

ORIGINAL RESEARCH

Open access

Privacy-Aware Analysis of Encrypted Medical Image Sequences Using Convolutional Bi-LSTM Networks

Hassan Ali^{1*}, Mariam Farooq¹, Usman Shah², Bilal Khan³, Sana Malik⁴

Abstract

The integration of deep learning (DL) algorithms offers significant potential to enhance healthcare applications. Specifically, DL has advanced medical imaging diagnostics, therapeutic interventions, and disease management. However, deploying deep learning algorithms on sensitive medical imagery introduces substantial challenges regarding data security and patient privacy. Optimizing medical imaging performance while simultaneously ensuring patient anonymity remains a complex task. Consequently, privacy-preserving techniques designed for deep learning model training and inference are increasingly gaining traction. Currently, these medical image sequences are interpreted using state-of-the-art computer-aided detection/diagnosis (CAD) methodologies. Yet, frameworks that require uploading medical photographs to external servers inherently create privacy vulnerabilities.

To address this, this article introduces a convolutional Bi-LSTM network designed to evaluate fully homomorphic-encrypted (HE) time-series medical images. Within this framework, convolutional blocks extract selective spatial features from the secure image sequences, while analytical sequence layers based on Bi-LSTM encode the temporal data. To optimize efficiency and minimize misdiagnoses, a weighted unit and sequence voting layer incorporates spatial data using varying weights. The effectiveness of this framework is demonstrated using two rigorous public benchmarks: the CheXpert and BreakHis datasets. The proposed technique surpasses several competing methods, achieving an accuracy exceeding 0.99 across both datasets. These findings indicate that the introduced methodology can successfully extract visual representations and sequential dynamics from encrypted medical image series, thereby safeguarding patient privacy while maintaining high performance in medical image analysis.

Keywords Deep learning, Healthcare, Medical image, Security, Zero watermarking

*Correspondence:

Hassan Ali

hassan.ali@uaf.edu.pk

¹ Department of Privacy-Preserving Medical Image Analysis, Faculty of Engineering, University of Agriculture Faisalabad, Faisalabad, Pakistan

² Department of Encrypted Image Sequences and Bi-LSTM, Faculty of Engineering, National University of Sciences and Technology, Islamabad, Pakistan

³ Department of Convolutional Networks for Medical Imaging, Faculty of Engineering, University of Punjab, Lahore, Pakistan

⁴ Department of Time-Series Medical Image Privacy, Faculty of Engineering, University of Karachi, Karachi, Pakistan

Introduction

The widespread adoption of digitization within the medical sector has been heavily driven by the standardization of hospital systems [1]. Modern medical machinery generates digital medical images on a daily basis [2, 3]. Concurrently, rapid developments in information technology have accelerated the maturity of intelligent medicine and remote diagnostics [4-6]. Consequently, transmitting vast quantities of medical imagery over the internet has become a routine practice [7]. Modalities such as X-rays, CT scans, MRI scans, and ultrasounds deliver critical insights regarding patient

health. However, these files often contain sensitive personal information, including patient identifiers, which face the risk of unauthorized access if exposed. Developing methodologies that safeguard patient privacy without reducing the quality or clinical utility of medical images is therefore essential.

Deep learning models excel in this domain because they can automatically process massive datasets to uncover intricate patterns and features, making them highly suitable for privacy-preservation tasks in medical imaging. Researchers and developers are actively investigating diverse approaches to utilize deep learning techniques to guarantee the confidentiality and privacy of medical imagery. Medical records transmitted digitally are vulnerable to theft, unauthorized exploitation, and alteration [8]. Under these conditions, a patient's medical image can easily expose confidential information.

While the evolution of healthcare IT infrastructure has enhanced remote diagnosis and medical image sharing [9], the expanding use of these systems intensifies the need to protect sensitive patient information, such as MRI scans, diverse medical imagery, and electronic health records [10, 11]. Hence, establishing robust mechanisms to safeguard sensitive patient data is imperative.

During clinical assessments, time-series medical photographs capture the dynamic progression of lesions. However, uploading these sequences to cloud servers can compromise patient privacy, especially given rising concerns over the sharing of medical and healthcare data [12, 13]. Notably, standard approaches like image scrambling encryption [14], Advanced Encryption Standard (AES) cryptosystems [15], and Rivest–Shamir–Adleman (RSA) encryption [16] only secure data during transit; the cloud server must decrypt the data before executing artificial intelligence algorithms. Because the cloud server retains access to the unencrypted data, these techniques fail to resolve the underlying privacy dilemma.

Recent studies have explored the use of neural networks to evaluate encrypted images directly. Homomorphic encryption-based privacy-preserving deep learning models have gained popularity due to their capacity to compute directly on encrypted imagery while maintaining high performance. Nevertheless, most existing algorithms only process isolated encrypted images, which hinders their ability to encode discriminative, time-dependent data.

Time-series medical imagery is vital for analyzing lesion dynamics. When developing these models, the unique constraints of medical scenarios and the risks of missed diagnoses must be prioritized. From a clinical perspective, lowering the rate of incorrect diagnoses is more critical than merely maximizing overall accuracy. Missed evaluations can cause patients to miss optimal treatment windows, complicating subsequent therapeutic interventions and ultimately reducing 5-year survival rates.

Deep learning models are frequently deployed to anonymize or de-identify medical imagery. This is achieved by removing or masking identifiable details—such as patient names, birth dates, and other personal characteristics—while preserving the diagnostic integrity of the images. By utilizing deep learning algorithms, sensitive regions can be automatically detected and blurred or excised, rendering the images safe for research, distribution, or analysis without violating patient privacy.

Furthermore, deep learning models can generate synthetic medical images that replicate real patient data while ensuring absolute privacy. These models are trained on existing medical image datasets to learn the underlying structural patterns and characteristics. The resulting synthetic images can then be utilized for tasks such as algorithm development without exposing actual patient data.

The primary contributions of this work are outlined below:

1. This paper introduces a convolutional Bi-LSTM network designed to evaluate homomorphic-encrypted time-series medical images. Discriminative spatial features are extracted from the encrypted frames using specialized

convolutional blocks.

