MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 1928-1941

Attack-Resilient Medical Image Watermarking Using Convolutional Neural Networks and Fuzzy Inference Decoding Mechanisms

[1]Barkha Sahu*, [2]Dr. Neeraj Kumar Rathore, [3]Dr. Abhishek Bansal
[1]PhD research scholar, [2]Associate Professor, [3]Associate Professor
[1, 2] Affiliation Address: Indira Gandhi National Tribal University Amarkantak, Anuppur, Madhya Pradesh, India
[3]Affiliation Address: Dr. Harisingh Gour Vishwavidyalay, Sagar, Madhya Pradesh, India
[1]Email: barkhaigntu@gmail.com, [2]Email: neeraj.rathore@igntu.ac.in,
[3]Email: abhishek.bansal@igntu.ac.in
***Corresponding Mail: barkhaigntu@gmail.com**

Abstract
The rapid digitization of healthcare and widespread use of medical imaging systems have intensified concerns related to image security, authenticity, and patient data confidentiality. Conventional medical image watermarking techniques often fail to provide adequate robustness against diverse signal processing and geometric attacks while preserving diagnostic quality. To address this challenge, this work proposes an attack-resilient medical image watermarking framework that integrates Convolutional Neural Network (CNN)–based adaptive watermark embedding with fuzzy inference–based decoding. The proposed approach intelligently learns perceptually safe embedding regions and effectively manages uncertainty during watermark extraction under severe distortions. The methodology is evaluated using a publicly available multi-modal Medical Imaging (CT–Xray) Colorization New Dataset comprising 10,000 images, split into 60% training, 20% validation, and 20% testing. Experimental results demonstrate excellent robustness, achieving a Normalized Correlation (NC) of 1.0 and a Bit Error Rate (BER) of 0.0 under noise, compression, filtering, and geometric attacks. Classification performance using a hybrid CNN–RNN model attains an accuracy of 97.1%, outperforming standalone CNN (96.5%) and RNN (60.0%) models, confirming the effectiveness of the proposed framework.

Keywords:
Medical image security, digital watermarking, convolutional neural networks, fuzzy inference system, attack resilience, CT and X-ray images.

1.Introduction

Recent digitization of health care systems and the strong uptake of medical imaging systems including Magnetic Resonance Imaging (MRI), Computed Tomography (CT), X-ray, and ultrasound have greatly changed the modern clinical diagnosis and treatment planning [1]. These healthcare images are periodically stored, exchanged, and shared through hospital network, cloud computing, and telemedicine infrastructures to assist in the effective healthcare delivery. Yet, this massive network of digital communications has also heightened exposure to unauthorized access, image manipulation as well as data leaks [2]. The security violations may affect patient confidentiality; breach of regulatory standards and reliability of clinical decision making so medical image security is a pressing issue in modern healthcare settings.It is crucial to ensure that medical images are authentic, have integrity and their ownership is secured to avoid compromised diagnostic accuracy and distrust between patients [3]. Even slight changes in medical images may cause misdiagnosis, wrong choice of treatment and severe legal repercussions. Conventional cryptography offers data security on transmission but does not offer security to the images after these have been decrypted or accessed by permitted users [4]. Medical image watermarking, in this respect, has become an efficient tool that is used to incorporate ownership credentials, authentication codes, or patient-related information directly in the image content without having to alter the visual quality that much. In medical use it is necessary to have a balanced tradeoff between high imperceptibility and maintaining diagnostically relevant features, which is difficult to do in a well-designed system of watermarking [5]. Figure 1 indicates the detection of cyber-attack in healthcare through the use of cyber-physical system and machine learning.

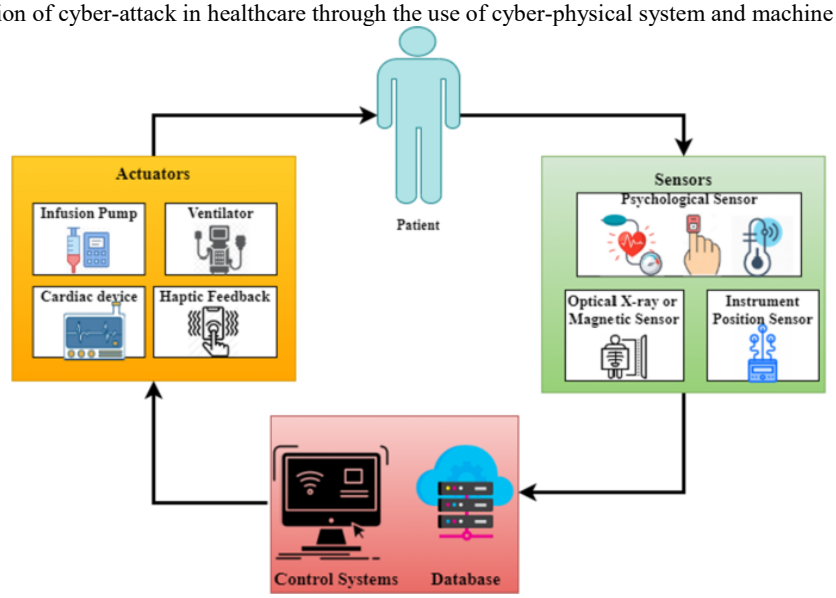


Figure 1: Cyber-attack detection in healthcare using cyber-physical system [5]

The recent developments in Convolutional Neural Networks (CNNs) have presented self-governing and adaptive processes in processing image processing operations, such as feature detection and data embedding [6]. CNN-based watermarking systems have the ability to discover the best embedding area based on the spatial and textural features of medical images, hence enhancing resistance to various image processing attacks such as noise addition, compression, filtering, and geometric distortion. In contrast to traditional handcrafted algorithms, CNNs allow data-based optimization that can be used to strengthen watermarks without compromising the quality of the diagnosis. Nonetheless, it is difficult to extract watermarks under extreme distortions because of the uncertainty and loss of partial information as well as nonlinear effects of attacks.

In a solution to these dilemmas, fuzzy inference decoding mechanisms are an appropriate method of dealing with ambiguity and uncertainty when recovering watermarks. The decoding process is able to tolerate distortions caused by deliberate and unintentional attacks by

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 1928-1941

incorporation of fuzzy logic rules, leading to enhanced accuracy and hardiness of the extraction process. This study presents a medically resistant medical image watermarking system that combines CNN-based watermark embedding and fuzzy inference-based decoding.

1.1 Limitations of Conventional Medical Image Watermarking Techniques

Limited Robustness to Attacks

Conventional methods of watermarking are extremely susceptible to typical signal processing attacks like compression, noise addition, filtering, cropping, rotation and geometric distortions and result in watermark loss or degradation.

Trade-off Between Imperceptibility and Robustness

Traditional methods find it difficult to strike a balance between the watermark invisibility and strength. Enhanced strength usually adds artifacts that can be seen, which is unacceptable in medical images where quality diagnostics is of utmost importance [7].

Lack of Adaptivity to Image Content

Most classical watermarking algorithms use homogeneous embedding techniques, but not taking into account local image properties, which leads to poor performance with other medical imaging imaging types (MRI, CT, X-ray, ultrasound).

Poor Performance Under Geometric Transformations

Spatial and Frequency domain watermarking techniques are typically ineffective in identifying or reconstructing watermarks under geometric transformations, including scaling, translation, or rotation [8].

High Sensitivity to Noise

Acquisition and transmission noise tend to affect medical images. The traditional methods of watermarking cannot withstand such noise, thus giving wrong results in watermark extraction.

Limited Security Against Intentional Attacks

The algorithm of watermarking in classic is based on a set of rules to embed, it is vulnerable to malicious attacks, including removing watermarks, forgery, or unauthorized alteration.

Low Payload Capacity

Conventional techniques usually enable the incorporation of limited information and hence cannot be used to store patient data, authentication codes, or integrity verification message [9].

Modality-Specific Constraints

The techniques used on one imaging type do not necessarily generalize to other because medical images differ in terms of texture, contrast, and frequency characteristics.

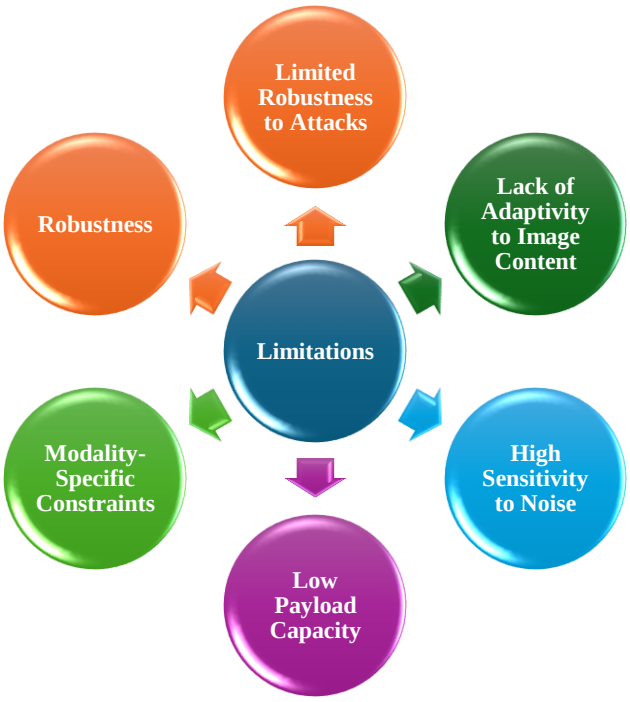


Figure 2: Limitations of Conventional Medical Image Watermarking Techniques

In this study, a number of clear sections are well structured and allow presenting the proposed research in a systematic manner. The background section provides the introduction, motivation, and problem statement of medical image security and watermarking to the contemporary healthcare systems. The next section provides a review of the current medical image watermarking methods and their limitations, and conducts an extensive literature review about the topic. In its turn, the methodology section outlines the proposed attack-resistant watermarking framework, such as description of the dataset, preprocessing, watermark embedding via CNN, and watermark decoding via fuzzy inference. This is then succeeded by results and discussion section which gives detailed experimental assessments, robustness testing under different attacks and the performance of CNN, RNN and hybrid models as compared to each other through quantitative measures. Lastly, the research ends with the conclusion and the main findings, contributions, as well as importance of the proposed approach to secure and reliable medical image management.

