuzzy sets (Zadeh, 1965), Rough Sets (Pawlak, 1996), Vague Sets (Bustince & Burillo, 1996), Probability theory (Rényi, 2007). All these methods have some limitations, i.e., they can't handle parameterization. To deal with such problems the concept of Soft Sets was presented by Molodtsovin 1999(Molodtsov, 1999).One application of Soft Set is, Boolean Valued Information System (BIS), which aligns the values of Soft Set in tables in bit format (Ali, Khan, Shah, Garg, & Alsheddy, 2023). It has many applications in various fields e.g., Games Theory(Deli & Cagman, 2013), Decision Making, Medical Diagnosing(Majumdar & Samanta, 2008), and Financial Problems(Kharal, 2010), etc.Data can be lost in BIS due to different factors e.g. Virus Attacks, Improper Data entry, or other Factors. Such systems are called Incomplete Information Systems (IIS). They need to be recalculated to make them usable again. They have two categories. They are Preprocessed Categories and Unprocessed

Categories (Mohd. Rose, Hassan, Awang, Herawan, & Deris, 2011). In the Processed Category Data about the existing data is stored before the data loss occurs and, that recorded information about the data is used to reconstruct the original data. The preprocessed techniques include Parity bits (Mohd. Rose et al., 2011), Supported Set, and Aggregates Set (Rose et al., 2011). A heuristic method is presented by Khan et al which used available aggregates to recalculate the entire BIS(Khan, Herawan, Wahab, Mujtaba, & Al-Garadi, 2016).Salman et al [6] presented another such Preprocessed Category, which recalculates the whole BIS using 4 aggregates.TheUnprocessed category uses the available information to recalculate the data that has been lost. This category includes Probability (Kong et al., 2014), Weighted Average (Zou & Xiao, 2008), DFIS (Qin, Ma, Herawan, & Zain, 2012), and ADFIS (Sadiq Khan, Al-Garadi, Wahab, & Herawan, 2016).A simplified method of data filling in the soft set is also presented, which presents a simpler way to recalculate the missing data in the Soft set(Kong, Lu, Wang, & Guo, 2023).The unprocessed approaches can only recalculate the missing values. Salmanet al. introduced a new preprocessed technique of, BooleanValued Information System recalculation from Aggregates using Genetic Algorithm (BISGA). This approach uses Genetic Algorithm (GA) and a set of 4 aggregates (Row Aggregates, Column Aggregates, Left – Right Diagonal Aggregates, and Right – Left Diagonal Aggreges) to recalculate the entire BIS. This approach selects 4 chromosomes from the available population with a special constraint, which makes the solution close to the actual results making it powerful than other methods (Ali et al., 2023). It completely regenerates the lost (unknown) data from aggregates which makes it applicable to data integrity (Ali et al., 2023). Hence, this concept can be applied in cryptography.Genetic Algorithm (GA) is a bioinspired algorithm, which was presented in 1975 by JH Holland (J. Holland, 1975). It is utilized to address optimization problems. It is based on Darvin's theory of Survival of the Fittest model. GA performs three main genetic processes or operations to produce optimal results. These operations are Selection, Crossover, and Mutation. Selection involves the selection of chromosomes (Individuals) from the population for the next generation, by evaluating their fitness value. Chromosomes that have large fitness values would have a higher chance of being selected. Crossover involves the mixing of two or more chromosomes to produce a new offspring (solutions). Mutation Involves diversifying the solutions by making changes to the individual solutions. In BISGA, GA is applied to reproduce the lost data, this concept can be used in Data Encryption for security purposes. Encryption is the process of converting plain text into cipher text. It scumbles the human-readable plain text so unintended people won't be able to understand it. Encryption can be categorized into two main categories: classical and modern. The classical ciphers are simpler and, follow the simple techniques of Substitution and Transposition. Modern ciphers are more complex and combine Substitution and Transposition to provide strong security. Examples of Classical Cyphers are Monoalphabetic and Polyalphabetic ciphers. Examples of Modern Cyphers are Data Encryption Standards (DES), Advanced Encryption Standards (AES), Rivest Cipher 4 (RC4), etc. (Goyal & Kinger, 2013).In the base paper BISGA, a new way of recalculating the entire data in a Boolean Valued Information System Using Genetic Algorithm is presented. That approach presented a novel way of restoring lost data in BIS, but its potential applications in other fields like Data Encryption remain unexplored. This paper seeks to demonstrate how the concept of BISGA can be applied to data encryption. To address the mentioned issue, this paper presents an application of BISGA in data encryption. Here the data is initially processed to calculate the aggregates, and those aggregates would act as the encrypted data. Those data will be sent to the receiver. Upon receiving the aggregates as the encrypted data,the receiver produces the original text after processing those aggregates. The proposed study presents a new data encryption technique using Genetic Algorithm through aggregates of BIS. Examples have been presented to show how this method works. This approach is implemented in Python and experiments on data of different lengths have been conducted to show how this method encrypts and decrypts the data, and what behaviors it shows when the

length of the data is increased. The highlighted contribution of this study is to apply the concept of BISGA in data encryption.

This paper has been divided into 5 sections. Section 2 containsthe literature review. Section 3 covers the methodology. It contains pseudo-codes and solved examples. Section 4presents the Results and adiscussion of the proposed method. Section 5 contains the conclusion and final remarks.

2. Literature Review

This section has four subsections. The Soft Sets literature review is included in the first subsection. The second subsection discusses the Genetic Algorithms (GA). The literature review of Boolean-valued information systems using Genetic Algorithms (BISGA) is presented in the third subsection. The final subsection discusses data encryption.

2.1 Soft Set: Soft Set is defined as,Let's say we have a universal set U which shows a collection of cars U = {C1, C2,C3,C4,C5} e.g., Toyota, Suzuki, Honda, Tesla, Ford, and A = {A1,A2,A3,A4} represents attributes or parameters, e.g., affordable, comfortable, electric, high speed.