2. The proposed methodology integrates varying spatial weights through a dedicated weighted unit and sequence voting layer.
3. This study validates the recommended technique by comparing it against a robust zero-watermarking system that addresses security challenges in medical image storage and transmission, with a particular focus on lesion zone protection. This comparison demonstrates that the proposed framework successfully ensures privacy while enhancing the performance of medical image analysis.

Related Works

Wang *et al.* [17] applied conventional machine learning (ML) to diagnose breast cancer within digital mammograms, utilizing a dataset gathered at the Tumor Hospital of Liaoning Province. Their study involved two distinct ML approaches: a single-layer extreme learning machine (ELM) and a traditional support vector machine (SVM). Although a deep neural network (DNN) architecture was not deployed in their research, it established a foundational pathway for utilizing deep learning models to conduct automated breast cancer screenings in subsequent studies. To improve breast cancer detection accuracy, Shen *et al.* [18] applied deep convolutional neural networks (DCNNs) to mammography data. They trained models using ResNet-50 and VGG-16 architectures, conducting evaluations on the CBIS-DDSM [19] dataset, which comprises 2478 mammographic images. Incorporating the ResNeSt [20] framework, Zhang *et al.* [21] introduced ResNetSA using a novel brain MR dataset provided by the Ruijin Hospital, affiliated with the Shanghai Jiao Tong University School of Medicine. This specialized attention-driven deep convolutional neural network successfully identified malignancies, a capability facilitated by the spatial-attention component of the Convolutional Block Attention Module (CBAM).

As a rapidly evolving auxiliary diagnostic tool, computer-aided detection/diagnosis (CAD) algorithms hold substantial potential for the broad analysis of time-series medical imagery. For instance, the authors in [22] integrated a CNN with an LSTM network to improve surgical workflow recognition by leveraging discriminative visual representations alongside temporal components. The effectiveness of LSTM has also been demonstrated in the classification of mammography images [23]. Furthermore, Reference [24] combined convolutional, deconvolutional, and LSTM layers to categorize breast cancer images. Existing literature indicates that recurrent neural networks, such as LSTM and the Gated Recurrent Unit (GRU), are capable of instinctively identifying conditions like prostate cancer and myocardial infarction [25, 26]. Building upon these findings, the current study utilizes deep learning-based CAD frameworks to interpret time-series medical photographs.

Homomorphic encryption permits mathematical operations to be executed directly on ciphertexts without prior decryption, thereby preventing the exposure of plaintext data [27]. As noted in Reference [28], fully homomorphic encryption (FHE) marked the first instance where unrestricted computations could be performed natively on encrypted data. Throughout the past decade, numerous FHE variations have emerged to optimize processing efficiency while maintaining data confidentiality. The Brakerski/Fan-Vercauteren (BFV) scheme [29] stands out as an exceptionally efficient fully homomorphic encryption model, supporting arbitrary multiplication and addition operations on encrypted messages [30]. Because of its elegant and straightforward structure, the BFV approach delivers strong performance in cloud computing and secure technology environments [31, 32].

Natsheh *et al.* [33] introduced an efficient methodology for encrypting and decrypting DICOM medical images via the Advanced Encryption Standard (AES). The key sequences, generated through chaotic maps, exhibit distinct cryptographic strengths due to their pseudo-randomness, ergodicity, and sensitivity to initial values. Similarly, a medical image encryption strategy based on selective chaos was outlined by Kanso *et al.* [34], where each iteration incorporates block-based shifting and masking sequences. Here, chaotic cat maps are utilized to shuffle and mask the input image. Song *et al.* [35] demonstrated another secure medical image encryption model based on chaos theory, which safeguards imagery during the permutation phase using a bit-level shuffling algorithm coupled with a replacement mechanism.

Furthermore, Ding et al. [36] introduced DeepEDN, a deep neural network framework engineered for medical image encryption and decryption. To protect medical visuals, this system first deploys a Cycle-Generative Adversarial Network (Cycle-GAN) as its primary learning engine to convert data from the plaintext domain into the target domain. The decryption is subsequently managed by an updated network structure. Rather than decrypting the entire image, a region of interest (ROI)-mining network is utilized to extract only the pertinent components.

Since the introduction of the concept by Goodfellow et al. [37] in 2014, numerous researchers have directed their attention toward GAN-based frameworks across various domains. A standard GAN network consists of an adversarial discriminator and a generator [38]; the generator models the underlying data distribution, whereas the discriminator learns to identify anomalies within the data. GAN-based algorithms have consistently achieved state-of-the-art results across several disciplines, including image generation [36], image segmentation [37], image super-resolution [38], and image-to-image translation. To convert images from one domain to another, Yi et al. [39] utilized a conditional generative adversarial network (CGAN). Their research demonstrated that this approach outperforms existing methodologies in tasks such as image synthesis from label maps, object reconstruction from edge maps, and image colorization.

An epistemological framework [40] establishes the essential principles and perspectives that govern how knowledge is conceptualized, acquired, verified, and articulated within a specific academic field or area of inquiry. It fundamentally defines the philosophy of knowledge in that domain, shaping the methodologies and techniques utilized to produce new insights. In [41], an investigation offered valuable insights into both the theoretical underpinnings and practical consequences of ethical factors and collective accountability at the intersection of healthcare and technological integration.

Using the DualGAN [42] methodology, a learning network can be trained using two sets of unlabeled images. DualGAN accepts unlabeled images from two domains as input to facilitate various image-to-image translation tasks, simultaneously optimizing two reliable image transformation networks. To execute image translation using unpaired data, Cycle-GAN was introduced in [43]. The Cycle-GAN architecture trains two distinct GAN models concurrently: one model maps from class A to class B, while the other maps the inverse relationship. By linking these two mappings, the system reformulates the loss function. Adversarial loss remains crucial to the success of GANs, ensuring that generated images are indistinguishable from target images. To achieve successful "Image-to-Image transformation," this negative loss is used to learn the direct mapping from source domain images to target domain images.