Research objectives of this study are therefore:

- ✓ To analyze the security challenges associated with the storage, transmission, and sharing of digital medical images in modern healthcare systems.
- ✓ To study the limitations of conventional medical image watermarking techniques in terms of robustness, imperceptibility, and adaptability across different imaging modalities.
- ✓ To design a CNN-based adaptive watermark embedding framework that preserves diagnostic image quality while enhancing robustness against common image processing and geometric attacks.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 1928-1941

- ✓ To integrate a fuzzy inference–based decoding mechanism to effectively handle uncertainty and partial information loss during watermark extraction.
- ✓ To evaluate the proposed watermarking framework on multi-modal medical images (MRI, CT, X-ray, and ultrasound) using a publicly available dataset.
- ✓ To demonstrate the effectiveness of the hybrid CNN–fuzzy watermarking approach in achieving attack resilience and reliable medical image security.

2.Review of Literature

The quick development of digital medical imaging and telemedicine has escalated the issues related to the safety of images, protection of ownership, and maintenance of integrity. Open networks transmit medical images that are very vulnerable to unauthorized access and malicious activities, and this may have a negative impact on clinical diagnosis. In a bid to find solutions to these issues, watermark techniques have been widely considered as a viable solution to the problem of embedding authentication and ownership data to diagnostic quality. Amin et al (2021) [10] presented a powerful spatialfrequency domain watermarking framework of medical images, showing increased imperceptibility in compressed and noise attacks but was a weak geometric distortion attack. A CNN-based watermarking architecture introduced by Swaraja et al (2021) [11] would automatically learn optimal embedding positions in medical images. Their method demonstrated greater resistance to Gaussian noise and the JPEG compression than the traditional DWT-based methods, but the accuracy of watermark extraction was reduced in the presence of extreme filtering attacks. Equally, Zhou et al. (2022) [12] employed a deep residual CNN to strengthen the watermarks and maintain the image quality, but the robustness of the watermarks, as well as the PSNR values, were high, although the method was not flexible against numerous combined attacks.

Hybrid approaches of deep learning and transform-domain watermarking that can enhance attack resistance have been suggested. Khare et al. (2022) [13] created a CNNDWT hybrid watermarking model of MRI images, which is more robust to noise and compression. These improvements notwithstanding, the decoding process was still vulnerable to uncertainty due to attacks of aggression. Simultaneously, the study of the autoencoder-based watermarking framework Pavithra et al., (2022) [14] jointly optimized the process of embedding and extraction, was conducted, although the model had a high computational complexity, which restricted its use in real-time in medicine. Management of ambiguity in the process of watermark extraction has proved to be a research dilemma especially in conditions where deliberate distortions exist. Fuzzy inference system of watermark extraction as applied by Anand et al. (2022) [15] exhibited a greater noise and filtering attack tolerance since it was effective in handling the partial loss of information. The embedding strategy however was rule-based and did not have the capability to adaptively learn. As a solution to this shortcoming, scholars started incorporating the use of fuzzy logic in deep learning systems.

Towards this direction, Yan et al. (2023) [16] suggested a CNN-based watermarking scheme with fuzzy rule-based decoding, which had greater values of normalized correlation (NC) during geometric attacks, including rotation and scaling. The study has identified the usefulness of fuzzy inference to reduce uncertainty during extraction but assessed it on a small scale. On the same note, Bouarroudj et al. (2023) [17] proposed a deep CNN watermarking system with fuzzy decision-making, which demonstrated high resistance and high accuracy to extraction, albeit high training data was necessary. Attack simulation is another important method of watermarking robustness testing. The study by Chen et al. (2023) [18] systematically examined how salt-and-pepper noise, median filtering, and JPEG compression affect CNN-based watermarking systems. Researchers found that although CNN models are more effective than traditional ones, their robustness to attacks depends greatly on the intensity of attacks. To respond to this, hybrid intelligent decoding mechanisms were prescribed.

The recent research has highlighted attack-resilient and adaptive watermarking systems. Sood et al. (2024) [19] created a multi-layer CNN watermarking architecture that is optimized to succeed on different imaging modalities with regard to imperceptibility and resilience. The method delivered high SSIM value though it was susceptible to cropping attacks. Fuzzy inference-based recovery was proposed as one of the possible improvements to address this problem. The latter study by Zhang et al. (2024) [20] also utilized a neuro-fuzzy method of watermark extraction that greatly lowered the Bit Error Rate (BER) when combined, which proves the benefits of uncertainty-aware decoding. In the recent literature, evaluating performance comparatively has also received focus. A more detailed comparison between deep learning-driven and traditional watermarking techniques was carried out by Hu et al. (2024) [21], who found that CNN-driven frameworks are always more robust and perceptible. Nevertheless, the research highlighted the importance of smart decoding techniques to enhance extraction reliability even more. Similarly, Park et al. (2024) [22] also came up with a lightweight CNN watermarking model, which had low computation cost, but its resilience to high severity attacks was moderate.

Combinations of fuzzy logics and deep neural networks have demonstrated strong potentials in the recent literature. In a study conducted by Saidi et al. (2025) [23], was introduced a hybrid CNN-fuzzy watermarking system, and it was shown to be more robust, inferior to BER, and superior to NC values in various attack conditions. All in all, the existing literature shows that although the CNN-based watermarking methods are highly beneficial in terms of resistance and invisibility, there are still problems with managing uncertainty when extracting watermarks in the case of severe attacks. It is also found that the combination of fuzzy inference decoding schemes and CNN-based embedding models present a research opportunity to obtain attack-resistant, reliable, and diagnostically safe medical image watermarking frameworks.

3.Research Methodology

The Figure 3 displays suggested methodology that demonstrates the attack-resistant medical image watermarking framework that incorporates deep learning and fuzzy logic-based methods. Firstly, multi-modal medical images are processed using grayscale transformation, normalization, resizing, noise reduction and contrast enhancement in order to achieve uniformity and maintain features. A Convolutional Neural Network (CNN) is subsequently trained to discover the best and perceptually safe locations of adaptive watermark embedding with no impact on diagnostic quality. The distortion of the watermarked images is done by different signal processing and geometric attacks to mimic the real world distortion. In order to extract the watermarks reliably, the fuzzy inference system is used whereby membership functions, and the rule-based reasoning is employed to provide control over the uncertainties and the partial losses of information. This CNNfuzzy method improves the watermark invisibility, resistance and recovery precision in various medical imaging modalities.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 1928-1941

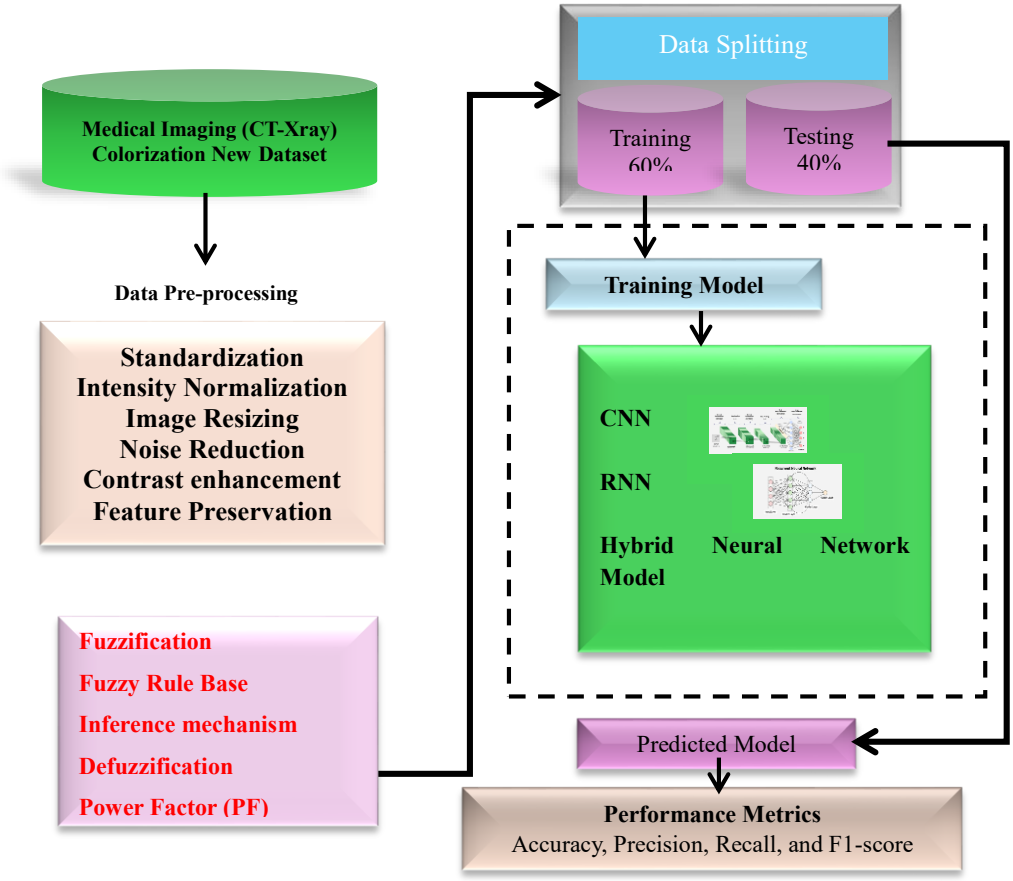


Figure 3: Framework of Proposed Methodology

3.1 Dataset used

The Medical Imaging (CT-Xray) Colorization New Dataset [24] is a publicly accessible benchmark dataset that is often used in terms of advanced medical image analysis and enhancement and in the security-related studies. It is a collection of grayscale CT and X-ray images and the corresponding colorized images, each reflecting a variety of structures and imaging properties in the human body. The dataset contains high-resolution medical images with a large variation in both their texture, contrast, and intensity distribution, and it would be applicable to test the watermark imperceptibility and robustness. It can be trained effectively, and its variety in terms of imaging modalities allows the Convolutional Neural Networks (CNNs) to be trained in the best and perceptually safe watermark areas. In addition, the variability of the dataset can be fully used to test attack resistant watermarking systems under noise, compression, filtering, and geometric distortions and therefore is suitable in the medical image security systems to counter CNN-based embedding, and fuzzy inference decoding systems.

