Abstract: - Securing private information during its transmission over the internet is of paramount importance, as it safeguards against unauthorized access and guarantees the integrity of data. Steganography serves as a technique for enhancing security by concealing sensitive information within digital content such as images, videos, audio, and textual carriers, thus thwarting unauthorized interception. Current methodologies involve manipulating pixels to create stego-images, effectively embedding confidential messages within the image structure. However, this approach has led to steganalysis experts uncovering these hidden messages. To address this issue, an innovative strategy for covert data concealment is proposed. This paper proposes a robust, secure, and embedding capacity coverless image steganography method using OMR and Mapping Rules. The secret information is represented by a binary string, and the bubble sheet is used to represent binary fragments. This approach ensures the confidentiality of sensitive data communicated over the internet. Notably, this approach surpasses previous alternatives in terms of embedding capacity, as substantiated by result analysis. Furthermore, it is adaptable for refinement as needed. Significantly, it exhibits remarkable resilience, safeguarding against a spectrum of attacks including scaling, color space conversion, JPEG compression, thresholding, "salt and pepper" noise, file format conversion, and steganalysis tools, among others.

Keywords: Coverless information hiding, Optical Mark Recognition (OMR), Image steganography, Machine Learning.

I. INTRODUCTION

All in a variety of industries, including network broadcasting, social production, and television, portable multimedia devices and 4G network apps are quickly gaining popularity. Information concealing is a secure communication method used in network broadcasting, social production, and television that uses encryption such as DES and RSA. Confidential information is encrypted by the sender and decrypted by the recipient using a key [1]. However, because of data processing and computer speed, this approach is vulnerable to hacking. Information concealing technology embeds hidden information that is difficult for attackers to decode by using human visual sensitivity and carrier signal redundancy [20]. The modifications made to the stego-image by image steganography algorithms, whether in the frequency or spatial domains, may be recognized by any detection system. Passive and aggressive steganalysis techniques can be used to identify payloads. Active techniques retrieve, alter, or tamper with the concealed message, whereas passive methods just reveal secret data. Artifacts including network transmission faults or picture alterations like filtering, scaling, file format conversion, noise, and data compression are among the unresolved issues in image steganography [2, 3, 4, 5].

In order to fundamentally resist steganalysis, Sun and other scholars propose a coverless information hiding scheme to resist steganalysis by establishing a mapping relationship between secret information and hidden carriers based on carrier characteristics. This unique anti-steganalysis ability prevents attackers from obtaining the information, even if they obtain the original carrier containing the secret information [1]. Although theoretical research and maturity are still in its early stages, coverless is a promising field with potential value. Similar to image retrieval, coverless image steganography techniques make use of images from a local database that has been pre-defined to represent the payload. These methods are not reliable since they are capable of handling some steganalysis tools but not all of them. They are unable to solve problems brought on either corrupted images or network transmission faults [1]. Based on an image's pixel values, features, or histogram, it is chosen to represent the hidden message. Part or the entire secret message is lost if image information is altered through steganography, modification, or network transmission fault. Due to fixed-length binary streams and the unequal distribution of chosen features in images, coverless image steganography algorithms have a limited embedding capacity. A novel coverless image steganography method is proposed to improve security, robustness against active attackers, and overcome challenges like network transmission errors and image manipulations.
Section 2 of the paper's remaining sections introduces various studies on coverless image steganography techniques that are relevant to the rest of the paper's structure. The suggested technique is thoroughly explained in Section 3. Section 4 summarizes the outcomes of the suggested approach before contrasting them with those of other ways. The paper's conclusion is covered in Section 5 lastly.

II. RELATED STUDIES

Zhou et al\[6\] paper uses a pixel feature to obtain a hash sequence for each image in the database. The algorithm uses the average intensity of each block and inverts the images according to their hash sequence. The index structure retrieves the image with the same hash sequence as the secret information sequence as a stego image. The images are transmitted to a receiver, who obtains a hash sequence through the same algorithm at the receiving end, storing the secret information. The algorithm has a capacity of 8 bits.

Chen et al\[26\] improved the hash algorithm and increased the secret information length to 18 bits, enhancing the hidden capacity. Yuan et al. proposed coverless information hiding using SIFT and BOF, converting secret information into binary and segmenting it. The image hash sequence was obtained through SIFT feature extraction and clustering, and the matched sequence was transmitted as a stego image to the receiver.