Materials and Methods

This section discusses the properties of deep neural networks that prevent the leakage of private data. The introduced framework is anchored upon the MORE homomorphic encryption scheme, which permits conventional neural network models to be trained and executed directly on homomorphically protected data [44, 45].

Problem formulation

The problem of preserving privacy in medical imagery using deep learning can be formulated mathematically as follows: Consider a collection of sensitive medical images $I = \{I_1, I_2, \dots, I_n\}$ linked with corresponding patient identifiers $P = \{P_1, P_2, \dots, P_n\}$, where I_i denotes an individual image and P_i denotes the patient identifier mapped to image I_i . The objective is to construct a deep learning framework F capable of safeguarding the privacy of these medical images while preserving their full clinical and diagnostic utility. The framework F should incorporate a collection of privacy-preserving methods applied directly to the medical imagery to secure confidential patient details.

Let the privacy-preserving function be represented as $PP(I, P)$, which accepts the medical image collection I and the corresponding patient identifiers P as input variables, yielding a transformed image collection $I' = \{I'_1, I'_2, \dots, I'_n\}$ with integrated privacy protections. The modified images I' must fulfill specific criteria: the patient identifiers $P' = \{P'_1, P'_2, \dots, P'_n\}$ mapped to the transformed images I' must not expose the identities of the individuals in the original dataset.

Consequently, there must be no explicit connection between the modified images and their original patient identifiers. Furthermore, the transformed images I' must retain a high level of diagnostic information to allow for valid analysis and clinical diagnosis. The privacy-preserving transformations executed on the imagery must not degrade the overall quality or utility of the medical data.

To achieve deep learning-driven privacy preservation for medical images, the framework F must utilize the capabilities of deep learning algorithms to formulate techniques that transform the dataset I while satisfying the demands for both anonymity and utility conservation. The target is to establish an optimal privacy-preserving function $PP * (I, P)$ that maximizes data privacy while preserving the diagnostic worth of the altered images, subject to any domain-specific constraints or application requirements. Mathematically, the optimization problem can be formulated as:

$$PP * (I, P) = \arg \max PP(I, P), \quad (1)$$

subject to constraints and requirements specific to privacy preservation, such as anonymity and utility preservation. Resolving this problem requires engineering and training deep learning architectures, creating suitable privacy-preserving methodologies, and assessing the performance of the framework F regarding both privacy protection and diagnostic accuracy through appropriate evaluation metrics.

Dataset

The empirical investigations in this study utilize the CheXpert dataset [46] (illustrated in **Figure 1**), which constitutes a massive repository comprising 224,316 chest radiographs obtained from a cohort of 65,240 patients. This dataset encompasses five distinct clinical categories corresponding to various thoracic pathologies, namely: (a) atelectasis, (b) cardiomegaly, (c) consolidation, (d) edema, and (e) pleural effusion. Crucially, the process of re-initializing the fully connected layer alongside the modifications applied to the remaining convolutional layers introduces no vulnerabilities regarding data privacy leaks [1]. For the experimental pipeline, a subset of 10,000 radiographs is designated for model training, whereas 234 images are reserved for evaluation purposes.

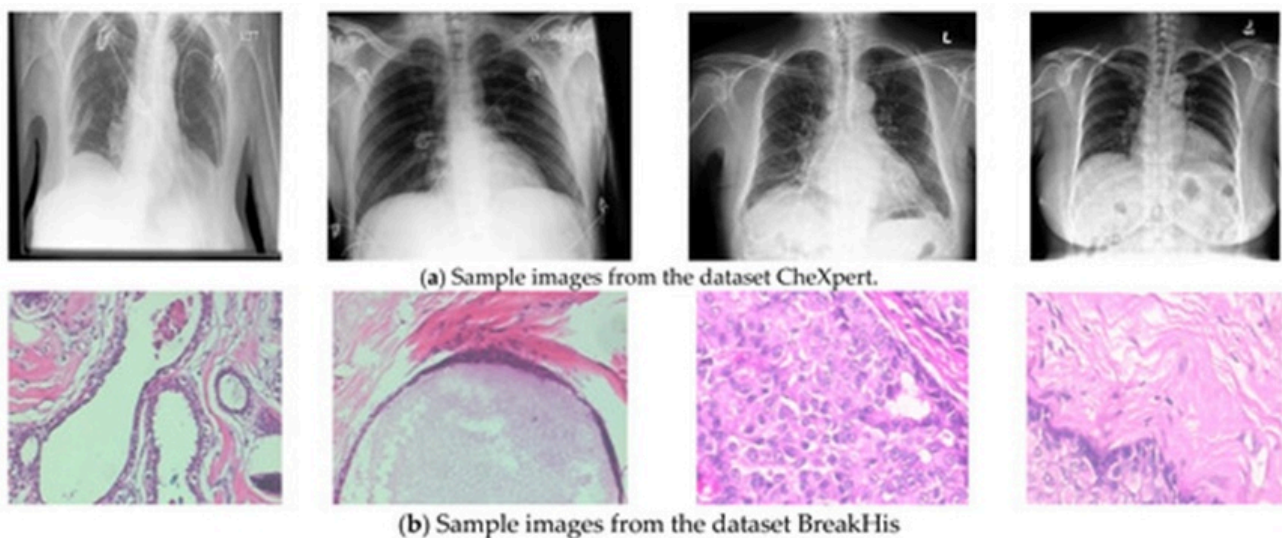


Figure 1. Sample images from the dataset used in this study.