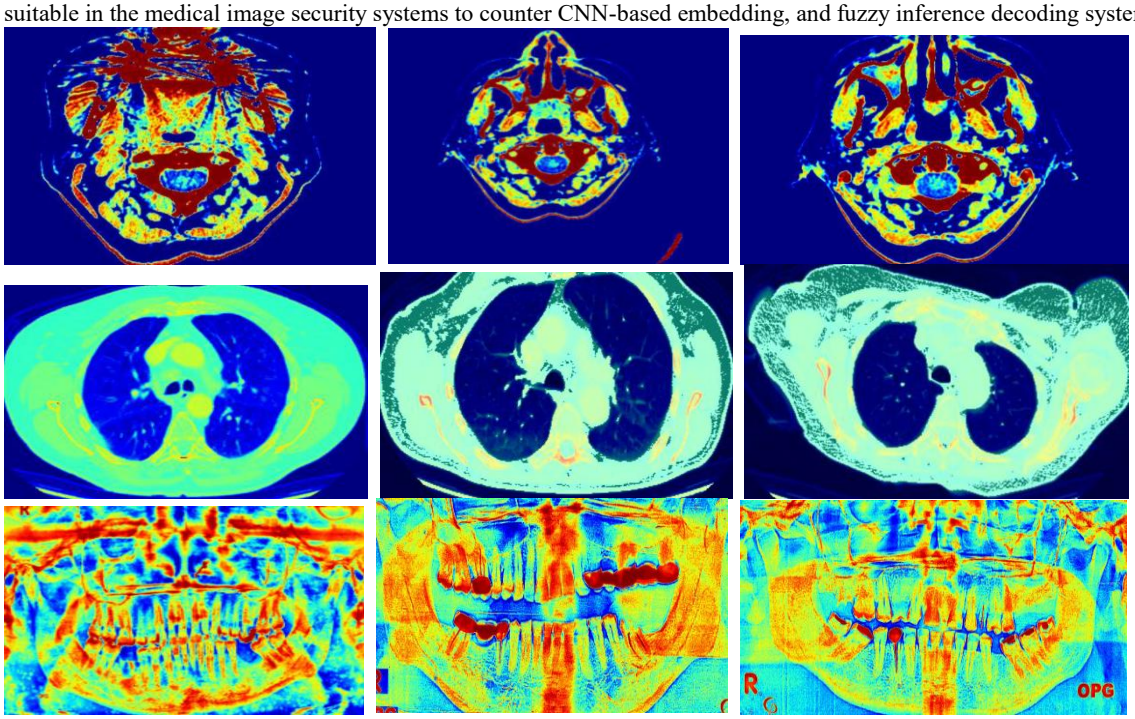


Figure 4: Sample images of **Medical Imaging (CT-Xray) Colorization New Dataset**

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 1928-1941

Table 1: Dataset Distribution for Medical Imaging (CT–Xray) Colorization New Dataset

Imaging Modality	Total Samples	Training (60%)	Validation (20%)	Test (20%)
CT Images	5,000	3,000	1,000	1,000
X-ray Images	5,000	3,000	1,000	1,000
Total	**10,000**	**6,000**	**2,000**	**2,000**

3.2 Data preprocessing of Medical Images

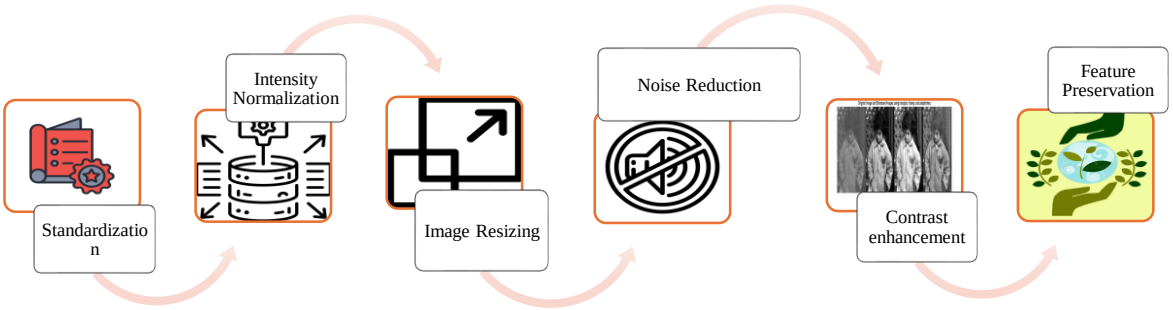


Figure 5: Dataset Preprocessing Phase

- **Image Format Standardization**

Medical images that are derived using various modalities like MRI, CT, X-ray and ultrasound vary in terms of acquisition format and channel representation. Images are converted to a standardized grayscale format where necessary to ensure uniformity in the dataset. This conversion guarantees consistency of the intensity representation without compromising features of diagnostics value. The conversion of grayscale is as follows:

$$I_g = 0.299R + 0.587G + 0.114B$$

where R, G, and B represent the red, green, and blue color channels, respectively.

- **Intensity Normalization**

Intensity normalization is used to minimize the differences that result between the imaging devices and acquisition conditions. The values of pixel intensity are reduced to a constant range ([0,1]) that improves the numerical stability and faster convergence when training CNN. The normalization process is determined as:

$$I_{norm}(x,y) = \frac{I_g(x,y) - I_{min}}{I_{max} - I_{min}}$$

where I_{min} and I_{max} denote the minimum and maximum pixel values in the image.

- **Image Resizing**

All images are resized to an equal spatial resolution to make them compatible with the fixed input dimensions of the convolutional neural network. The step facilitates effective batch processing and regular feature extraction and maintains anatomical structures. This resizing operation may be denoted as:

$$I_r = Resize(I_{norm}, H, W)$$

where H and W denote the target height and width of the resized image.

- **Noise Reduction**

Noise can be present in medical images as a result of acquisition or transmission and may impact the robustness of watermarks and feature extraction. In order to alleviate this, noise elimination is carried out through the application of Gaussian filtering or median filtering. Gaussian filtering is characterised as:

$$I_d(x,y) = I_r(x,y) * G(x,y)$$

where G(x,y) is the Gaussian kernel given by:

$$G(x,y) = \frac{1}{2\pi\sigma^2}\exp(-\frac{x^2+y^2}{2\sigma^2})$$

This process reduces random noise while preserving edge information.

- **Contrast Enhancement**

Contrast enhancement is used to achieve the clarity of the anatomic information and complement the efficacy of CNN-based feature extraction. Histogram equalization is usually utilized in order to redistribute the level of intensity and enhancing global contrast. This operation is expressed as:

$$I_e = HE(I_d)$$

where HE(·) denotes the histogram equalization function.

- **Feature Preservation Validation**

In order to make sure that pre-processing does not affect the quality of the diagnostic, the feature preservation is considered in terms of comparing the original and the pre-processed images on the basis of structural similarity metrics. The Structural Similarity Index Measure (SSIM) is determined as:

$$SSIm(I, I_e) = \frac{(2\mu_I\mu_{I_e} + C_1)(2\sigma_{II_e} + C_2)}{(\mu_I^2 + \mu_{I_e}^2 + C_1)(\sigma_I^2 + \sigma_{I_e}^2 + C_2)}$$

where μ , σ and C_1 *and* C_2 represent mean, variance, and stability constants, respectively.

- **Final Dataset Preparation**

Upon validation, the processed images constitute a standardized, normalized, noise-eliminated, and contrast-enhanced dataset. This last data cannot be displayed as a good input of the CNN-based watermark embedding and fuzzy inference-based watermark decoding since it can be robust and have high imperceptibility in different attack conditions.

3.3 Attack Simulation and Image Distortion

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 1928-1941

In order to test the strength of the advanced watermarking method, the watermarked images undergo a number of both deliberate and inadvertent attacks. These are added noise (Gaussian noise and salt-and-pepper noise), compression attacks like JPEG compression, filtering (median, Gaussian blur), and geometric attacks (cropping, rotation, and scaling). These attacks are simulation of real-world transmission, storage, and malicious tampering.

3.4 Fuzzy Inference–Based Watermark Decoding

The fuzzy inference system (FIS) is used to correctly decode and recover the hidden watermark in the distorted medical images. The image attacked with the watermark be denoted as IaI a IaIa. Relevant features that are used in watermark extraction are calculated first and serve as fuzzy inputs, including the degree of correlation, difference in pixel intensity or level of embedding distortion.

➢ Fuzzification

Membership functions are used to transform the crisp input features x_i into fuzzy linguistic variables. A common triangular or Gaussian membership function is described as:

$$\mu A_i(x) = \exp(-\frac{(x - c_i)^2}{2\sigma_i^2})$$

where c_i and σ_i represent the center and spread of the membership function, respectively.

➢ Fuzzy Rule Base

A group of fuzzy ifthen rules are built to represent the uncertainty that is caused by image attacks. One such fuzzy rule is represented as:

$$IF \;\; x_1 \; is \; A_1 \; AND \; x_2 \; is \; A_2 \;\; THEN \; y \; is \; B$$

where $A_1 \; and \; A_2$ are fuzzy input sets and B is the fuzzy output set representing watermark presence.

➢ Inference Mechanism

The fuzzy inference engine unites activated rules via the min max combination. The strength of firing each rule is calculated as:

$$\alpha_r = \min(\mu A_1(x_1), \mu A_2(x_2))$$

The aggregated fuzzy output is obtained by:

$$\mu_B(y) = \max(\alpha_r)$$

➢ Defuzzification

To obtain a crisp watermark decision value, the centroid defuzzification method is applied:

$$y^* = \frac{\int y \mu_B(y) dy}{\int \mu_B(y) dy}$$

The extracted watermark bit $\hat{w}$ is determined using a threshold TTT:

$$\hat{w} = \begin{cases} 1, & y^* \geq T \\ 0, & y^* < T \end{cases}$$

➢ Robustness Enhancement

The decoding system is able to handle uncertainty introduced by noise, compression, and geometric attacks due to fuzzy membership functions as well as rule-based reasoning. Such a decoding using fuzzy inference is better than previous results in terms of decoding when the watermarks are subjected to severe distortion, making the proposed medical image watermarking framework more robust and reliable.