Now let's say affordable cars are C1, C2, and C3, comfortable cars are C4, and C5, electric cars are C4, and high-speed cars are C4 and C5. C4 has most of the features that make it the most appropriate choice.

$$(\mathrm{F},\mathrm{A}) = \begin{Bmatrix} (f_1,\{\mathrm{C1},\mathrm{C2},\mathrm{C3}\}) \\ (f_2,\{\mathrm{C4},\mathrm{C5}\}) \\ (f_3,\{\mathrm{C4}\}) \\ (f_4,\{\mathrm{C4},\mathrm{C5}\}) \end{Bmatrix}$$

Boolean Valued Information System (BIS) is an application of a Soft set thataligns the values of the soft set in tables in bit format(Ali et al., 2023). BIS has many applications, e.g., games theory(Deli & Cagman, 2013), decision making, medical diagnosing(Majumdar & Samanta, 2008), and financial problems(Kharal, 2010) etc.

2.2 Incomplete Information System (IIS): Sometimesinformation can be lost in the BIS. This can happen due to a virus attack, incorrect data entry, or other factors. Such a system is called the Incomplete Information System (IIS) (Ali et al., 2023). Wrong decisions can be made if the data is incomplete, which could lead to the loss of an individual or organization.The data needs to be recalculated in the IIS. This Recalculation of Incomplete Boolean Valued Information System has two categories. Preprocessed Categories and Unprocessed Categories.

In the preprocessed category some records about the existing data are stored before the data loss occurs and, that recorded information about the data is used to regenerate the lost data. Some of the preprocessed techniques are Parity bits (Mohd. Rose et al., 2011), Supported Set, and Aggregates Set (Rose et al., 2011). Recalculating BIS using available aggregates and heuristic algorithm (Khan et al., 2016). Salman et al(Ali et al., 2023), introduced a method to recalculate the complete BIS using 4 aggregates. Row aggregates, Column aggregates, Left Right Diagonal Aggregates, Right Left Diagonal Aggregates. (Ali et al., 2023)

Unprocessed categories are those which use the available data to recalculate the lost data. They include Probability (Kong et al., 2014), Weighted Average (Zou & Xiao, 2008), DFIS (Qin et al., 2012), and ADFIS (Sadiq Khan et al., 2016). Also the simplified data filling approach in Soft Set (Kong et al., 2023).These approaches can't recalculate the entire data but they work with the available data to recalculate the missing ones.

2.3 Genetic Algorithm:

Genetic Algorithm is also a bio inspired Algorithm. It is based on how living things are evolved into more successful organisms (Mathew, 2012). It is a bio inspired heuristic algorithm, which was presented in 1975 by JH Holland (J. Holland, 1975). It is based on the survival of the fittest model. It is used in optimization problems in various fields, such as economics, engineering, and computer science, to find an optimal solution to problems.

In optimization problems, the Initial Population refers to all the available sets of solutions to a problem. Each solution that can take part in the next generation is called a chromosome. Fitness values are given to these individual chromosomes which shows how much fit these chromosomes are to take part in the next generation.

The initial population represents the set of all the available solutions or chromosomes to the problem.

The fitness values show how much fit the individuals in the population are for taking part in the next generation.

Selection involves selecting the best individual chromosomes from the available solutions for the next generation. (Shukla, Pandey, & Mehrotra, 2015) Selection can be performed by different methods. E.g., roulette wheel selection, tournament selection or random selection.

In crossover, the selected chromosomes are combined to create new offspring or individuals. The purpose of crossover is to exchange information among individuals. (Kora & Yadlapalli, 2017). Some of the well-known crossover methods are: Single-Point, Two-Point, K-Point, etc.(Katoch, Chauhan, & Kumar, 2021)

Mutation involves making changes in the individual chromosome to make it more fit or to pass the fitness criteria. They maintain genetic diversity from one population to the next population. Some well-known mutation operations are displacement, simple inversion, and scramble mutation. (Katoch et al., 2021)

There are many applications of Genetic Algorithms. Genetic Algorithms are used as optimization algorithms in various medical fields, including Radiology, Cardiology, and Oncology (Ghaheri, Shoar, Naderan, & Hoseini, 2015). It has applications in network routing, Image Processing, Data Mining, and Computer games. (Lambora, Gupta, & Chopra, 2019)

2.4 BISGA: BISGA is a method of recalculation of lost data in BIS using a set of 4 aggregates through Genetic Algorithm. The aggregate set contains Row Aggregates, Column Aggregates, Left – Right Diagonal Aggregates, and Right – Left Diagonal Aggreges. This method is based on Genetic algorithm which performs Selection, Mutation, and Crossover operations until the required results are generated.

It starts with the initial population, which shows the total number of possible solutions to the problem. It creates a table of the original BIS size, then it fills the table with 1s but applies a constraint that says the sum of the 1s in each row of the table should be equal to the Row aggregate values. Once the 1s are filled the remaining table is filled with 0s.

Steps in BISGA

a. Population

The initial population of BISGA contains the set of all possible solutions. All the individuals would have the same size as the original data. The first contribution of BISGA is to narrow down this large possible solution set or population. So BISGA applies Row Aggregates constraints which reduces the overall population by 25%.

b. Fitness Values

Each individual solution has a fitness value which shows how much fit this chromosome is for the next generation. And that the fitness of the individual chromosome is calculated using the aggregate values. Fitness values in simple words are the absolute difference between the original data aggregates and the new individual aggregates. If the difference is close to 0 the more fit it is.

c. Selection

Selection involves selecting the best chromosomes or individuals for the next generation. Two or more chromosomes can be selected, and the selection would be based on the fitness values. Roulette wheel selection or tournament selection methods are used for the selection purposes. Individuals having high fitness values would have a higher chance of being selected.

d. Crossover

Once the selection is done, two or more chromosomes will be selected for the crossover stage. In BISGA crossover is performed by taking the two selected chromosomes and mixing them. The first half of the first chromosome and the second half of the second chromosome are taken row wise, and a new chromosome is generated. Similarly, the second half of the first chromosome is taken, and the first half of the second chromosome is taken row wise, and another chromosome is generated.

e. Mutation

Mutation is performed by swapping values in rows. It diversifies the individual chromosomes.

It keeps the initial row constraint and only performs row wise mutation to create new diversified chromosomes without disturbing the initial row constraint.