Concurrently, the Breast Cancer Histopathological Image Classification (BreakHis) repository contains a collection of 9109 digital micrographs displaying breast tumor tissue samples. These samples were acquired from 82 distinct patients

across four discrete optical magnification levels: 40×, 100×, 200×, and 400×. The dataset is partitioned into 5429 malignant and 2480 benign image instances, each characterized by a resolution of 700 × 460 pixels, a 3-channel RGB color space, an 8-bit depth, and saved in PNG format. The assembly of this database took place at the P&D Laboratory of Pathological Anatomy and Cytopathology located in Parana, Brazil. Within the BreaKHis framework, lesions are classified into two primary pathological groups: benign and malignant. A tumor is pathologically classified as benign when it exhibits an absence of malignant indicators, including cellular atypia, mitotic figures, basement membrane degradation, or metastatic potential. Generally, benign growths progress slowly and remain localized to their site of origin. Conversely, malignant neoplasms—frequently termed cancer—are distinguished by their capacity to invade and destroy adjacent anatomical structures (local invasion) and disseminate to distant regions of the body (metastasis).

Methodology

Over the past several years, the deployment of deep learning paradigms for medical data analytics has yielded highly successful outcomes. Although deep learning architectures present a surface-level impression of extreme complexity, their core mechanisms can be broken down into recurrent computational operations relying on basic arithmetic functions over rational numbers. Indeed, a significant portion of state-of-the-art benchmarks achieved in deep learning rely on neural architectures that utilize a highly restricted set of fundamental operations. By exploiting the inherent homomorphic attributes of the MORE framework, it becomes feasible to expand the functional scope of these neural network models to execute computation directly on ciphertext.

The conceptualized pipeline integrating homomorphic encryption (HE) and deep learning models is illustrated schematically in **Figure 2**. Prior to ingestion and processing, a cryptographic private key is applied to secure the training observations. Consequently, the plaintext information is decoupled from the processing architecture, remaining strictly isolated within the environment of the data provider, whereas the deep learning network operates exclusively on the encrypted ciphertext. Because every internal network function is tailored to support ciphertext inputs, and given that the MORE encryption methodology natively supports homomorphic floating-point arithmetic, the system can undergo direct optimization on encrypted data using standard training protocols.

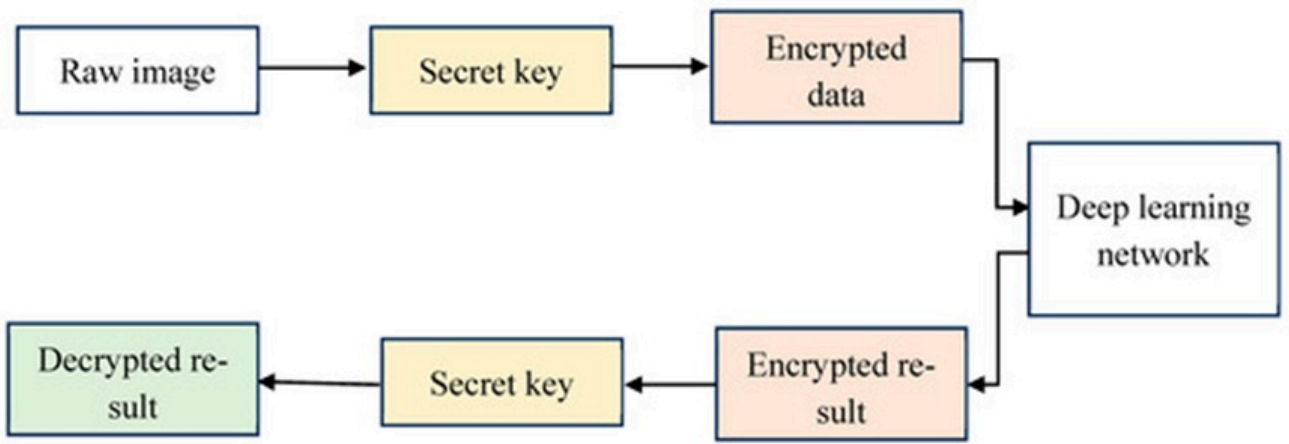


Figure 2. Workflow of the recommended deep learning-based application that protects the privacy and uses homomorphic encryption.

The resulting inferences generated by the model are similarly encrypted and remain unreadable to any entity lacking the corresponding secret key. Once the optimization phase concludes, the encrypted parameters of the model can evaluate novel, encrypted data instances using the identical cryptographic key deployed during the training sequence. The MORE framework is built upon symmetric cryptography. Consequently, this methodology produces a singular secret token

responsible for both the encryption of plaintext records and the decryption of ciphertext results, as formalized in **Algorithm 1**.

Secret Key Generation	
Input:	None
Output:	Secret key <i>SK</i>
Steps:	1. Random Matrix Generation: $R \in R^{(n \times n)}$ 2. Inverse Matrix: $R_{inv} = R^{-1}$ 3. Secret Key: $SK = R_{inv}$ 4. Generate a random matrix <i>R</i> of size $(n \times n)$ with elements from a suitable key space. 5. Compute the inverse matrix R_{inv} of <i>R</i>. 6. Set $SK = R_{inv}$. 7. Output <i>SK</i> as the secret key.
MORE Encryption:	
Input:	Plain text matrix <i>P</i> , Secret key <i>SK</i>
Output:	Encrypted matrix <i>C</i>
Steps:	1. Plain Text Matrix: $P \in R^{(m \times n)}$ 2. Encrypted Matrix: $C = P * SK$ 3. Compute the matrix multiplication $C = P * SK$. 4. Output <i>C</i> as the encrypted matrix.
MORE Decryption:	
Input:	Encrypted matrix <i>C</i> , Secret key <i>SK</i>
Output:	Decrypted matrix <i>P</i>
Steps:	1. Encrypted Matrix: $C \in R^{(m \times n)}$ 2. Decrypted Matrix: $P = C * SK$ 3. Compute the matrix multiplication $P = C * SK$. 4. Output <i>P</i> as the decrypted matrix.

Algorithm 1. of MORE (Matrix Operation for Randomization or Encryption)

Convolutional Bi-LSTM

Within the domains of visual pattern analysis and object localization, Convolutional Neural Networks (CNNs) serve as a foundational methodology. The primary advantage of a CNN lies in its capacity to autonomously learn hierarchical feature representations from raw input graphics, effectively bypassing manual feature engineering workflows which are inherently laborious and error-prone. Standard CNN configurations are built upon a triad of structural layers: convolution, pooling, and fully connected layers. Through the integration of these components, convolutional blocks containing alternating CLs and PLs are formed to capture meaningful spatial representations from an incoming image. A complete CNN framework is subsequently assembled by chaining multiple convolutional blocks sequentially. When deploying a CNN for predictive

regression or categorical classification, FCLs are typically appended to constitute the terminal layer. The architectural paradigm combining a standard CNN with a Bi-LSTM network is presented in **Figure 3**.