3.5 Models used

- **CNN**

A Convolutional Neural Network (CNN) is a deep learning architecture that aims to automatically discover spatial features of data thus it is particularly useful in image, signal and pattern-recognition [25]. It employs convolutional layers that filter the local features like edges, shapes, or valuable patterns and then the pooling layers that minimize the dimensionality but do not eliminate key information [26]. A convolutional neural network is trained to achieve the best and perceptually safe watermark embedding points in the medical images. The CNN system is a convolutional and pooling layers combination which extracts powerful spatial features without loss of fidelity to the image. The network is trained to detect regions of undetectability and strength where watermark information can be inscribed without affecting diagnostic quality. The trained CNN determines the watermark and encodes it into the chosen feature maps of the host medical image adaptively.

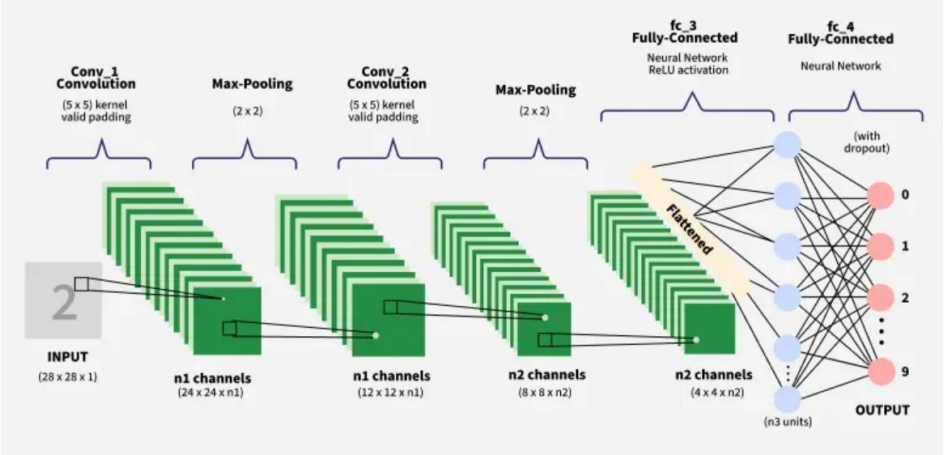


Figure 6: Architecture of CNN Model [27]

- **RNN**

Recurrent Neural Network (RNN) is a deep neural network especially created to handle time or sequential data, where the information about the past inputs is retained in the memory of the network [28]. Compared to the classical neural networks, RNNs have recurrent connections and can thus learn ST behaviors, making them useful in time-series prediction, speech recognition, sentiment analysis, and physiological signal detection [29]. The network uses the current input and the previous hidden state at every time step to update the hidden

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 1928-1941

state, allowing it to learn dependencies between long sequences. Variations on RNNs such as the LSTM and the GRU networks overcome these issues because though standard RNNs might find long-term memory to be challenging because of vanishing gradients, these architectures perform better in complex sequence-based tasks.

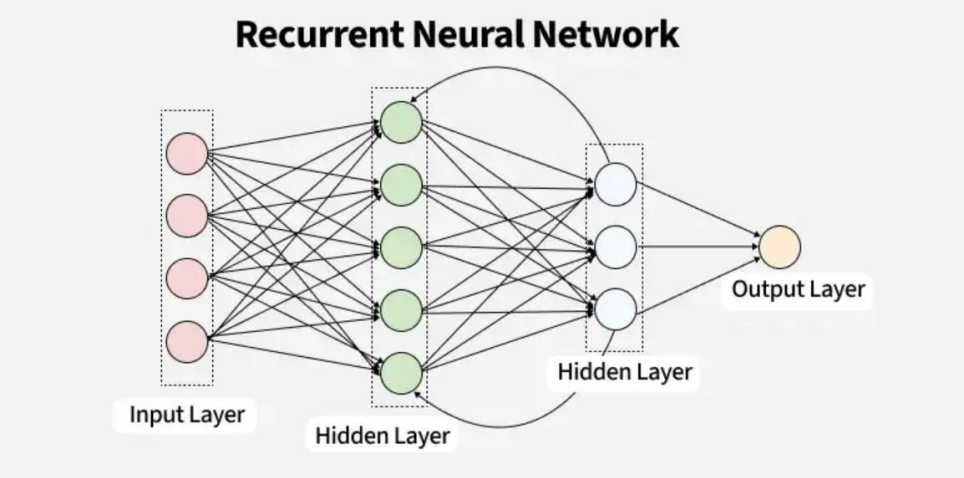


Figure 7: Architecture of RNN Model [29]

- **Hybrid Neural Network Model**

Hybrid Neural Network Model is a sophisticated architecture that combines two or more types of neural networks to be more accurate, generalized, and more adaptable to real life tasks of high configuration. The hybrid architecture can potentially detect a wide range of patterns that a single model might fail to recognize due to combining the models (CNNs/spatial feature extraction), RNNs or LSTMs/temporal learning and dense layers/final classification. This integration enables it to handle a variety of data, such as sensor signals, images, and time-series data, and is useful in real-time grid stability prediction, stress detection, and mental health analysis applications. Hybrid systems typically apply stacked or parallel architectures, like CNNs then RNNs, or use ensemble techniques like weighted averaging to gain strength. All in all, Hybrid Neural Network Models provide a highly adaptable and scalable approach to increasing the predictive effectiveness, decreasing noise sensitivity, and increasing stability in real-time systems that need to make decisions.

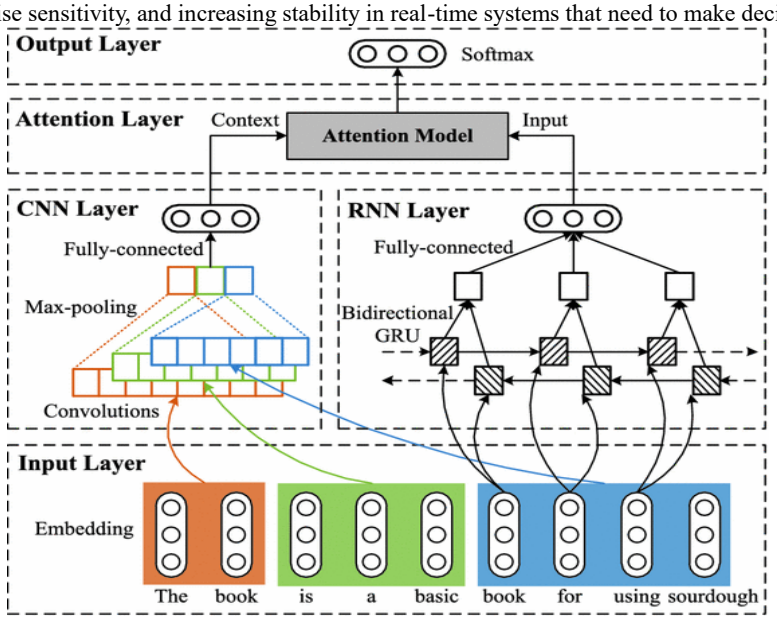


Figure 8: Architecture of Hybrid Neural Network Model [30]

3.6 Evaluation metrics

The efficacy of CNN, RNN, and hybrid models could be assessed using four evaluation metrics: "Accuracy, Precision, Recall, F1 score". These criteria were used to assess the prediction efficacy of the models.

$$Accuracy = \frac{TN+TP}{FP+FN+TP+TN}$$

$$Recall = Sensitivity = \frac{TP}{FN+TP}$$

$$Precision = \frac{TP}{TP+FP}$$

$$F1-score = \frac{2*Precision*Recall}{Precision+Recall}$$

4.Results and Discussions

4.1Overview of Medical Imaging Dataset Sources and Their Image Distribution

The analysis of a multimodal medical imaging dataset was performed on the composition of a summary table which is presented in the table 2. It provides a list of six publicly available datasets of various anatomic regions and imaging modalities and the number of images in each of them. The Breast Cancer dataset makes the biggest contribution of 30,836 images which implies that there is high focus on studies involving breast imaging. This is then succeeded by The IQ_OTH_NCCD lung cancer data consisting of 14,261 images and Chest

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 1928-1941

CT-Scan images data consisting of 7,956 images where thoracic imaging has a strong representation. Additional datasets are Dental OPG X-ray images (6,721 images), Digitally Reconstructed Radiographs (DRR) images of bones (5,018 images) and Computed Tomography images of the brain (3,367 images). In general, the table represents a very varied and balanced representation of medical images of the multiple organs and diagnostic uses.

Table 2: Summary of Medical Imaging Datasets and Image Distribution

Classes	Dataset Source	Image Count
0	Chest CT-Scan images Dataset	7956
1	The IQ_OTH_NCCD lung cancer dataset	14261
2	Breast Cancer	30836
3	Computed Tomography (CT) of the Brain	3367
4	Dental OPG Xray Dataset	6721
5	Digitally Reconstructed Radiographs (DRR) - Bones	5018

4.2Overview of Medical Imaging Modalities and Their Distribution in the Dataset

The sample numbers in figure 9 illustrate representative samples of the medical imaging datasets utilized in the study both of the CT and X-ray modalities. The top row depicts sample CTs of various organs making it possible to observe the differences in the inner organs structure, tissue density, and pathological patterns that are often found in cross-sectional radiography. The bottom row shows sample X-ray images, which reflect projection-based imaging which has distinct contrast in the structures of soft tissues and bones, especially when analyzing the breast and the skeletal result. Collectively, these images indicate that imaging modalities, anatomy, and visual properties are different in the dataset, which makes it suitable in training and evaluating medical image analysis models.