2.4 Encryption: Cryptography usually referred to as "the study of secrets" is nowadays related to data encryption (Al Tamimi, 2006). Encryption is the process of converting plain text into cipher text. Encryption or cipher is divided into two main categories. These are Classical Cipher and Modern Cipher.

a. Classical Ciphers: They are the oldest Ciphers, these existed from the very early ages. Here the encryption is based on simple Substitution and Transposition. It has two types.

1. Substitution Ciphers
2. Transposition Ciphers

In substitution ciphers, each content of a plain text is replaced with other content, which could be letters, symbols, or even blocks(Wenbo, 2004) Some examples of Substitution ciphers are Caesar Cipher, Monoalphabetic Cipher, Playfair Cipher, Hill Cipher, and Polyalphabetic Cipher.
In Transposition Cypher some sort of distortion is applied on the plain text in such a way that it would rearrange the bit pattern. A simple example of such ciphers isthe Rail Fence technique and Column Transposition. (Wenbo, 2004)

b. Modern Ciphers

Modern Encryption can be categorized into two main categories, Symmetric Encryption and Asymmetric Encryption (Stallings, 2006).
Symmetric Encryption uses the same key for encryption and decryption. It has five components. Plain Text, Encryption Algorithm, Cipher Text, Key, Decryption Algorithm.
Asymmetric Encryption uses different keys for encryption and decryption. It is also known as public key encryption it used a pair of keys, if one is used for encryption the other would be used for decryption. Its components are: Plain text, Encryption Algorithm, Public and Private keys pair, Cypher text, Decryption Algorithm.

3. Proposed Method

This section has 3 parts. The first part presents a working example of the proposed method. The second part discussed the generalization of this example with algorithms and pseudo code. In the third part, important functionalities and procedures are discussed.

3.1 Methodology

The base paper BISGA (Ali et al., 2023)introduces a method to recalculate the wholeBIS using four aggregates through Genetic Algorithm.That method regenerates data from some available information, making it a promising technique for data security.In that method, four aggregates are calculated and they are used to regenerate the original data. So, the aggregates can act as encrypted data which then be processed to regenerate the original data.
The proposed method is based on the concept of BISGA. It first generates BIS for each word. For that, firstly, the ASCII values of each character in the words are generated. Secondly, the ASCII value of each character is converted into binary values. Binary values are arranged in a table and the BIS would be produced. Finally, Aggregates would be generated from those BIS. Those aggregates would be the encrypted messages. Those aggregates are processed to regenerate the original ASCII codes and hence the corresponding characters.

3.2 Example

Suppose the message is: "Kill the King!"
It will first split the sentence into its component words, and each word will be processed individually.

$$\boldsymbol{MESSAGE} = [\text{"}\boldsymbol{Kill}\text{"},\text{"}\boldsymbol{the}\text{"},\text{"}\boldsymbol{King}\text{"},\text{"!"}]$$

ENCRYPTION

Step 1: Find the ASCII of the Character

BIS would be generated for each word where each row represents the binary value of that ASCII code of each character. Table 1 is the BIS for the word "Kill".

Table 1:BIS of the word "Kill"

K	0	1	0	0	1	0	1	1
I	0	1	1	0	1	0	0	1
L	0	1	1	0	1	1	0	0
L	0	1	1	0	1	1	0	0

STEP 2: Encrypt the Data
Find the four aggregates of the BIS.
For the word "Kill" aggregates are:

$$\textbf{Row Aggregates} = \{4, 4, 4, 4\}$$
$$\textbf{Column Aggregates} = \{0, 4, 3, 0, 4, 2, 1, 2\}$$
$$\textbf{Left} - \textbf{Right Diagonal Aggregates} = \{1, 2, 0, 1, 2, 2, 3, 2, 2, 1, 0\}$$
$$\textbf{Right} - \textbf{Left Diagnal Aggregates} = \{0, 1, 1, 2, 3, 2, 2, 3, 2, 0, 0\}$$

These four aggregates are the encrypted messages.
STEP 3: Generate A Key
A key must be generated to show the length of the word and the order in which the aggregates are sent. It could be:
K = 4RLRRLC
Priority can be assigned to the order of the Aggregates because sometimes the combination can be confusing.
STEP 4: Send the encrypted Data and the Key
Send the encrypted message to the receiver and the Key through a secure channel.

DECRYPTION
Step 1: Receive the Key and the Encrypted Data
K = 4RLRRLC
D = [{4, 4, 4, 4}, {1,2,0,1,2,2,3,2,2,1,0 }, {0,1,1,2,3,2,2,3,2,0,0 }, {0, 4, 3, 0, 4, 2,1, 2}]
Receive the Encrypted data and the Key from the receiver.
Step 2: Understand the length of the Word and the Order of the aggregates
Use key to find out the length of the word and the order in which the aggregates are sent.
Step 3: Apply BISGA
Apply Pseudo Code 3 (BISGA) and recalculate the BIS from the provided Aggregates. The BISGA produces BIS with the highest fitness value, in each iteration. It starts with the initial population and calculates each chromosome's fitness values. Then, the fittest individuals with the highest fitness values will be selected for mutation and crossover to generate new individuals closer to the intended solution. If a chromosome is produced with the highest fitness value this process will end. BISGA produced the first fittest BIS shown in Table 2. The produced BIS is not the exact solution so BISGA needs to be re-executed.

Table 2:The first fittest BISproduced by BISGA

I	0	1	0	0	1	0	1	1
N	0	1	1	0	1	1	0	0
K	0	1	1	0	1	1	0	0
G	0	1	1	0	1	0	0	1

Table 3: The second fittest BIS produced by BISGA

K	0	1	0	0	1	0	1	1
I	0	1	1	0	1	0	0	1
L	0	1	1	0	1	1	0	0
L	0	1	1	0	1	1	0	0

Step 4: Replace the bits of BISGA with its corresponding Characters

BISGA produces a BIS with the highest fitness value, which can be either the same solution or could be close to the same solution. So, if that is not the same solution, BISGA needs to be re-executed until the original solution is formed.

To check if BISGA has produced the same solution, first, the bits in each row of the BIS produced by BISGA are replaced by their corresponding characters as in Table 3.

Step 5: Compare the word in the dictionary

A word will be formed by combining the characters from Step 4. Now that word will be checked in the user-defined dictionary, (which will contain all the words that the user agrees to use for their communication). If it doesn't exist in the dictionary, it goes back to Step 3 and recalculates a new word.