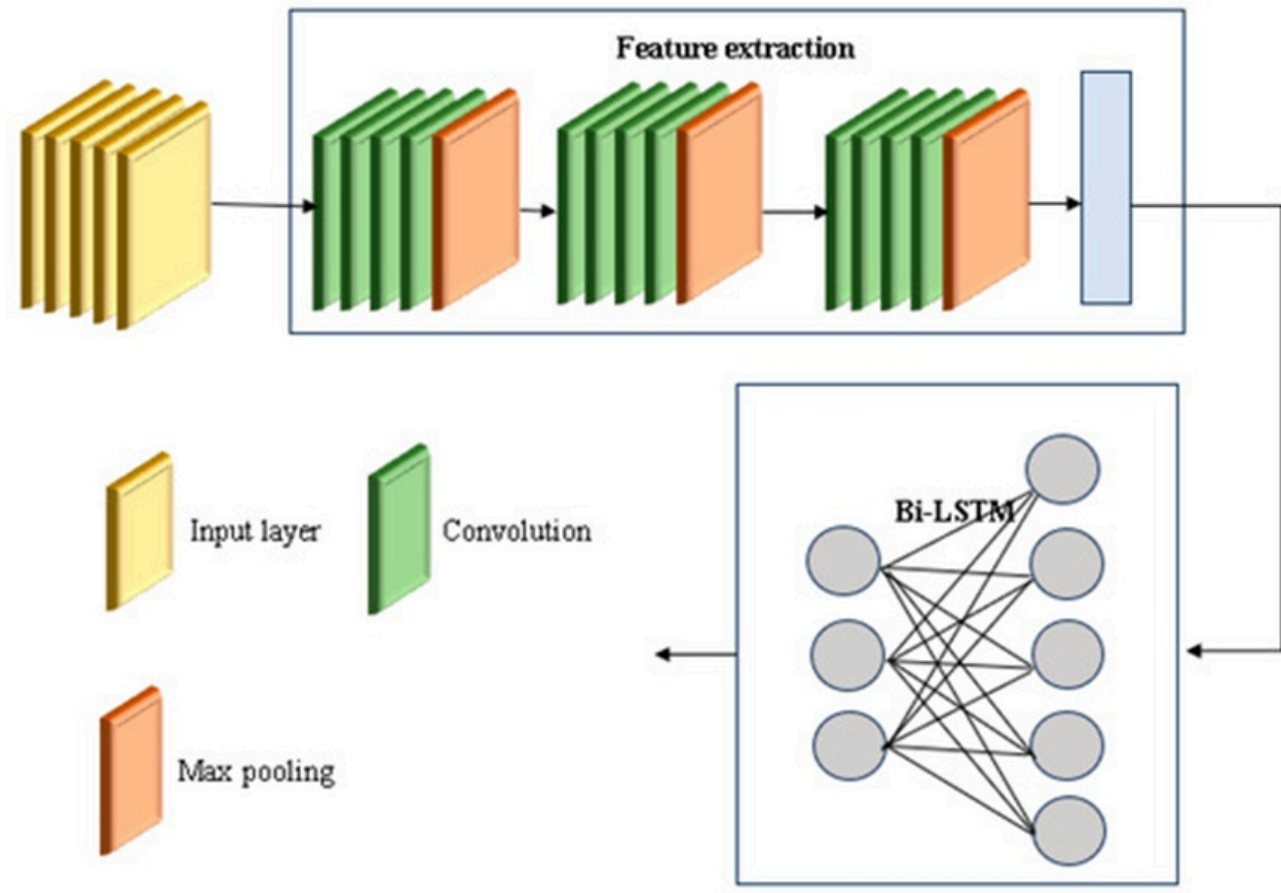


Figure 3. CNN + Bi-LSTM structure.

The convolutional layer serves a critical function within the broader CNN arrangement. This layer implements a series of analytical filters, known as the 'convolution kernel,' across the spatial dimensions of the input graphic or the preceding feature maps. As noted previously, the convolution kernel operates as an intrinsic feature extractor, identifying structural patterns embedded within the input image or intermediate feature spaces. From a mathematical perspective, convolution describes a transformation that maps an input image matrix to a corresponding output kernel space.

$$y_k = X \otimes K_k + b_k. \quad (2)$$

In the following formula, 'X' represents the input image, 'K_k' represents the *k*th convolution kernel in CL, 'b_k' represents the bias term, 'y_k' represents the *k*th output feature map, and '⊗' represents the convolution operation. A non-linear activation function was then applied to the final feature map after the convolution procedure to introduce the non-linearity. The aforementioned process may be stated mathematically as:

$$S_{x,y} = a \left(\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{p=0}^{N-1} w_{p,n,m} X_{x+n,x+p,m} + b \right). \quad (3)$$

Non-linear activation function (·); output feature map node at (x, y) designated by S_{x,y}; input pixel value x+n,x+p,m designating weight and bias of convolution kernel; convolution kernel size, k×k. Note that the picture at (x, y) at *p*th depth

can be significantly affected by the size of the kernel; w_p , and b represent the network's performance. Large kernel functions can generate duplicate processing and an increase in the computational complexity of a network, while tiny kernel functions can result in considerable information loss.