Zhou\[8\] et al.’s CIS-PDVR algorithm uses partial-duplicate visual retrieval to achieve coverless information hiding, using image as the transmitted secret information. The algorithm is inspired by similarities in image blocks between two images, allowing for the formation of similar visual images. However, the secret information cannot be extracted completely correctly.

For finding hidden information in cover photos, Liu suggested a CIS technique utilizing DenseNet feature mapping \[9\]. High-dimensional CNN features are extracted by deep learning and converted into hash sequences using a binary tree hash index for quicker searching. The secret information is retrieved by computing the cover image's DenseNet hash sequence from the matched pictures that are provided. The cover pictures are kept constant throughout the steganography process.

Saad [10] proposes a coverless image steganography method using jigsaw puzzle image generation using secret message bits. The image is divided into equal rows and columns, creating blocks with tabs/blanks. The mapping function and secret message bits create a fully shaped jigsaw puzzle stego-image, which is then sent to the receiver.

Liu [24] proposes a Coverless image steganography based on DenseNet feature mapping. ResNet extracts semantic features and segments object areas using Mask RCNN for information hiding. These areas have ethical structural integrity and are not in the visual center, reducing information loss. They are then binaries to generate hash sequences for information mapping. The model resists steganalysis and has excellent robustness, making it suitable for pre-training through supervised learning.

A. Optical Mark Recognition

First, OMR is an electronic technique that collects data that has been manually marked by identifying particular markings on a bubble sheet paper (Fig. 1). Typically, a specific scanner is used to evaluate the light reflection through the sheet; marked bubbles and circles reflect less light than blank ones, resulting in lower reflectivity. Currently, several other forms of exams are given using the OMR procedure [25]. With this technique, data is gathered from fill-in-the-bubble worksheets, such as tests, surveys, multiple-choice (MCQ) worksheets (Fig. 1). In OMR, the respondent or student selects the response to a question by filling in the bubble or circle to the appropriate response. OMR data collection is more accurate than handwritten response recognition. OMR is not an entirely new technology, but it has advanced; traditional OMR systems call for preprinted forms and unique scanners.
B. Mapping Rules using Machine Learning

A subfield of artificial intelligence (AI) and computer science called machine learning focuses on using data and algorithms to simulate how humans learn, gradually increasing the accuracy of the system.

Machine learning is crucial in data science, using statistical methods to classify and predict data, driving decision-making and impacting growth metrics. As big data expands, demand for data scientists increases, identifying relevant business questions and data. ML can be categorized into supervised, unsupervised, and reinforcement learning approaches, and is employed in the detection of faces and the understanding of spoken orders.

There are several machine learning algorithms that may be used, such as Bayesian, clustering, decision trees, dimensionality reduction, deep learning, linear regression and regularization. Choosing the optimal approach involves both art and science. Machine learning (ML) systems, which create a model that can be extensively applied to all incoming data, can be likened to this approach. In summary, these systems are straightforward: If A is input data, then execute B. It is critical to use caution while establishing a strategy in order to prevent overcomplicating a system, which may include more than 100 pre-defined rules.

The fundamental contribution of this study is the proposal of a novel and extremely robust coverless image steganography approach based on an OMR system and Mapping Rules that uses a bubble sheet as a cover in order to improve the security, robustness, and capacity of coverless image steganography.

III. PROPOSED METHOD

This method focuses on OMR and Mapping Rules, providing greater capacity than previously proposed methods.

A. Proposed Method Components (Heading 2)

The following are the elements of the suggested method:

- Bubble sheet: This corresponds to the cover image in traditional coverless steganography methods. The carrier is a specially made answer/bubble sheet.
- Secret message (payload): The bubble sheet has sensitive information mapped on it.
- Answered/mapped bubble sheet (stego-image): the secret message response or mapping bubble sheet that was created.
- Mapping algorithm (embedding algorithm): The bubble sheet replies were embedded using the secret message bits-based technique.
- Detection algorithm: The secret message was created and retrieved using the extraction method, which was used to collect and recognize the replies from the finished bubble sheet.
B. Mapping Phase

Generate a single bubble-sheet template using free OMR software before mapping phase. Then, the generated bubble sheet template is saved as an image in any format, which is the image used instead of the local database currently used in coverless image steganography methods. No image library is needed, no images need to be generates again on the sender’s or receiver’s end.