BISGA produced the first fittest BIS in Table 2, which formed the word "iNkg" which doesn't exist in our dictionary, so it went back to Step 3, and after the 35thiteration, it formed Table 3. The word formed from that table is "Kill" in the dictionary.

Step 6: End the Process

Once the word is found in the dictionary the process comes to an end.

The same steps would be followed for the remaining words in the sentence "Kill the King!"

3.2 Pseudo Codes

The pseudo-code 1 shows how the encryption is performed. For the encryption, it takes a text, e.g., a sentence. Its output is the aggregate set that would be used as encrypted data. The flowcharts of Pseudo Code 1 (Encryption) and Pseudo Code 2 (Decryption) are presented in Figure 1 and 2 respectively.

Pseudo Code 1 (Encryption)
Input: A text data to be encrypted
Output: Aggregates (encrypted data)

1. Split the input text into separate words and store it in a list.
2. Find the ASCII value of each character in the word.
3. Put each word's ASCII code in the row of a table to create a BIS.
4. Encrypt the BIS by replacing the original data by computing its Aggregate values
5. Generate the Key.
6. Send the Encrypted Data and the Key to the receiver.
7. Repeat step 2 to 6 for each word in the list.

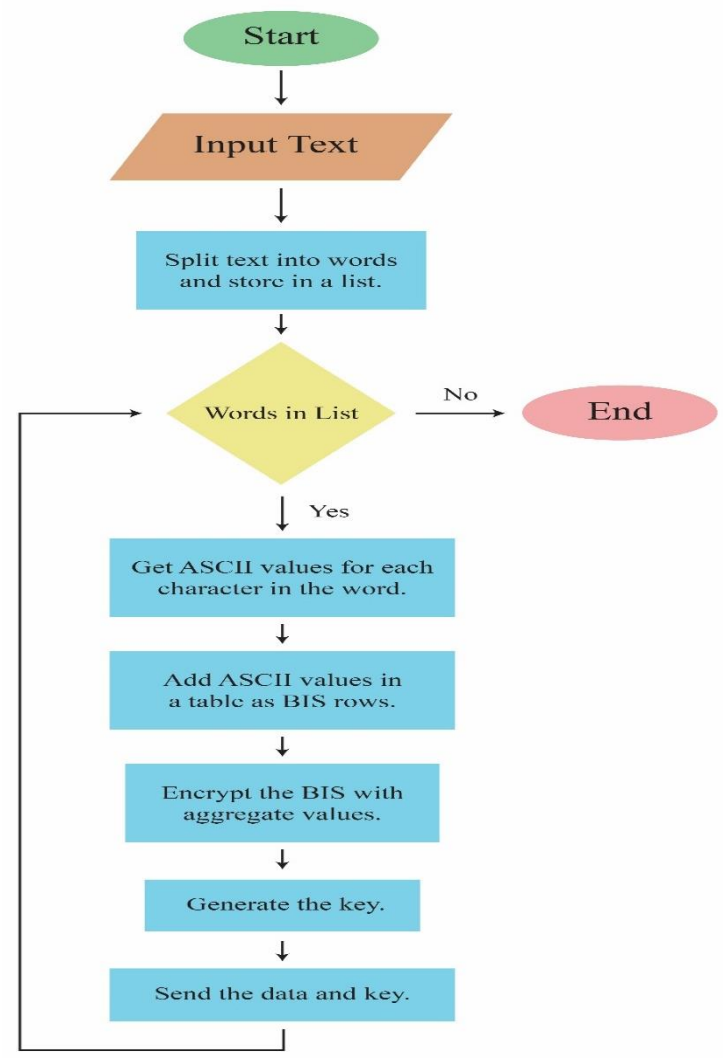


Figure 1: Flowchart of Pseudo Code 1 (Encryption)

Pseudo Code 2 (Decryption) **Input:** Aggregates (encrypted data) **Output:** Message
1. Receive the Encrypted Data and the Key 2. Apply the key to understand the word's length and the Aggregates' order, and create a BIS. 3. Apply BISGA to the BIS to regenerate the data by using the Aggregates. 4. Replace each row of the fittest BIS with its corresponding character. 5. Checks the recalculated word from the character, if that exists in the dictionary. 6. If the recalculated word exists in the dictionary, terminate the process; otherwise, go to step 3 and recalculate the data. If it reaches 500 iterations and still does not find the word, exit the process.

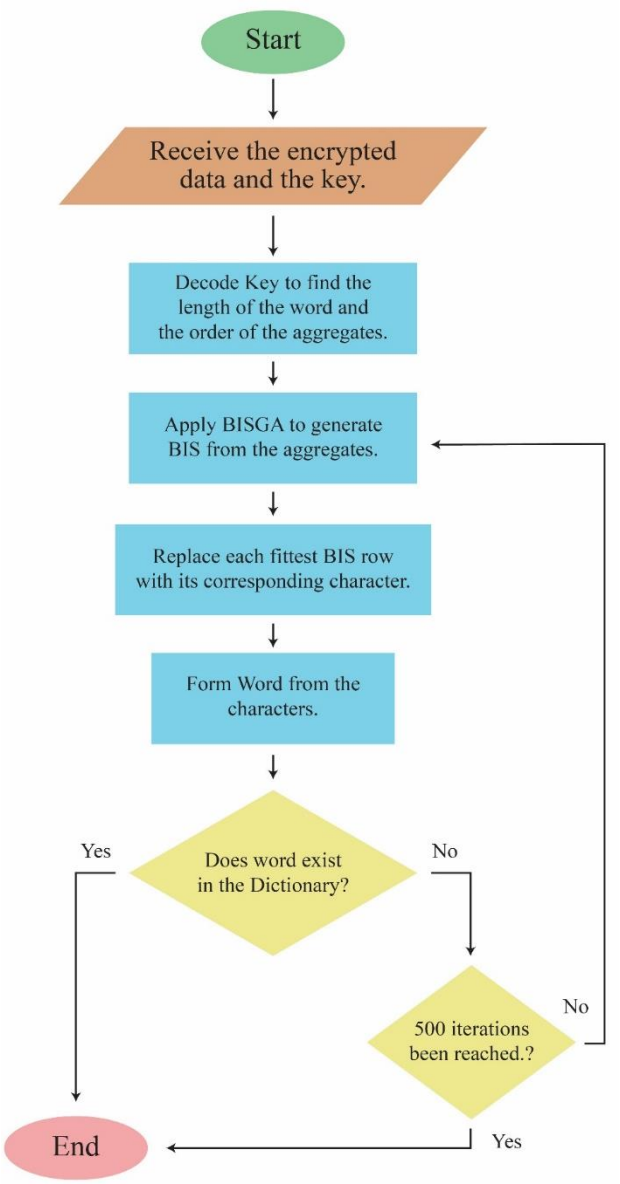


Figure 2: Flowchart of Pseudo Code 2 (Decryption)