Subsequent to the CL operations, a PL is introduced to downsample the generated feature maps, reducing their dimensionality while preserving crucial localized spatial contexts and uniform attributes. The mathematical formulation governing this pooling mechanic is given by:

$$P_{x,y,z} = L_{(m,n) \in r_{x,y}}(X_{m,n,x}). \quad (4)$$

$L(\cdot)$ represents the pooling operation; $P_{x,y,z}$ represents the updated value for the node located at coordinates (x, y) in the z -th feature map; $r_{x,y}$ represents the pooling region encompassing coordinates (x, y) ; and $X_{m,n,x}$ represents the node at coordinates (x, y) inside the pooling region. There are several kinds of pooling operations. Maxpooling is the best option available. The maxpooling procedure takes a set of convolved features and chooses the one with the highest value inside the pooling window as the output feature.

FCLs are utilized to handle both continuous regression values and categorical classifications. Within the FCL, the multi-dimensional outputs derived from the preceding CL/PL sequence are flattened into a unified 1D feature vector. For classification environments, the final FCL outputs are routed through a softmax activation function. Group labels are then projected based on the FCL outputs, with the softmax function computing the corresponding probability distribution. The mathematical formulation of the softmax activation function is established as follows:

$$F = \sigma(h_n \circ w^T + b). \quad (5)$$

Estimated class is represented by F , the total number of hidden neuron values is represented by (' hn '), the element wise multiplication operator is (' $\circ$ '), the weight matrix is (' w^T ') between FCL and output layer and bias (' b ').

The Bi-LSTM model represents a specialized extension of the Long Short-Term Memory (LSTM) architecture found within recurrent neural networks (RNNs). By integrating a dual-directional processing paradigm into the classical LSTM framework, the Bi-LSTM model enhances the system's capacity to synthesize historical dependencies alongside future contextual cues when generating sequential estimations. Mathematically, the Bi-LSTM architecture extracts both forward and backward latent states, denoted as hif and hib , from the sequential input sequence I_t at any given temporal index t . The calculation of these respective latent variables is handled by independent forward and reverse LSTM modules.

$$hif = LF(I_t, hif_{t-1}) \quad (6)$$

$$hib = LB(I_t, hib_{t+1}). \quad (7)$$

Here, I_t represents the input at time step t , hif_{t-1} is the previous hidden state for the forward LSTM unit, and hib_{t+1} is the prior hidden state for the backward LSTM unit. Information about the past is stored in the forward hidden states (hif), and data about the future are stored in the backward hidden states (hib). In BiLSTM, the forward normal LSTM uses the same threshold calculation equations as a traditional LSTM, while the reverse normal LSTM uses the threshold design formulas described below.

These hidden states are concatenated to obtain the final hidden state h_t :

$$in_t = \sigma(\omega_{int} \cdot [hi_{t-1}, I_t] + b_{int}) \quad (8)$$

$$fr_t = \sigma(\omega_{frit} \cdot [hi_{t-1}, I_t] + b_{frit}) \quad (9)$$

$$op_t = \sigma(\omega_{opt} \cdot [hi_{t-1}, I_t] + b_{opt}) \quad (10)$$

$$mt_t = fr_t * mt_{t-1} + in_t * \tanh(\omega_{mtt} \cdot [hi_{t-1}, I_t] + b_{mtt}). \quad (11)$$

The resulting hidden state h_t produced by the Bi-LSTM architecture is subsequently transferred to downstream modules for prediction or additional computational operations. By processing information simultaneously across both temporal directions, the Bi-LSTM framework gains a significant advantage in evaluating the extended context surrounding an inference target. This attribute proves highly advantageous in domains such as traffic forecasting, where historical patterns and upcoming scheduled events exert a combined influence on real-time metrics. Consequently, the incorporation of bidirectional sequential processing allows Bi-LSTM models to achieve elevated accuracy levels across diverse sequence estimation tasks—particularly traffic flow modeling—by identifying long-range dependencies latent within the underlying data structures.

Optimizing the training phase is critical for ensuring the peak performance of any data-driven model. Inadequate optimization can lead to networks that fail to accurately capture the training dataset or struggle to generalize to unseen data. Overfitting and underfitting represent two prevalent learning-related challenges that substantially undermine a model's efficacy on novel datasets. Consequently, identifying the appropriate juncture to halt training is vital to circumventing these issues. A widely adopted strategy involves establishing early termination criteria based on validation dataset errors to prevent performance degradation. Specifically, training can be discontinued if the error on a validation set ceases to decrease over time, or if the divergence between training and validation errors expands. In both methodologies, error analysis dictates the suspension threshold. However, while these techniques are readily applied during plaintext training, they become increasingly impractical when processing ciphertext data. The underlying cryptosystem prevents the evaluation of encrypted error metrics within conditional logic statements.

To bypass this constraint, privacy-preserving models are typically trained for a predetermined, fixed duration. Because the primary objective of this study is to evaluate whether a deep neural network can operate effectively on ciphertext data without requiring subsequent training, a suitable termination point can be designated beforehand. For simplicity and practical utility, we conducted evaluations and reported outcomes over an extended number of epochs. Every experimental task was evaluated in both its unencrypted and encrypted variants. In the initial phase, the neural network was trained and evaluated using plaintext data; in the second phase, ciphertext data were utilized alongside fully encoded trainable parameters. Both the plaintext and ciphertext architectures maintained identical training methodologies, hyperparameters, and initialization protocols. Additionally, identical baseline weights were used to train the models across both data states. To evaluate the performance of the neural network algorithms on the concealed testing dataset under ciphertext conditions, all assessment metrics were derived from the decrypted outputs.

Experimental Setup

The proposed hybrid deep neural network was implemented using TensorFlow2 and the Keras module in Python. Training was executed on a hardware configuration featuring an Intel(R) Core(TM) i7 2.2 GHz CPU and an NVidia giga texel shader extreme (GTX) 1050 GPU. The recommended network architecture was constructed using the following input parameters: a learning rate of 0.0001, a minibatch size of 256, and a cross-entropy loss function. Following each training step, the loss function was optimized via the Adam optimization algorithm. Notably, an early stopping mechanism was employed to regulate the total epoch count during training. If the validation loss failed to decrease by a threshold of at least 0.001 over 10 consecutive epochs, the training process was terminated. The final evaluation utilized the specific epoch weights that yielded the lowest validation loss. Furthermore, a 4-fold cross-validation technique was implemented to validate the operational efficacy of the network.