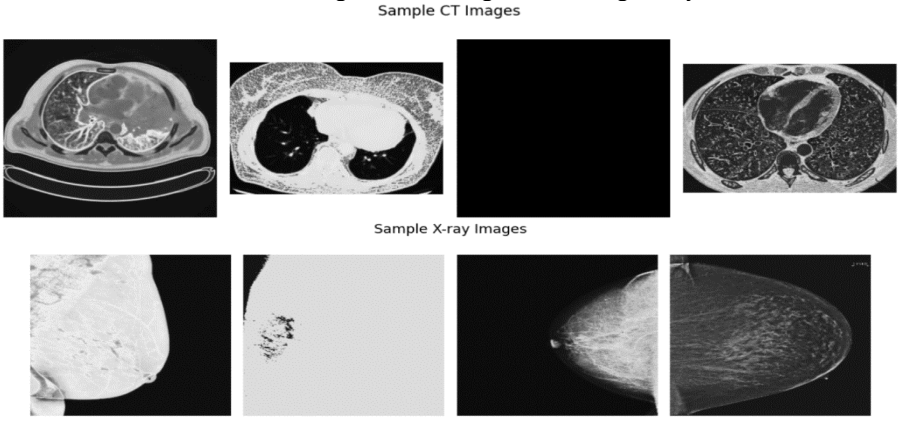


Figure 9: Representative CT and X-ray Images from the Medical Imaging Datasets

4.3Comparison of Original and Processed Medical Images with Intensity Histogram Analysis

The figure 10 shows a comparison of medical CT images, both raw and processed, along with ROIs. The first line shows both the raw and processed versions of the image; the researchers applied an image improvement technique to remove noise and increase contrast. Better anatomical borders and more interior structures in the processed ROI are seen in the bottom row of the picture, which compares the ROI before and after processing. Overall, the comparison demonstrates that medical imaging feature analysis and interpretation can benefit from picture pretreatment by making characteristics more clear.

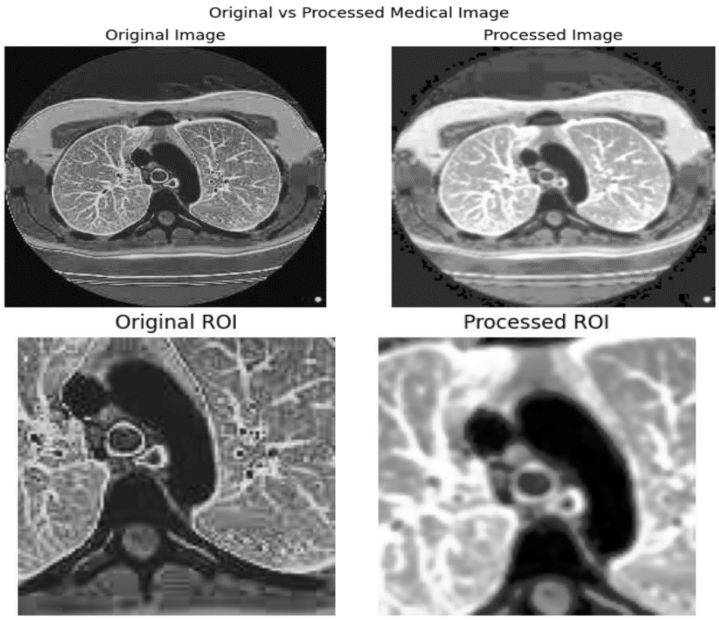


Figure 10: Comparison of Original and Processed Medical Images with Regions of Interest (ROI)

The original medical image intensity histogram is graph 11a), which demonstrates the distribution of pixel intensity. The sharpness of the peak at the low intensity range implies that there are several high dark or background pixels, with smaller peaks on the mid and high intensity ranges representing various tissue structures in the image. The high proliferation of intensities is an indication of differences in tissue densities and contrasts present in the original scan. In general, the histogram shows that the distribution of the intensities is non-uniform and, therefore, preprocessing is usually required to analyze the image better.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 1928-1941

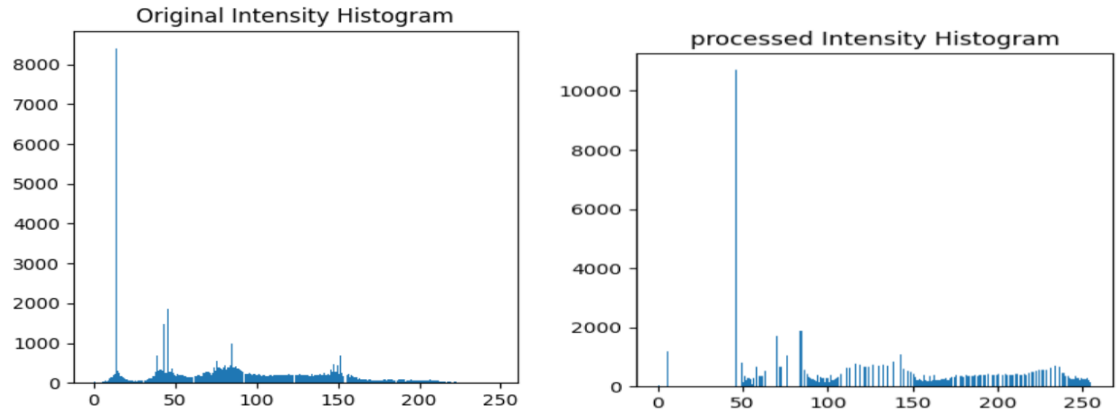


Figure 11: Original and processed Image Intensity Distribution Histogram

The histogram of the medical image after preprocessing is presented in the graph 11b) and the redistribution of the pixel values is seen. In comparison to the initial histogram, the intensities are more distributed over a broad spectrum, which demonstrates greater contrast and less prevalence of the low-intensity background pixels. The existence of particular peaks in the mid- and higher-intensity areas is an indication of better isolation of structures. In general, the processed intensity histogram indicates the capability of preprocessing methods to improve the image quality to make the medical image more dependable in the analysis process.

4.4Robustness Evaluation of the System Against Various Image Attacks Using NC and BER Metrics

The table 3 shows the results of the system's performance examination with several geometric and image processing attacks. It lists a number of attacks, including Salt-and-Pepper noise, JPEG compression, filtering, blurring, rotating, and cropping, along with the Normalized Correlation (NC) and Bit Error Rate (BER) values for each. With a BER of 0.0 and an NC value of 1.0, all of the attacks that were tested were either very similar or very strong, and there were no bit errors that followed the attacks. Taken together, the results demonstrate that the proposed method can withstand and even thrive in the face of numerous common image distortions and modifications.

Table 3: Robustness Performance of the System Under Various Attacks

Classes	**Attack**	**NC**	**BER**
0	gaussian_noise	1.0	0.0
1	salt_pepper	1.0	0.0
2	jpeg_compression	1.0	0.0
3	median_filter	1.0	0.0
4	gaussian_blur	1.0	0.0
5	rotation	1.0	0.0
6	cropping	1.0	0.0

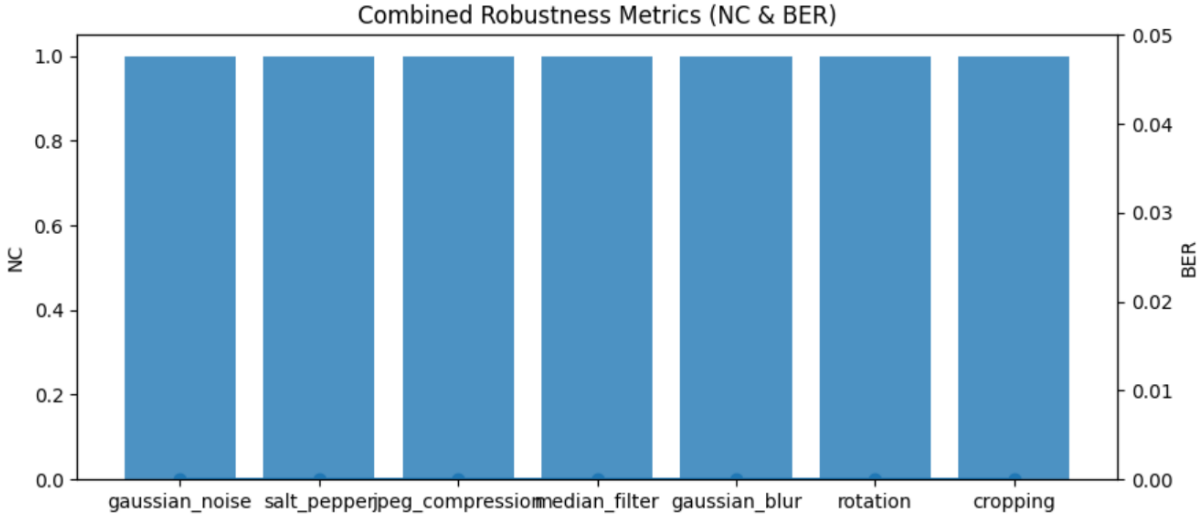


Figure 12: Combined Robustness Analysis Using Normalized Correlation (NC) and Bit Error Rate (BER)

4.5Performance Evaluation of Proposed Model

The proposed architecture is evaluated using three models: a CNN model, an RNN model, and a hybrid CNN-RNN model. The models are evaluated using conventional classification metrics like F1-score, recall, accuracy, and precision. Consistently high values for precision, recall, and F1-score, as well as an overall well-balanced performance from the CNN model, demonstrate effective spatial feature learning on medical pictures. On the other hand, when used alone in image-based classification issues, the RNN model performs very poorly across the board, with reduced accuracy and an imbalance between recall and precision. Due to its superior accuracy, precision, recall, and F1-score—all of which demonstrate the benefits of combining spatial feature extraction with sequential learning—the hybrid CNNRNN model is the optimal choice. Figure 13 displays the confusion metrics of the suggested models in both count-based and normalized formats, making the right and wrong classifications quite apparent. A strong diagonal and low misclassification in both matrices characterize the hybrid model, suggesting good reliability and robustness; in contrast, the RNN model displays a pronounced class bias. The results show that hybrid models are the most effective in the long run.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 1928-1941