3.3 Genetic Algorithm

In this context, a Genetic Algorithm operates by treating potential solutions as a population of chromosomes. Each chromosome is evaluated using a fitness function to determine its suitability for the next generation. The selection process, often through Roulette Wheel selection, favors chromosomes with higher fitness. Selected "parent" chromosomes undergo crossover, where segments from two parents are combined to form new offspring. To introduce diversity, mutation is applied by altering elements within a chromosome. This iterative process continues to refine solutions over successive generations.

3.4 Fitness Function

Fitness Function is used to find the fitness value of individual chromosomes. It shows the absolute difference between the aggregates set of the original data and the input aggregates. Here first of all the fitness value of the Aggregates "x" is calculated.

$$\text{Fitness (aggX)} = \sum \text{abs(act} - \text{ind)}$$

Once those are calculated the total Fitness is calculated by taking the sum of all the individual Fitness values. So, in this case, there are 3 aggregates set and the sum of the fitness values of those 3 aggregates will give the overall fitness value of that chromosome.

$$\begin{aligned}\text{Fitness(ind)} = {} & \text{fitness(aggLR)} \\ & + \text{fitness(aggRL)} \\ & + \text{fitness(aggC)}\end{aligned}$$

3.5 Dictionary

This encryption technique maintains a list of words, numbers, and letters, which are used to constitute the message, and named it a dictionary. The dictionary needs to be both on the sender side and on the receiver side. Words for the message must be chosen from the existing dictionary, or the word first needs to be added to both dictionaries before carrying out the encryption and decryption. For our experiments, the dictionary used is shown in Figure 3.

Dictonary.txt

```
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9 ! @ # $ % ^ & * ( ) _ + - = { } [ ] : ; ' " , . < > / ? | \ ~ `
Kill the King zz To Go Be Cod Jet Agent Strike
Hi The sun is shining today, Hi To Go
No Be Spy Air Gun Cod Red Code Base Rank Unit Fire Agent
Intel Alert Drone Recon Target Strike Attack Weapon Threat
call code plan snap dive work fast find move ship base list safe
task hand zone stop save grow home shop pack link fire done
goal lock grid send flag red snap load snap coin key done
help item park team drop note map info calm aim pick data
coin tape fast calm phone hide keep task list jump done snap
find meal note stop dive bank lock plan move test pass work
park stop zone pack code hand snap grow done snap post flag
keep read snap list fast move plan fire item task calm hide
key lock snap food snap note load test snap post pack done
goal flag link park snap plan zone stop drop snap grow fire
coin pick task hide snap done list calm move snap pass zone
fire drop task snap bank hand calm snap move work fast plan
save snap task plan grid snap park done code drop pack flag
snap task find load post code map safe team fire drop info
item stop calm hand task note snap plan load grid move fast
```

Figure 3:A sample Dictionary used for experiments

4. Results and Discussions

In this section, the results of the proposed method after implementation in Python are given and include a discussion about the results and the proposed methods.

4.1 Results

The proposed method was tested with words of different lengths and sentences and it worked as a functional encryption method. However, its performance drops when the size of the words becomes large. The proposed method was tested with words of length two, three, four, five, and six. Results show that it's a functional method with some limitations with words of large lengths.

4.1.1 Results for words of length 2

A group of length 2 words were taken and this method is applied 10 times for each and the results are shown in Figure 4.

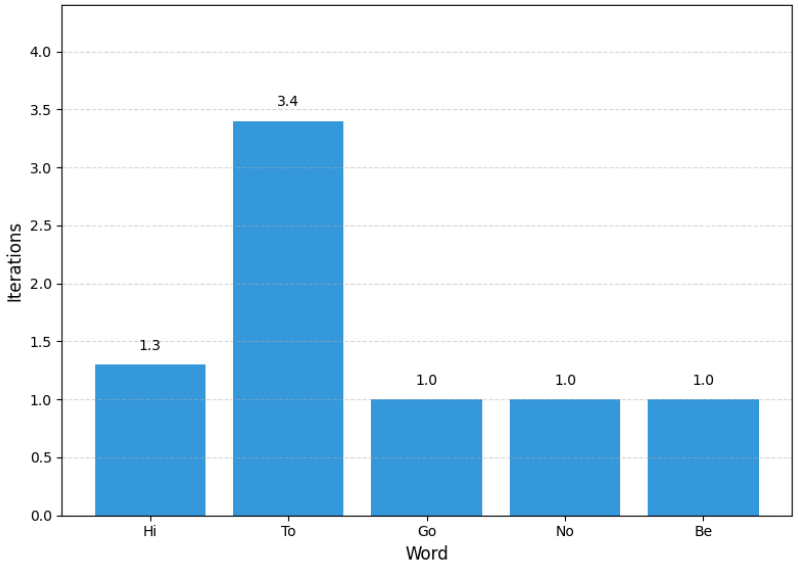


Figure 4:Iterations for words of length 2

4.1.3 Results for words of length 3

Figure 5 shows iteration for 5 words of length 3.