Results and Discussion

This study assesses the performance of the proposed methodology using four evaluation metrics: correctness, exactness, specificity, and the F1 score. The mathematical representations of these metrics are outlined in the subsequent expressions. Within these equations, the variables ' T_r ', ' T_{nr} ', ' F_{pr} ', and ' F_{nr} ' correspond to true positives, true negatives, false positives, and false negatives, respectively.

$$Accuracy = \frac{T_{pr} + T_{nr}}{T_{pr} + T_{nr} + F_{pr} + F_{nr}} \quad (13)$$

$$Precision = \frac{T_{pr}}{T_{pr} + F_{pr}} \quad (14)$$

$$Specificity = \frac{T_{nr}}{T_{nr} + F_{nr}} \quad (15)$$

$$F1 - score = \frac{2 \times T_{pr}}{2 \times T_{pr} + F_{nr} + F_{pr}} \quad (16)$$

The training trajectories for the proposed hybrid network are illustrated in **Figure 4**. As suggested by **Figure 5**, the convergence of training and validation losses toward zero at the end of the curves indicates that the network has been sufficiently trained. The total training duration for the proposed hybrid network is approximately 95 minutes.

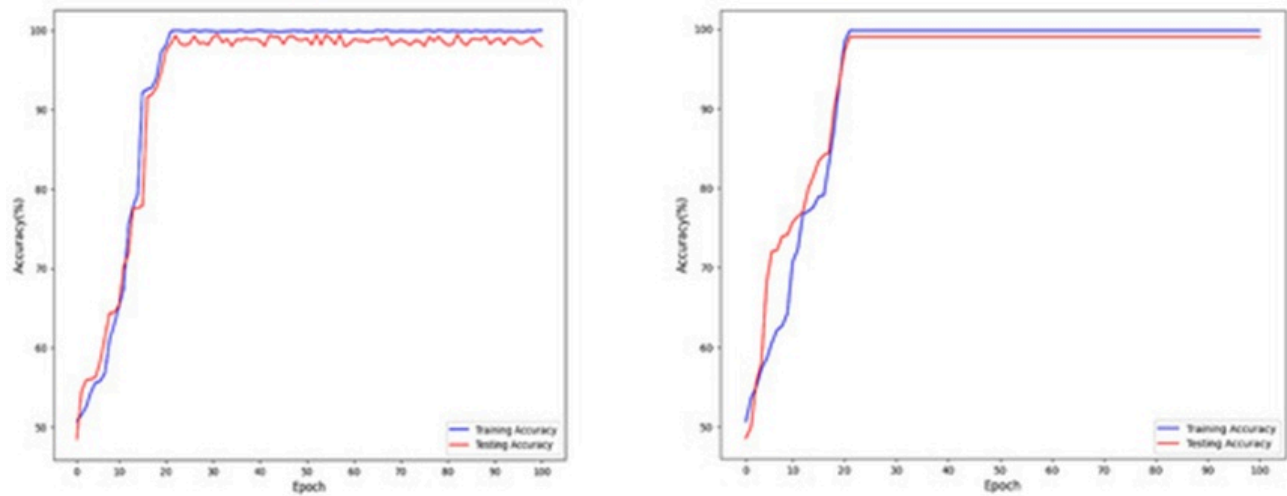


Figure 4. Accuracy of CNN-Bi-LSTM model on CheXpert and BreakHis datasets.

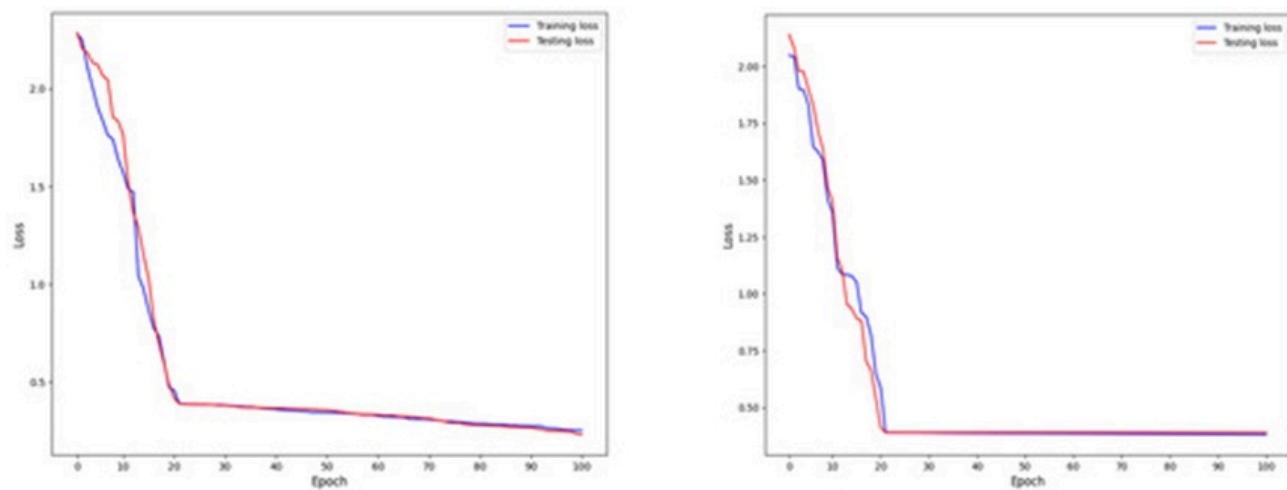


Figure 5. Loss of CNN-Bi-LSTM model on CheXpert and BreakHis datasets.

The final loss metrics recorded for the CNN-Bi-LSTM model on the CheXpert and BreakHis datasets are 0.39 and 0.29, respectively. In machine learning, loss serves as a standard metric to quantify the variance between a model's predicted outputs and the actual ground truth values, where lower values signify superior alignment. As shown in the loss curves over successive training epochs in [Figure 5](#), the error begins at an elevated level and systematically declines as the model learns from the underlying data. Minor fluctuations during the training phase are expected as the network continually modifies its parameters to optimize predictions. The overall downward trajectory of the loss confirms steady performance improvements on the CheXpert dataset.