The Mapping Rules mapping method is given the secret message and the generated bubble sheet as inputs, as shown in Fig. 2. The hidden message is then transformed into a binary stream of 0s and 1s. This bit stream is then divided into blocks of 2 bits each, with block’s length being equal to the number of questions in the newly produced sheet.

Second, The Mapping Rules algorithm originated and is currently known to carry out the following functions.
1. Counting the number of questions on the bubble sheet and identifying them;
2. Finding every circle in the bubble sheet that was created;
3. Calculating the bubble sheet’s maximum capacity given that each question will include two bits of hidden information;
4. Marking and reacting to the bubbles/circles in the bubble sheet to map the hidden data.

In the mapping step, a block from the secret message (2 bits) and a question from the bubble sheet are utilized. The trained ML algorithm then marks the bubbles in the current question depending on the block’s 2 bits.

The first two bits is the answer to the first question, the second two bits is the answer to the second question and so on, with the 2 bits block corresponding to the question in created bubble sheet. So, the bubble sheet as in Fig. 3, there is four possible answers (A, B, C, and D) for each question. So there are four case of circle marked in the bubble sheet by the Mapping Rules algorithm which is given below.

Case 1: if the 2 bits value is 00, A option is correct answer and the A option circle is detected and marked in the bubble sheet.
Case 2: if the 2 bits value is 01, B option is correct answer and the B option circle is detected and marked in the bubble sheet.
Case 3: if the 2 bits value is 10, C option is correct answer and the C option circle is detected and marked in the bubble sheet.
Case 4: if the 2 bits value is 11, D option is correct answer and the D option circle is detected and marked in the bubble sheet.

As an example, if the secret message is ‘HI’ and convert into binary string is “01001000 01001001”, as shown in Fig. 3, first 16 bits secret message divides into 8 parts of 2bits length. Such as 01, 00, 10, 00, 01, 00, 10, 01. So based on above cases, the algorithm will marked the B option circle of the first question for first 2bits (01), same as for second 2 bits (00), the algorithm will marked the A option circle of the second question, for thirds 2 bits (10), the algorithm will marked the C option circle of the third question, and so on this process is repeated up to last block.
The algorithm will complete the mapping process by solving every answer on the sheet in accordance with the secret message, and the answered/mapped bubble sheet will then be prepared for sharing with the recipient.

C. Detection Phase

According to the OMR detection, the detection phase functions similarly to the mapping phase but in the opposite order. First, the circles for each question on the answered/mapped bubble sheet are recognized by scanning it question by question. After that, replies are gathered, secret bits are retrieved, and they are combined to form the bit stream. The bit stream is then split into 8-bit blocks, which are subsequently converted into ASCII characters. Finally, the gathered characters are merged to form the concealed message.

IV. Evaluation and Comparisons

This paper improves coverless image steganography properties by enhancing capacity, security, imperceptibility, and robustness. Experimental evaluation assesses effectiveness and efficiency, comparing it with existing methods.

A. Capacity

The hiding capacity of the suggested method is the highest among the available methods, as demonstrated in Tables 1 and 2. Notably, the capacity of the proposed technique is not restricted to 200 bits/cover; instead, depending on the actual requirements, more bubbles may be added until the sheet capacity is met. There are 100 questions on the experiment sheet.

The proposed method has the highest hiding capacity of 200 bits/cover, indicating significant embedding capacity which is shown in Table 1. This technique increases the hiding capacity of coverless image steganography techniques, with the bubble sheet containing more bubbles representing more binary bits. Table 2 shows that the least number of images are required to store a hidden message, indicating a reduced number of covers. Zhou et al. needed 1024 images, Zheng needed 457, and this approach only needed 40. Because each bubble sheet may map up to 20 bytes, the number of covers image is reduced.

<table>
<thead>
<tr>
<th>Method</th>
<th>Capacity (bits/cover)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z. Zhou et al. [22]</td>
<td>8</td>
</tr>
<tr>
<td>C. Yuan et al. [26]</td>
<td>8</td>
</tr>
<tr>
<td>S. Zheng et al. [23]</td>
<td>18</td>
</tr>
<tr>
<td>L. Zou et al. [12]</td>
<td>80</td>
</tr>
</tbody>
</table>

Table 1: Proposed Method Embedding Capacity

<table>
<thead>
<tr>
<th>Method</th>
<th>Secret Message Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 byte</td>
</tr>
<tr>
<td>Z. Zhou et al. [22]</td>
<td>1</td>
</tr>
<tr>
<td>S. Zheng et al. [23]</td>
<td>2</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2: Images Required When the Same Data are Hiding

V. Robustness

The steganography method's resistance to attacks is known as robustness. Algorithm failure in the secret data/payload transmission process is brought on by a number of attacks, including noise attacks, scaling attacks, and JPEG compression. Through trials and comparisons, the coverless proposed method's robustness was examined, assessed, and confirmed [17]. First, a bit error rate (BER) must be defined.