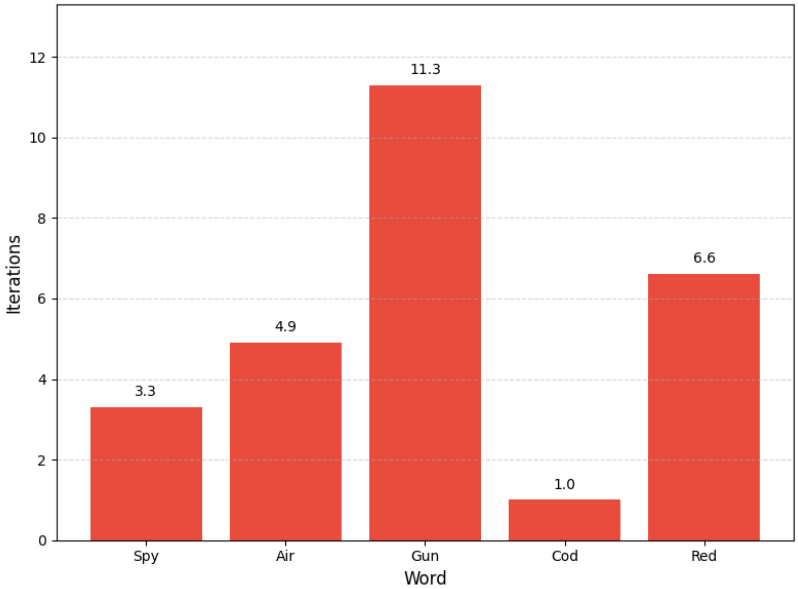


Figure 5:Iterations for words of length 3

4.1.4 Results for words of length 4

Results for 5 words of length 4 are shown in Figure 6.

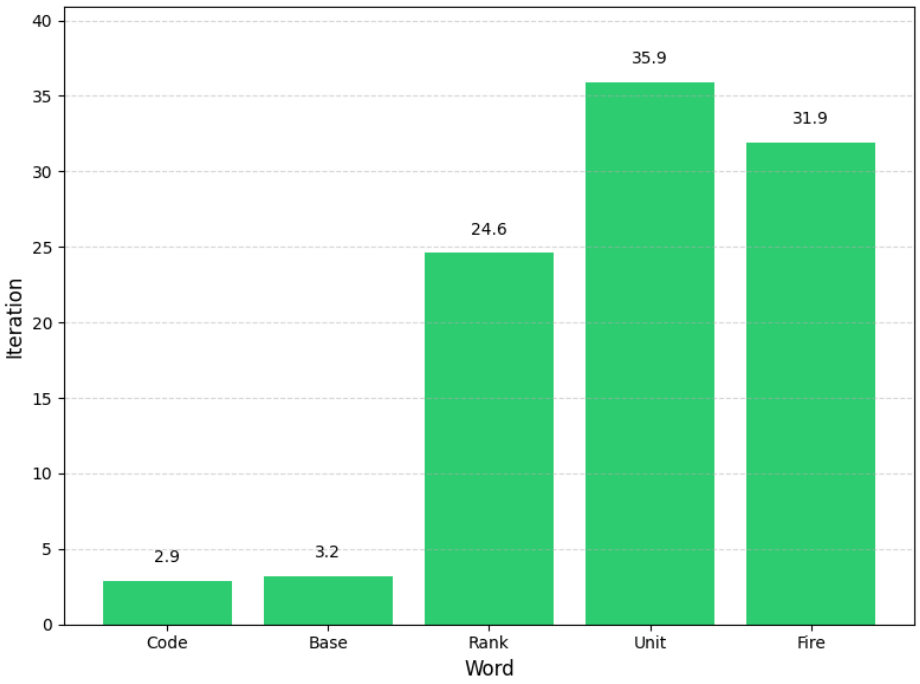


Figure 6:Iterations for words of length 4

4.1.5 Results for words of length 5

The results for the 5 words of length 5are shown in Figure7.

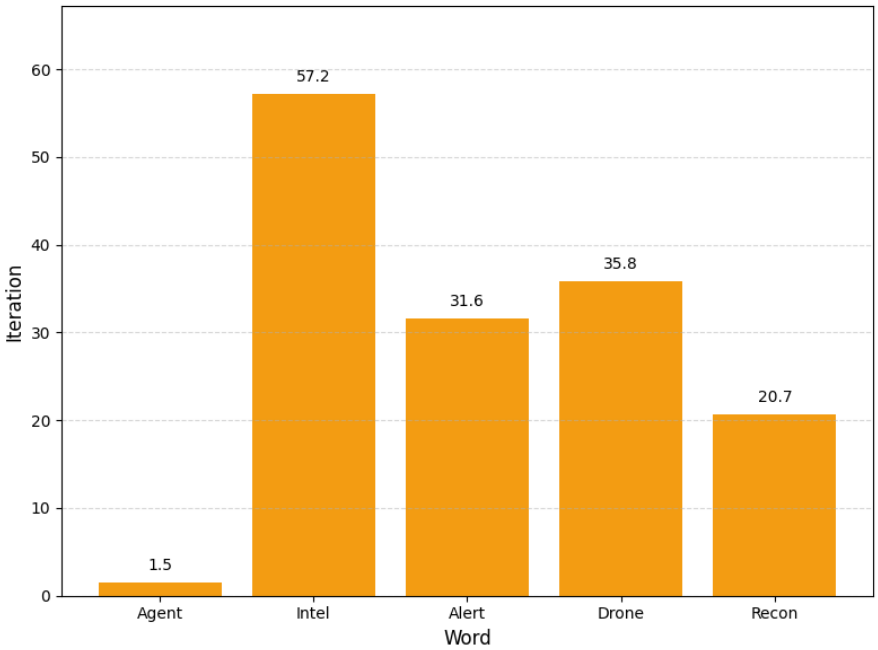


Figure 7: Iterations for words of length 5

4.1.5 Results for words of length 6

Figure 8 shows the results for 5 words of length 6. The results show the iteration has increased significantly as the word length increased from 2 to 6.