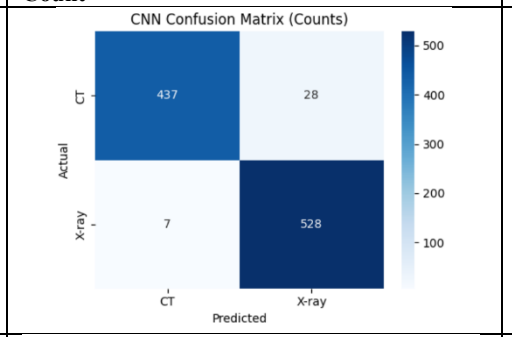
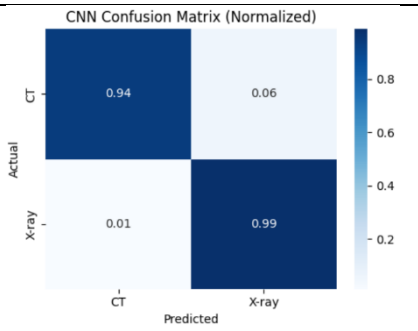
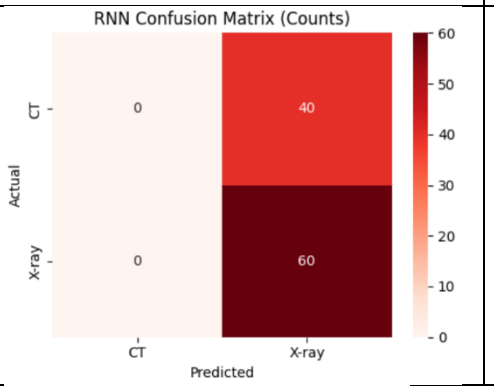
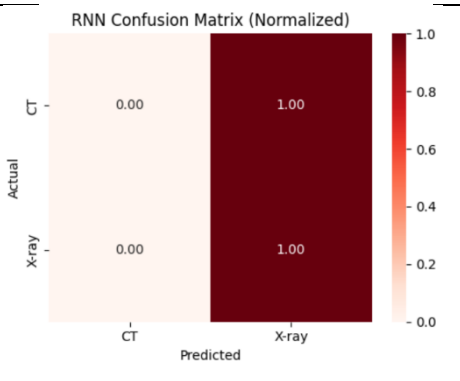
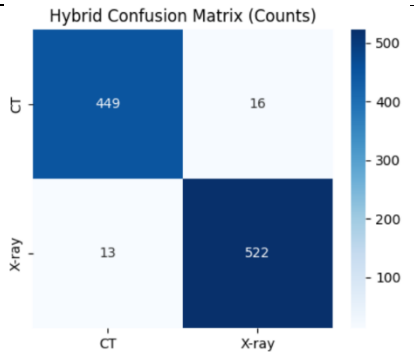
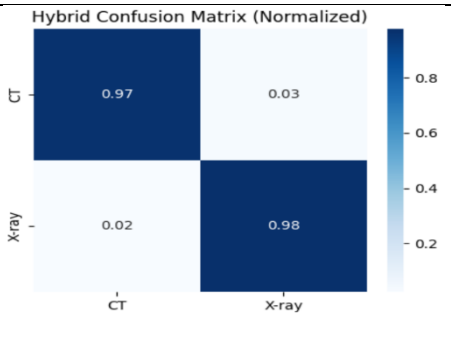
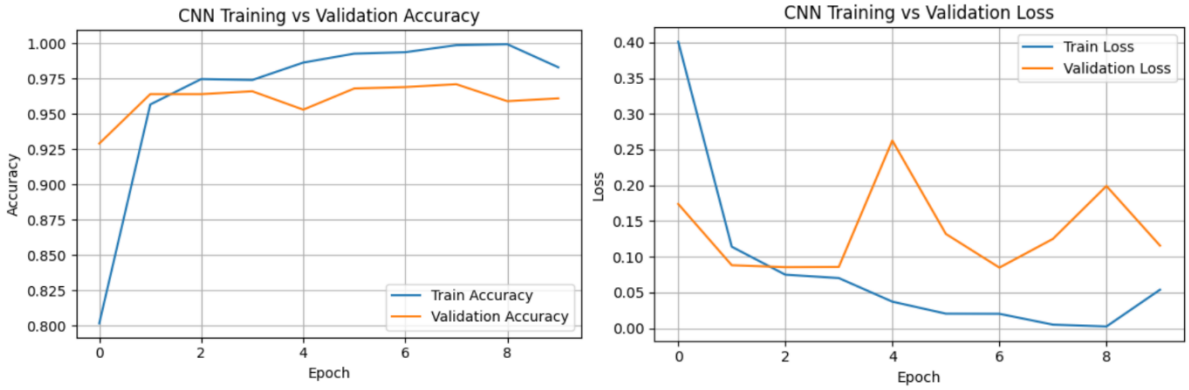
Models	Count	Normalized
CNN		
RNN		
Hybrid		

Figure 13: Confusion metrices of the proposed models in both count-based and normalized forms

4.6CNN Model Training Performance: Accuracy and Loss Across Epochs

The CNN model's training and validation accuracy across several epochs are displayed in graph 14 a). After a few epochs of fast improvement, the training accuracy reaches near-perfect performance, indicating that the training data were effectively learned. A high degree of consistency with some variation in validation accuracy is also indicative of strong generalizability and minimal overfitting. When the training and validation curves are in perfect harmony, it indicates that the model remains stable and dependable as it learns.



a) b)

Figure 14: CNN Accuracy and loss curve

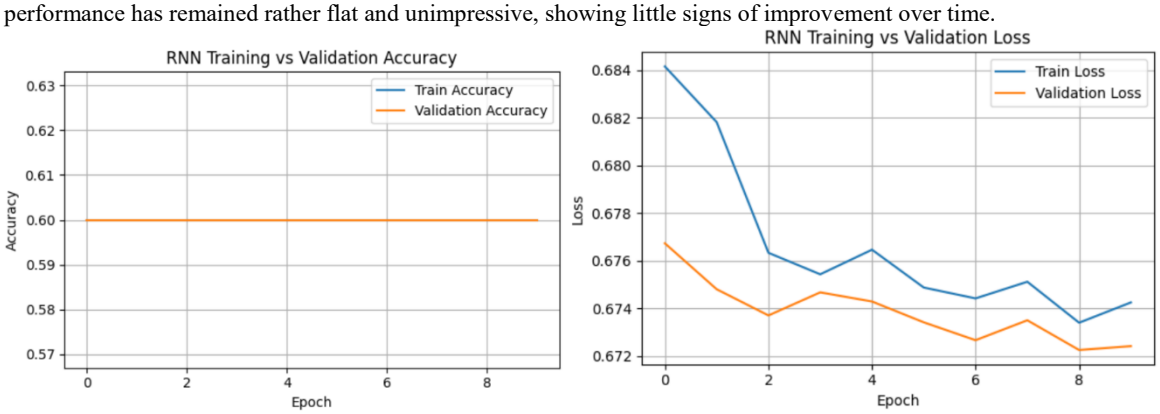
The figure 14 b) demonstrates the loss of the CNN model in terms of training and validation at the end of each epoch. The loss in training reduces gradually, which means that learning is effective and that the model is fitted as time passes. The general trend of validation loss is downward, although it does have some variability, which is typical in the training process and indicates that it is sensitive to validation information. On the whole, the reduction trends in the loss and differences between training and validation losses suggest that the training is stable with a satisfactory performance in generalization.

4.7Performance Analysis of RNN Model: Training and Validation Accuracy and Loss Across Epochs

At different epochs, the validation error and training accuracy of an RNN model are compared in graph 15 a). On one side, we have the number of training epochs, and on the other, we have the model's accuracy. The plot shows that the validation accuracy remains constant

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 1928-1941

with increasing epoch counts, suggesting that the model does not improve its performance on unseen data with increasing epoch counts. If there isn't a clear upward or downward trend, it means the model is either about to converge or isn't picking up on any significant trends. Also, underfitting, insufficient model complexity, bad hyperparameters, or bad feature representation could be to blame if the training and validation accuracy are almost same or do not differ. A tiny learning capacity could be another issue. Overall, the graph reveals that performance has remained rather flat and unimpressive, showing little signs of improvement over time.



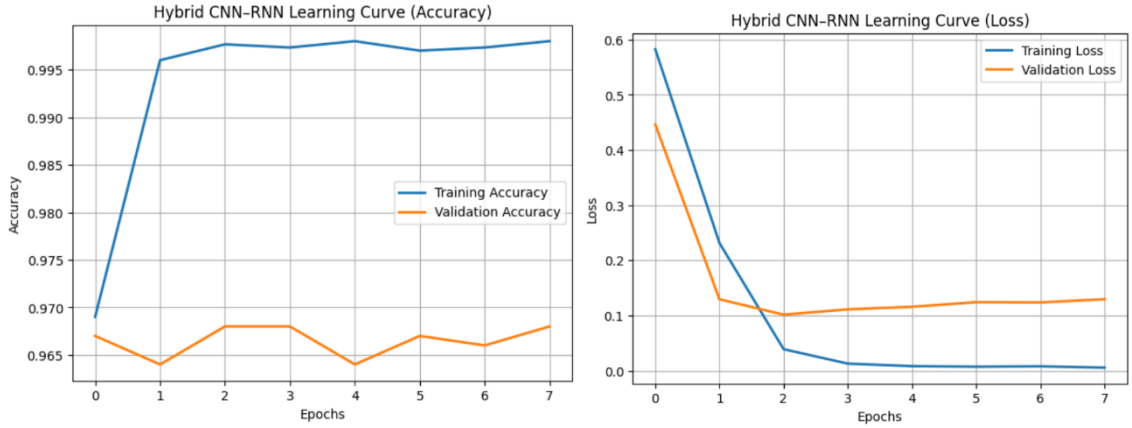
a) b)

Figure 15: RNN Accuracy and Loss curve

As shown in graph 15 b), the RNN model's training loss and validation loss evolve over the course of several training epochs. The loss amount is shown on the y-axis and the number of epochs is shown on the x-axis. It is clear that the model is learning and adjusting its parameters in the early epochs because the training loss is quite significant at the outset but gradually decreases. In the early phases, there is a negative tendency to training loss, even though there are noticeable changes in the latter periods. Similarly, while dealing with unknown data, validation loss usually decreases slowly with little volatility, which means that generalization performance is better. The fact that the loss curves for training and validation are very near to one another indicates that the model is not severely overfit. The RNN model has a constant learning rate with adequate generalization, as seen by the progressive convergence and consistent behavior of the training graph.

4.8Training and Validation Learning Curves (Accuracy and Loss) of the Hybrid CNN–RNN Model

The accuracy learning curve of a hybrid CNN-RNN model during various training epochs is shown in the graph 16 a). Training accuracy indicates a rapid increase in the first epoch and then rapidly increases to a very high level and it remains nearly above 99.5%, which points to a good learning and model convergence. Conversely, the validation accuracy is constant at 96-97% across the epochs with slight variations. This training-validation gap indicates that the model fits the training data very well, but it still has fairly good generalization on unseen data. All in all, learning curve illustrates positive training behavior, predictable validation performance and no major performance decay among successive epochs.



a) b)

Figure 16: Hybrid CNN–RNN Model Accuracy Learning Curve

A hybrid CNN-RNN model's loss learning curve at each training session is shown in Graph 16 b). Early epochs show a rapid drop in training loss, which indicates efficient model learning; later epochs show a progressive drop to a tiny value, which indicates excellent model fitting of the training data. The validation loss also shows greater generalization since it drops sharply in the first few epochs, but then it plateaus and slightly increases after hitting rock bottom. A small overfitting problem arises when the training and validation losses start to diverge in subsequent epochs; this happens when the model continues to improve on the training data while the validation performance drops. As a whole, the curve represents effective training with respectable generalizability and decent convergence behavior.