The following formula is used to compute BER, which is a measure of the durability of the communication process's steganography technique:

$$\text{BER} = e/n, e = \sum p_i \oplus q_i; \quad \text{where } i=1:n \quad (1)$$

The method's robustness to an attack is determined by the error rate (BER) in the extracted secret bits. If BER = 0, no errors were found, and the method extracted the secret bits with 100% accuracy. If BER > 0, an error rate exists, indicating the method's inability to recover from the attack.
A. JPEG Compression Attack

The most widely used method of compression for still images is JPEG. JPEG is a lossy compression format that permits data loss and is used to compress images before/during transmission [22]. If the stego-image is compressed, it is changed or destroyed during transmission. The proposed method's BER robustness was evaluated in comparison to the JPEG attack. The maximum compression ratio (JPEG quality factor) is 1, and the lowest compression ratio (JPEG quality factor) is 100.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>0.022</td>
<td>0.048</td>
<td>0</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>70</td>
<td>0.038</td>
<td>0.080</td>
<td>0.08</td>
<td>0.09</td>
<td>0.002</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>0.151</td>
<td>0.146</td>
<td>-</td>
<td>0.146</td>
<td>0.007</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3 compares CBD, CBZS, CSD, CIHRIH, Wu et al. methods, and the proposed method, revealing its 100% robustness to JPEG compression attacks and 100% secret message extraction accuracy.

B. Noise Attack

The proposed method analyzes random distributed "salt and pepper" noise in images with fixed depth, detecting simultaneous presence of salt and pepper noises in 0.01-0.04 increments.

The BER of several approaches, including CIHWE, CIHRIH, Wu et al., and the suggested method following "salt and pepper" noise attack, is compared in Table 4. At the same noise level, the suggested approach exhibited a zero BER, showing complete resistance to salt and pepper noise assaults.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.02</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.2</td>
<td>0.06</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.3</td>
<td>0.11</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.4</td>
<td>0.16</td>
<td>0.09</td>
<td>0.0005</td>
<td>0</td>
</tr>
</tbody>
</table>

VI. SECURITY

The suggested approach ensures great security by identifying a student's completed bubble-sheet for a final exam. This approach offers strong security and makes it difficult for attackers to discover the payload.

Data hiding techniques should be resistant to steganalysis assaults. The majority of extant image steganography algorithms, however, are traceable by steganalysis tools through pixel bit shifts. The suggested approach, which does not make use of a cover picture, discovers the circle and marks it in order to map hidden bits. OMR sheets are not suspicious because they are utilized in tests and surveys without being used as cover image earlier.

VII. CONCLUSION

This work suggests an OMR and Mapping Rules based coverless image steganography technique that is reliable, secure, and has a high embedding capacity. The binary string that encodes the secret information (payload) is represented by the bubble sheet as binary fragments using a mapping function. The version that has been responded and delivered to the recipient is the created bubble sheet. The technique makes use of optical mark recognition (OMR) and rule-based machine learning (MAPPING RULES) algorithms. In bubble-sheet examinations, the Mapping Rules algorithm simulates student behavior by locating circles and marking them according to the right responses. The algorithm solves the question and maps the secret bit, with the secret bit serving as the appropriate response.

The mapped bubble sheets are utilized to gather the students' correct responses using the OMR algorithm, which was created during the detection step. In comparison to existing techniques, the suggested coverless steganography approach has a number of advantages, including the lack of a database, the avoidance of revealing confidential information, and the reduction of search time. Compared to previously suggested approaches, this one has the maximum embedding capacity and may be scaled up as necessary. Additionally, it has a high level of resilience, fending off attacks from scaling, colour space conversion, JPEG compression, thresholding, "salt and pepper" noise, file format conversion, and steganalysis tools, among others. Since bubble sheets are not
suspicious and have never been used as cover files, the procedure is also very secure. Overall, compared to conventional techniques, the suggested coverless steganography approach is more effective and secure.

REFERENCES