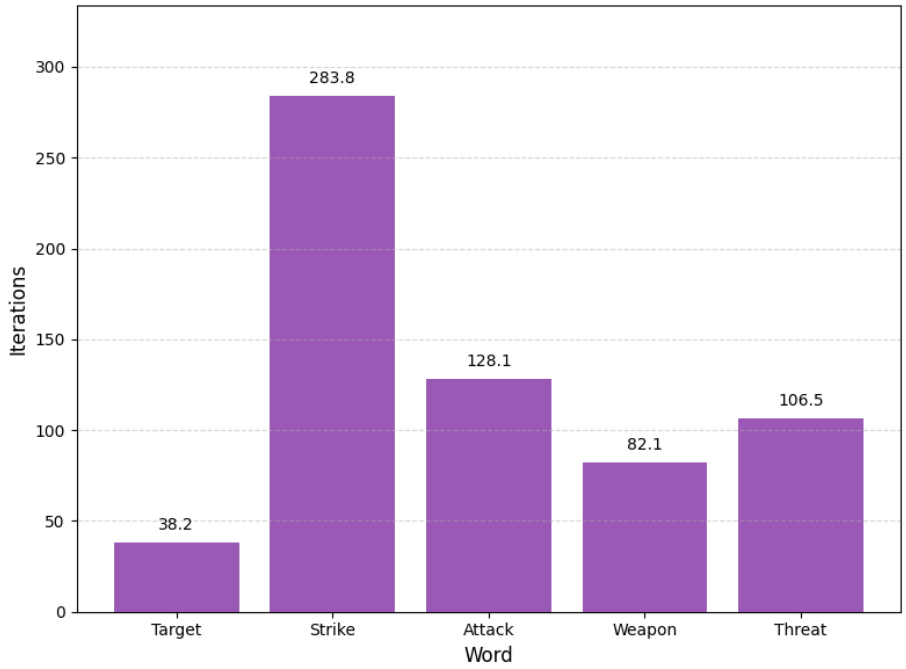


Figure 8:Iterations for words of length 6

4.1.6 Results for Letters, Numbers, Special Characters and Sentences

Letters, Numbers, and Characters always take just one iteration. So, making a dedicated table and figure for them is unnecessary. Sentences are the combination of words letters and special characters. So, iteration for sentences is the iteration for each word.

4.1.6 Average Results

The average results are shown in Figure 9. The results showed that this method would take more iteration and time when the length of the words becomes long.

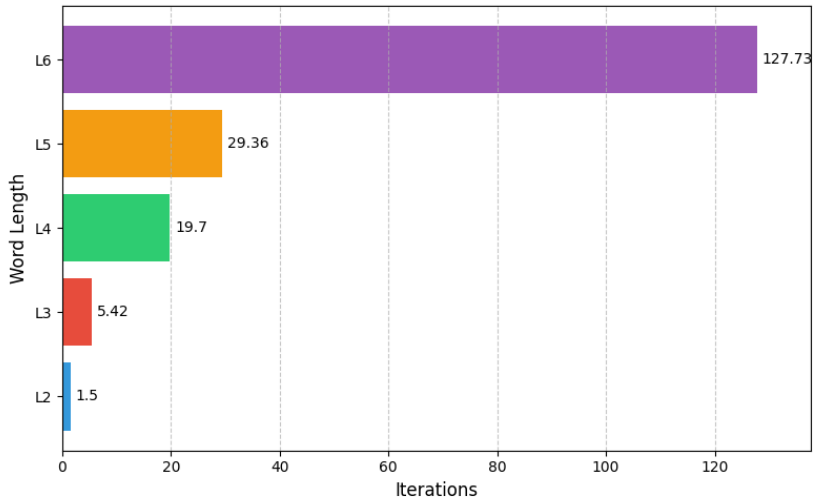


Figure 9:Overall results

4.2 Discussion

The proposed encryption method performs effectively, particularly with shorter words, as demonstrated through testing with varying word lengths. However, it encounters limitations with longer words, requiring more iterations and increased processing time. In our results, we applied a default limit of 500 generations.

Words that exceed this limit cannot be processed efficiently, making the approach less suitable for recalculating longer words.

The proposed method introduces a novel approach to data encryption, differing from traditional techniques based on Transposition and Substitution, such as Monoalphabetic Substitution or AES. It employs data aggregates for encryption, initially used to recalculate lost BIS in BISGA. The results highlight a novel perspective on encryption, offering a distinctive method through BISGA-based aggregates.

The proposed method has some limitations as well. Firstly, its efficiency drops with the increase in word length. This method works well with words of length up to 6, but beyond that, it takes more time and iteration. Secondly, the key is not very strong, it doesn't have a large key space so it can be found easily. So, work can be done to improve the key. Thirdly, it can't process sentences all at once. If a sentence is given, it first divides into component words and processes them individually.

A key concern of this approach is potential issues when verifying words in the dictionary. BISGA can generate multiple BISs with identical aggregates and fitness values, which may correspond to different words. While this isn't problematic with a small dictionary, larger dictionaries increase the likelihood of a BIS matching multiple words, creating ambiguity.

In the majority of languages, short-length words are used for communication. Short-length words make up around 76 to 80 percent of the English Language (Sigurd, Eeg-Olofsson, & Van Weijer, 2004). The proposed method works efficiently with words of shorter length, which makes it an effective encryption technique for day-to-day communication.

There are multiple potential future directions for the proposed method. Firstly, as mentioned its efficiency reduces as the length of the word increases. Secondly, it may face conflict while comparing the words in the dictionary. Thirdly, the results demonstrate that certain words take very less iterations and some words take very large iterations, even having the same word length. e.g., "Code" and "Unit". Although both have the same length the average iteration for "Code" is 2.9 and for "Unit" is 35.9. This could be due to some pattern in the internal bit combination of the BIS of each word, which can be explored in the future. Lastly, further investigation can be carried out to find out the real-world application of this cryptographic technique.

5.Conclusion

The present research introduced an innovative data encryption method that employs aggregates via Boolean Valued Information Systems in conjunction with a Genetic Algorithm (BISGA). The results of this research illustrate the capability of this approach in ensuring data integrity through a range of applications. The findings acquired during the experimental stage demonstrate that the proposed technique proficiently encrypts and decrypts words of varying lengths, highlighting its versatility and reliability.However, the research indicates that the efficacy of the method diminishes with the lengthening of the words. This constraint highlights the imperative for forthcoming investigations to concentrate on refining the methodology to enhance scalability and performance. Moreover, investigating the practical applications of this method and integrating it with alternative cryptographic techniques could substantially improve its functionalities and broaden its range of applications.This research yields several noteworthy contributions, which can be summarised as follows:

1. Presentation of an innovative data encryption methodology utilising Aggregates via BISGA.
2. An exhibition of the technique's efficacy in the encryption and decryption of words varying in length.
3. Recognition of the technique's constraints and potential routes for enhancement.