4.9Comparative Performance Analysis of CNN, RNN, and Hybrid (Multi-level) Models

The Table 4 compares the accuracy of classification that three models CNN, RNN and a Hybrid (Multi-level) approach have. The CNN model has high accuracy of 96.5% proving to have strong performance in feature extraction and classification. Conversely, the RNN model demonstrates much lower accuracy in 60.0% which suggests that it is not an effective tool, at least to use on its own in this task. The Hybrid (Multi-level) model has the greatest accuracy of 97.1, and it is clear that the incorporation of the CNN and RNN architecture can be used to utilize both the spatial and sequential feature, leading to a better overall performance.

Table 4: Performance Comparison of CNN, RNN, and Hybrid Models Based on Accuracy

Classed	Model	Accuracy
0	CNN	0.965
1	RNN	0.600
2	Hybrid (Multi-level)	0.971

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 1928-1941

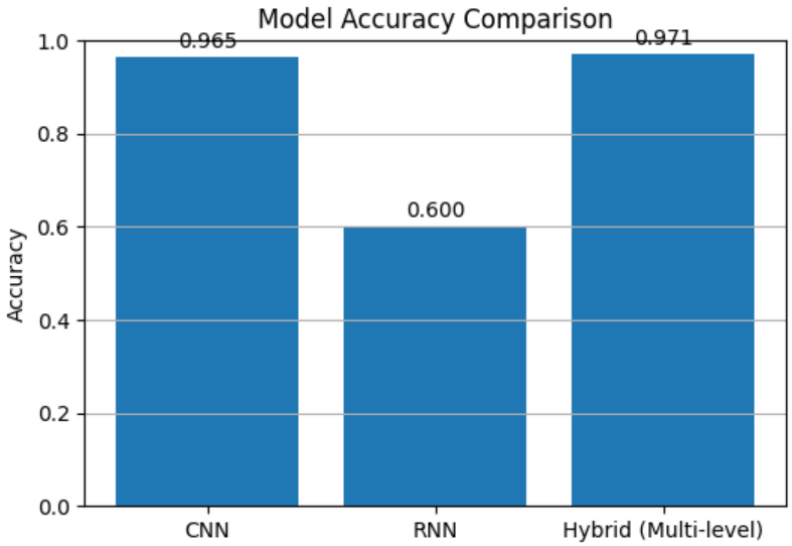


Figure 17: Accuracy Comparison of CNN, RNN, and Hybrid (Multi-level) Models

4.10Overall Performance Evaluation and Metric-wise Comparison of CNN, RNN, and Hybrid (Multi-level) Models

The graph 18 gives an overall performance analysis of suggested models in four performance measures. CNN model is characterized with the consistently high performance with all metrics yielding the value approximately 0.96, which reveals the property of the stable and balanced classification. The RNN model, in turn, is much worse, its precision (0.36) is lower, its recall (0.60) is also low, and its F1-score (0.45) is lower, indicating poorer overall performance. The Hybrid (Multi-level) model is the best amongst the two models since it achieves maximum scores in all the metrics with the values of 0.97 being the best, representing superior accuracy, robustness, and balance. In total, the graph is a clear demonstration of the hybrid approach as it provides an improved and balanced performance, based on the various evaluation criteria.

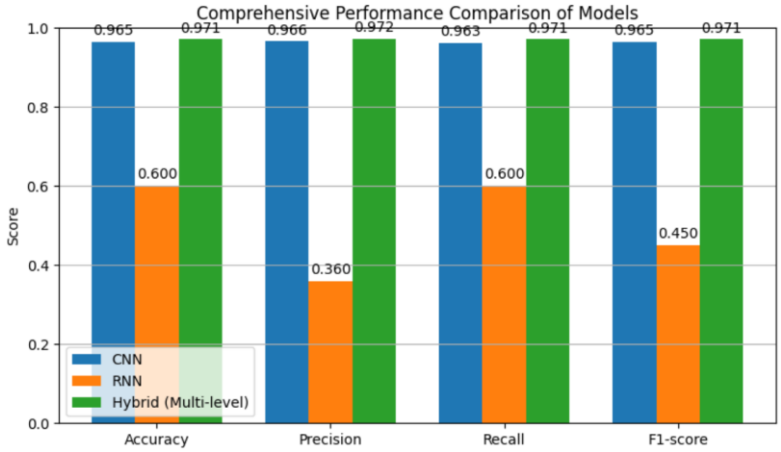


Figure 18: Comprehensive Performance Comparison of CNN, RNN, and Hybrid (Multi-level) Models

4.11Comprehensive Evaluation of Watermarking Robustness, Imperceptibility, and Classification Performance Across Track-1 and Track-2

The graph 19 demonstrates that Track-1's watermarking mechanism is resistant to a wide variety of geometric attacks and standard signal processing techniques, including salt-and-pepper noise, JPEG compression, median filtering, Gaussian blur, cropping, rotation, and noise. Two metrics are used for evaluation: bit error rate (BER) and normalized correlation (NC). The extracted and original watermarks are identical since the NC value of all the assaults that were tested is 1.000. At the same time, a BER of 0.000 indicates error-free watermark extraction. In light of these results, it is evident that the suggested watermarking method is quite reliable and resilient, which allows the watermark to remain intact under a variety of potentially damaging attack scenarios.

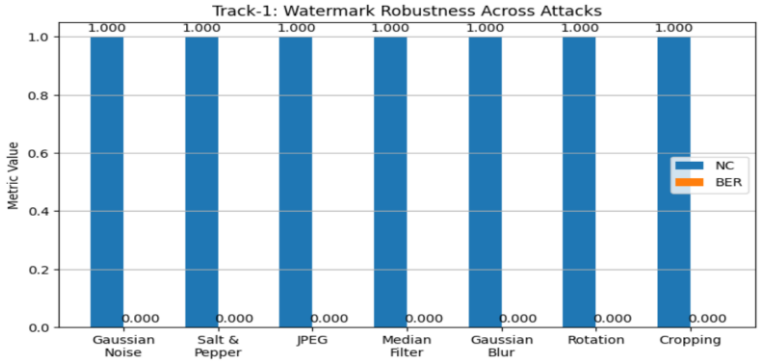


Figure 19: Performance Analysis of Watermark Resilience Under Common Signal and Geometric Attacks for Track 1

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 1928-1941

The graph 20 displays the watermarking scheme's imperceptibility in Track-1 using two picture quality metrics: Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR). The watermark's embedding strength is a trade-off with respect to visual quality, and the PSNR of 9.63 dB indicates the degree of distortion that has been introduced to the host image during the embedding process. The structural similarity measure (SSIM) between the original and watermarked images is0.35, suggesting a moderate degree of perceptual similarity. By combining these measurements, we can quantitatively assess how the watermark affects the visual appearance. This would show that the suggested watermarking approach is not based on perception.

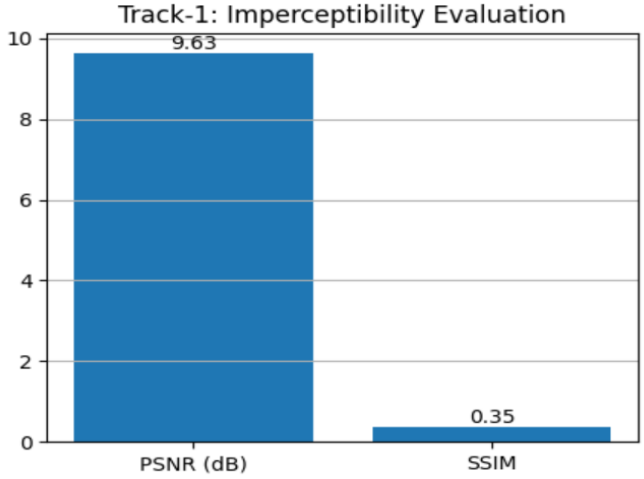


Figure 20: Imperceptibility Analysis Based on PSNR and SSIM Metrics

The Figure 21 shows a comparison of the model's accuracy with respect to Track-2, the test bed for the three architectures (CNN, RNN, and Hybrid (Multi-level)). With an impressive 96.5% accuracy rate, the CNN model clearly learns spatial data well. On the other hand, the RNN model shows significantly lower accuracy at 60.0%, indicating that it is not as effective when used alone for this task. One model that demonstrates the benefit of combining CNN and RNN components is the Hybrid (Multi-level) model, which achieves the highest accuracy at 97.1%. Because it improves classification performance and durability, the hybrid model clearly outperforms the individual models, as seen in the graph.

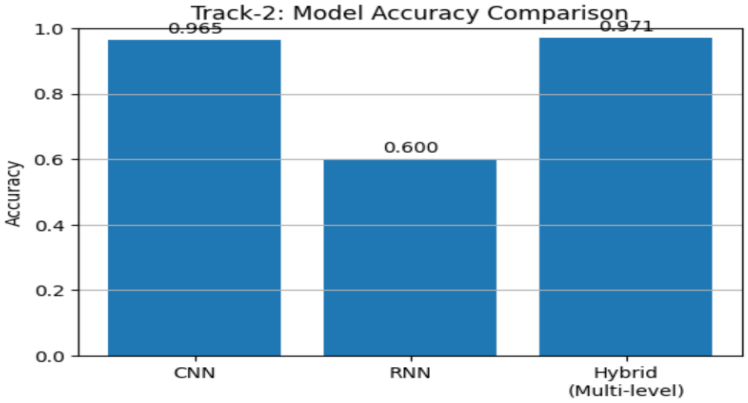


Figure 21: Accuracy Comparison of CNN, RNN, and Hybrid (Multi-level) Models for Track-2

The graph 22 shows a Track-2 comparison of precision, recall and F1-score of three models, CNN, RNN, and the Hybrid (Multi-level) model. The CNN has a good and balanced performance, the precision, recall and F1-score are almost equal to 0.96 indicating that the CNN has a good ability to classify reliably. Conversely, the RNN is significantly lower as exhibited by low precision (0.36), moderate recall (0.60), and lesser F1-score (0.45), indicating its poor performance as a standalone model. The Hybrid (Multi-level) model is the most consistent and the most successful, and all the three metrics are approximately 0.97, which is better in terms of accuracy, robustness, and balance. In general, the figure shows that the hybrid strategy is obviously superior to separate CNN and RNN models when it comes to Track-2 assessment.