In conclusion, this study has provided a framework for further investigations into the use of BISGA in the field of data encryption. Subsequent research may build on the findings of this study to develop more effective, flexible, and secure data encryption technologies.

Future research directions

In the forthcoming research, we intend to broaden this investigation to encompass additional fuzzy environments and to address its subsequent challenges.

1. Examine the potential for scaling the proposed technique to accommodate larger datasets and extended word lengths.
2. Investigate the integration of the suggested technique with alternative cryptographic methodologies to improve its security and efficiency.
3. Formulate pragmatic implementations of the suggested methodology across diverse domains, including data storage, transmission, and popularization.
4. In the upcoming research, we have also plane to expand this investigation to include other fuzzy frameworks andthe model can be applied to additional specialized decision-making challenges.

Through the exploration of these research fields, subsequent investigations can enhance and polish the suggested data encryption methodology, thus promoting the progression of data security and integrity.

Data Availability: All the data are available in the manuscript.

Author contributions: The authors Muhammad Sadiq Khan, Attaullah, Feroz Ali, Adila Tasnim, and Nizamud Dincontributed equally to the manuscript.

Competing interests: The authors declare no competing interests.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Funding declaration: There is no funding source available for this research.

REFERENCES

Al Tamimi, A.-K. (2006). Performance analysis of data encryption algorithms. available at weblink: http://www. cse. wustl. edu/~ jain/cse567-06/ftp/encryption_perf/index. html.

Ali, S., Khan, M. S., Shah, H., Garg, H., & Alsheddy, A. (2023). BISGA: Recalculating the Entire Boolean-Valued Information System from Aggregates Using a Genetic Algorithm. Complexity, 2023(1), 1539563.

Bustince, H., & Burillo, P. (1996). Vague sets are intuitionistic fuzzy sets. Fuzzy sets and systems, 79(3), 403-405.

Deli, I., & Cagman, N. (2013). Application of soft sets in decision making based on game theory. arXiv preprint arXiv:1302.4568.

Ghaheri, A., Shoar, S., Naderan, M., & Hoseini, S. S. (2015). The applications of genetic algorithms in medicine. Oman medical journal, 30(6), 406.

Goyal, K., & Kinger, S. (2013). Modified caesar cipher for better security enhancement. International Journal of Computer Applications, 73(3), 0975-8887.

J. Holland, M. M., S. Kumar, and R. Porwal,. (1975). Adaptation in natural and artificial systems," in Applying Genetic Algorithms to Increase the Efciency of a Data Flow-Based Test Data Generation Approach. University of Michigan Press, Ann Arbor, MI, USA.

Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: past, present, and future. Multimedia tools and applications, 80, 8091-8126.

Khan, M. S., Herawan, T., Wahab, A. W. A., Mujtaba, G., & Al-Garadi, M. A. (2016). Concept of entire Boolean values recalculation from aggregates in the preprocessed category of incomplete soft sets. IEEE Access, 5, 11444-11454.

Kharal, A. (2010). Distance and similarity measures for soft sets. New mathematics and natural computation, 6(03), 321-334.

Kong, Z., Lu, Q., Wang, L., & Guo, G. (2023). A simplified approach for data filling in incomplete soft sets. Expert Systems with Applications, 213, 119248.

Kong, Z., Zhang, G., Wang, L., Wu, Z., Qi, S., & Wang, H. (2014). An efficient decision making approach in incomplete soft set. Applied Mathematical Modelling, 38(7-8), 2141-2150.

Kora, P., & Yadlapalli, P. (2017). Crossover operators in genetic algorithms: A review. International Journal of Computer Applications, 162(10).

Lambora, A., Gupta, K., & Chopra, K. (2019). Genetic algorithm-A literature review. Paper presented at the 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon).

Majumdar, P., & Samanta, S. K. (2008). Similarity measure of soft sets. New mathematics and natural computation, 4(01), 1-12.

Mathew, T. V. (2012). Genetic algorithm. Report submitted at IIT Bombay, 53.

Mohd. Rose, A. N., Hassan, H., Awang, M. I., Herawan, T., & Deris, M. M. (2011). Solving incomplete datasets in soft set using parity bits of supported sets. Paper presented at the Ubiquitous Computing and Multimedia Applications: Second International Conference, UCMA 2011, Daejeon, Korea, April 13-15, 2011. Proceedings, Part I 2.

Molodtsov, D. (1999). Soft set theory—first results. Computers & mathematics with applications, 37(4-5), 19-31.

Pawlak, Z. (1996). Why rough sets? Paper presented at the Proceedings of IEEE 5th International Fuzzy Systems.

Qin, H., Ma, X., Herawan, T., & Zain, J. M. (2012). DFIS: A novel data filling approach for an incomplete soft set. International Journal of Applied Mathematics and Computer Science, 22(4), 817-828.

Rényi, A. (2007). Probability theory: Courier Corporation.

Rose, A. N. M., Hassan, H., Awang, M. I., Mahiddin, N. A., Amin, H. M., & Deris, M. M. (2011). Solving incomplete datasets in soft set using supported sets and aggregate values. Procedia Computer Science, 5, 354-361.

Sadiq Khan, M., Al-Garadi, M. A., Wahab, A. W. A., & Herawan, T. (2016). An alternative data filling approach for prediction of missing data in soft sets (ADFIS). SpringerPlus, 5, 1-20.

Shukla, A., Pandey, H. M., & Mehrotra, D. (2015). Comparative review of selection techniques in genetic algorithm. Paper presented at the 2015 international conference on futuristic trends on computational analysis and knowledge management (ABLAZE).

Sigurd, B., Eeg-Olofsson, M., & Van Weijer, J. (2004). Word length, sentence length and frequency–Zipf revisited. Studia linguistica, 58(1), 37-52.

Stallings, W. (2006). Cryptography and network security, 4/E: Pearson Education India.

Wenbo, M. (2004). Modern cryptography; theory and practice. (No Title).

Zadeh, L. A. (1965). Fuzzy sets. Information and control, 8(3), 338-353.

Zou, Y., & Xiao, Z. (2008). Data analysis approaches of soft sets under incomplete information. Knowledge-Based Systems, 21(8), 941-945.

IEER
Institute for Excellence in Education & Research