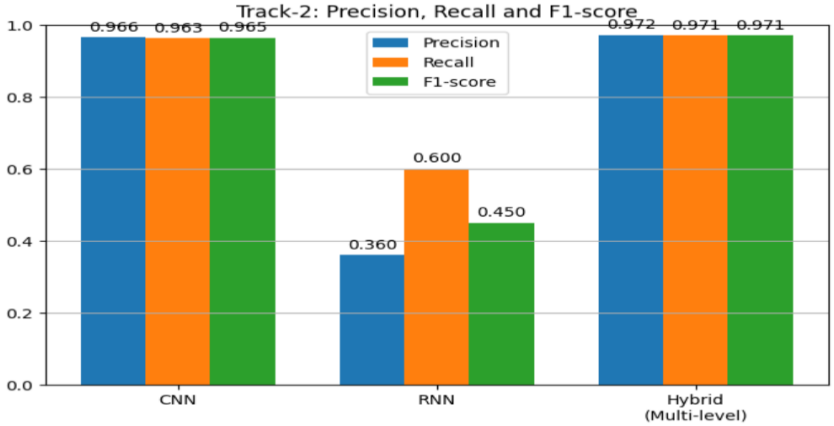


Figure 22: Precision, Recall, and F1-Score Comparison of CNN, RNN, and Hybrid (Multi-level) Models for Track-2

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 1928-1941

5.Conclusion

This research has managed to establish a watermarking model of medical images with a resistant attack mechanism, which is effective in overcoming the most important security, authenticity and integrity-related challenges in the current healthcare systems. Combining the adaptive Convolutional Neural Network-based watermark embedding with the fuzzy-based inference version of decoding, the solution methodology addressed the major drawbacks of traditional watermarking, especially their susceptibility to attacks and the lack of a mechanism to handle uncertainty when extracting information. The CNN also intelligently detects perceptually safe embedding areas to guarantee that the quality of diagnostic remains constant, and the fuzzy inference system compensates to make it more robust by reliably reconstructing the watermark information despite extreme signal processing and geometric distortions. The effectiveness of the framework is supported by extensive experiments carried out on a large and multi-modal CT and X-ray data set. The watermarking system is fully robust with the Normalized Correlation value of 1.0 and Bit Error Rate value of 0.0 in all the attacks that it has been tested in, which is an affirmation of high resilience. Also, the hybrid CNN-RNN classification model provides better results with 97.1% accuracy which demonstrates the effectiveness of learning spatial and sequential features.

References

[1] Bouarroudj, Riadh, Feryel Souami, Fatma Zohra Bellala, Nabil Zerrouki, Fouzi Harrou, and Ying Sun. "Secure and reversible fragile watermarking for accurate authentication and tamper localization in medical images." Computers and Electrical Engineering 123 (2025): 110072.

[2] Li, Daming, Lianbing Deng, Brij Bhooshan Gupta, Haoxiang Wang, and Chang Choi. "A novel CNN based security guaranteed image watermarking generation scenario for smart city applications." Information Sciences 479 (2019): 432-447.

[3] Kandi, Haribabu, Deepak Mishra, and Subrahmanyam RK Sai Gorthi. "Exploring the learning capabilities of convolutional neural networks for robust image watermarking." Computers & Security 65 (2017): 247-268.

[4] Wang, Xiaochao, Ding Ma, Kun Hu, Jianping Hu, and Ling Du. "Mapping based residual convolution neural network for non-embedding and blind image watermarking." Journal of Information Security and Applications 59 (2021): 102820.

[5] Hatoum, Makram W., Jean-François Couchot, Raphaël Couturier, and Rony Darazi. "Using deep learning for image watermarking attack." Signal Processing: Image Communication 90 (2021): 116019.

[6] Balasamy, K., and S. Suganyadevi. "Multi-dimensional fuzzy based diabetic retinopathy detection in retinal images through deep CNN method." Multimedia Tools and Applications 84, no. 18 (2025): 19625-19645.

[7] Balasamy, K., N. Krishnaraj, and K. Vijayalakshmi. "An adaptive neuro-fuzzy based region selection and authenticating medical image through watermarking for secure communication." Wireless Personal Communications 122, no. 3 (2022): 2817-2837.

[8] Shamia, D., K. Balasamy, and S. Suganyadevi. "A secure framework for medical image by integrating watermarking and encryption through fuzzy based ROI selection." Journal of Intelligent & Fuzzy Systems 44, no. 5 (2023): 7449-7457.

[9] Sun, Congcong, Hui Tian, Peng Tian, Haizhou Li, and Zhenxing Qian. "Multi-agent deep learning for the detection of multiple speech steganography methods." IEEE/ACM Transactions on Audio, Speech, and Language Processing 32 (2024): 2957-2972.

[10] Amin, Mahmoud, Fayez FM El-Sousy, Ghada A. Abdel Aziz, Khaled Gaber, and Osama A. Mohammed. "CPS attacks mitigation approaches on power electronic systems with security challenges for smart grid applications: A review." Ieee Access 9 (2021): 38571-38601.

[11] Swaraja, K., and K. Meenakshi. "An optimized blind dual medical image watermarking framework for tamper localization and content authentication in secured telemedicine." Biomedical Signal Processing and Control 55 (2021): 101665.

[12] Sinhal, Rishi, Sachin Sharma, Irshad Ahmad Ansari, and Varun Bajaj. "Multipurpose medical image watermarking for effective security solutions." Multimedia Tools and Applications 81, no. 10 (2022): 14045-14063.

[13] Khare, Priyank, and Vinay Kumar Srivastava. "A secure and robust medical image watermarking approach for protecting integrity of medical images." Transactions on Emerging Telecommunications Technologies 32, no. 2 (2022): e3918.

[14] Pavithra, V., and Jeyamala Chandrasekaran. "Developing security solutions for telemedicine applications: medical image encryption and watermarking." In Research anthology on telemedicine efficacy, adoption, and impact on healthcare delivery, pp. 612-631. IGI Global, 2022.

[15] Anand, Ashima, and Amit Kumar Singh. "Watermarking techniques for medical data authentication: a survey." Multimedia Tools and Applications 80, no. 20 (2022): 30165-30197.

[16] Yan, Fei, Hesheng Huang, and Xu Yu. "A multiwatermarking scheme for verifying medical image integrity and authenticity on the internet of medical things." IEEE Transactions on Industrial Informatics 18, no. 12 (2023): 8885-8894.

[17] Bouarroudj, Riadh, Feryel Souami, Fatma Zohra Bellala, Nabil Zerrouki, Fouzi Harrou, and Ying Sun. "Secure and reversible fragile watermarking for accurate authentication and tamper localization in medical images." Computers and Electrical Engineering 123 (2023): 110072.

[18] Chen, Yueh-Peng, Tzuo-Yau Fan, and Her-Chang Chao. "Wmnet: a lossless watermarking technique using deep learning for medical image authentication." Electronics 10, no. 8 (2023): 932.

[19] Sood, Rajat, Jyoti Rani, Ashima Anand, and Jatin Bedi. "Deep learning-based dual watermarking solution for securing medical images in e-healthcare." Knowledge-Based Systems (2024): 114750.

[20] Zhang, Xiaofei. "Detection and Resilience Mechanisms Against Cyber-attacks in Connected Automated Vehicle Platoons." PhD diss., University of Wollongong, 2024.

[21] Hu, Runyi, Jie Zhang, Shiqian Zhao, Nils Lukas, Jiwei Li, Qing Guo, Han Qiu, and Tianwei Zhang. "Mask image watermarking." arXiv preprint arXiv:2504.12739 (2024).

[22] Saidi, Hadjer, Okba Tibermacine, and Ahmed Elhadad. "High-capacity data hiding for medical images based on the mask-RCNN model." Scientific Reports 14, no. 1 (2024): 7166.

[23] Liu, Xinyun, Ronghua Xu, and Yu Chen. "Securing Digital Media Integrity: A Survey of Watermarking and Manipulation Detection for Image Authentication." Authorea Preprints (2025).

[24] https://www.kaggle.com/datasets/shuvokumarbasak2030/medical-imaging-ct-xray-colorization-new-dataset

[25] Bellido, Marlon Huamani, Luiz Pinguelli Rosa, Amaro Olímpio Pereira, Djalma Mosqueira Falcao, and Suzana Kahn Ribeiro. "Barriers, challenges and opportunities for microgrid implementation: The case of Federal University of Rio de Janeiro." Journal of cleaner production 188 (2018): 203-216.

[26] Akinyele, Daniel, Juri Belikov, and Yoash Levron. "Challenges of microgrids in remote communities: A STEEP model application." Energies 11, no. 2 (2018): 432.

[27] Faisal, Mohammad, Mahammad A. Hannan, Pin Jern Ker, Aini Hussain, Muhamad Bin Mansor, and Frede Blaabjerg. "Review of energy storage system technologies in microgrid applications: Issues and challenges." Ieee Access 6 (2018): 35143-35164.

[28] Wang, Shiqiang, Jianchun Xing, Ziyan Jiang, and Juelong Li. "Decentralized economic dispatch of an isolated distributed generator network." International Journal of Electrical Power & Energy Systems 105 (2019): 297-304.

[29] Chakraborty, Pratyush, Enrique Baeyens, Kameshwar Poolla, Pramod P. Khargonekar, and Pravin Varaiya. "Sharing storage in a smart grid: A coalitional game approach." IEEE Transactions on Smart Grid 10, no. 4 (2018): 4379-4390.

[30] Kristiansen, Martin, Magnus Korpås, and Harald G. Svendsen. "A generic framework for power system flexibility analysis using cooperative game theory." Applied energy 212 (2018): 223-232.