Comparatively, the loss curve for the BreakHis dataset initiates at a lower baseline than that of CheXpert, indicating a stronger initial performance by the model on this dataset. Despite normal training fluctuations, the loss consistently diminishes or stabilizes at a low threshold, demonstrating robust performance on the BreakHis dataset. To evaluate efficacy, the proposed hybrid framework was compared against two standard deep learning architectures. In the first baseline, the flattened and Bi-LSTM layers of the proposed design were substituted with a conventional CNN structure.

The second baseline combined a standard CNN with a traditional LSTM, replacing the Bi-LSTM components of the proposed hybrid model with standard LSTM layers. Each of these comparative networks underwent full training.

Table 1 evaluates the proposed hybrid (CNN-Bi-LSTM) framework against the baseline deep architectures regarding total adjustable parameters, recognition performance, and computational runtime. The data indicate that the hybrid integration causes a marginal expansion in the overall number of trainable parameters. However, the hybrid networks clearly deliver superior performance compared to standard CNNs. Furthermore, the CNN-Bi-LSTM hybrid design achieves higher accuracy than the companion CNN-LSTM network.

Table 1. Result of various baseline model comparison based on performance metrics.

	BreakHis					CheXpert		
Model	F1-Score	Recall	Precision	Accuracy	F1-Score	Recall	Precision	Accuracy
CNN	0.951	0.940	0.936	0.935	0.930	0.928	0.932	0.924
LSTM	0.943	0.948	0.942	0.945	0.944	0.952	0.945	0.944
Bi-LSTM	0.945	0.952	0.957	0.956	0.968	0.951	0.962	0.954
CNN-LSTM	0.970	0.963	0.962	0.964	0.976	0.977	0.984	0.972
CNN-Bi-LSTM	0.998	0.997	0.998	0.999	1.00	0.991	0.998	0.999

Limitation

The proposed methodology depends on fully homomorphic encryption (HE) to perform privacy-preserving processing on medical image sequences. However, HE introduces substantial computational overhead and increased complexity during encryption and decryption procedures. Deep learning frameworks, particularly those utilizing convolutional and LSTM layers, are inherently resource-intensive. Executing these operations directly on encrypted data streams can dramatically amplify computational demands, thereby extending processing times. Furthermore, the adaptability of this framework may fluctuate when applied to heterogeneous data sources or diverse imaging modalities, as the model might struggle to generalize to unfamiliar variations. Consequently, its resilience across varying acquisition parameters, image resolutions, and diagnostic devices requires deeper validation. Lastly, while the method is designed to protect patient privacy, legal and ethical challenges persist regarding the management of sensitive medical records, even in an encrypted state. Full alignment with data protection statutes and patient consent mandates remains mandatory.

Conclusion

In conclusion, deep learning methodologies have demonstrated substantial promise in advancing healthcare systems, particularly in medical image analysis, therapeutic planning, and disease monitoring. However, integrating sensitive clinical imagery into deep learning architectures introduces critical challenges regarding data security and privacy. Balancing diagnostic advancements with patient anonymity remains a complex dilemma. Consequently, privacy-preserving techniques for deep learning training and inference are gaining significant traction. While state-of-the-art CAD methodologies are currently used to evaluate sequential image streams, privacy risks associated with transferring

medical media to external servers endure. This study introduces an innovative approach using a convolutional Bi-LSTM network to process fully homomorphic encrypted, time-series medical imagery. The reliability of this framework was validated using two rigorous benchmarks: the CheXpert dataset and the public BreakHis dataset. The findings demonstrate that the proposed model outperforms multiple competing techniques, securing an accuracy exceeding 0.99 on both datasets. This confirms that the system effectively extracts visual representations and models sequential temporal dynamics from encrypted medical imagery while maintaining confidentiality.

Looking ahead, future research should continue to investigate and refine privacy-preserving deep learning methodologies for training and inference on sensitive clinical data. Frameworks such as federated learning can be explored to uphold patient anonymity without sacrificing the operational efficacy of healthcare algorithms. By advancing sophisticated cryptographic methods, including homomorphic encryption and secure multiparty computation, researchers can formulate resilient encryption models that protect data privacy while enabling precise longitudinal analysis of medical image sequences. The ultimate objective is to achieve a balance that secures privacy while maintaining the utility and fidelity of medical datasets during deep learning processing. The integration of health policy and IoT infrastructures introduces both new opportunities and shared ethical obligations. To address these complexities while maintaining privacy, the industry can deploy targeted strategies: moving forward, convolutional blocks can extract spatial features from encrypted image arrays, while Bi-LSTM sequence analysis layers model temporal characteristics. To boost performance and minimize false negatives, a weighted unit and sequence voting layer can capitalize on spatial and temporal parameters using distinct weighting schemes.

Acknowledgements

None

Conflict of interest

None

Financial support

This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444).

Ethics statement

None

Received: 20 Jul 2024

Revised: 17 Aug 2024

Accepted: 24 Sep 2024

Published online: 15 March 2025

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Anand A, Singh AK. An improved DWT-SVD domain watermarking for medical information security. *Comput Commun.* 2020;152:72-80.
<https://doi.org/10.1016/j.comcom.2020.01.038>.
- Garcia-Hernandez JJ, Gomez-Flores W, Loyola JR. Analysis of the impact of digital watermarking on computer-aided diagnosis in medical imaging. *Comput Biol Med.* 2016;68:37-48.
<https://doi.org/10.1016/j.combiomed.2015.10.014>.
- Fan TY, Chao HC, Chieu BC. Lossless medical image watermarking method based on significant difference of cellular automata transform coefficient. *Signal Process Image Commun.* 2019;70:174-83.
- Ali Z, Imran M, Alsulaiman M, Shoaib M, Ullah S. Chaos-based robust method of zero-watermarking for medical signals. *Future Gener Comput Syst.* 2018;88:400-12.
- Wang X, Wan L, Huang M, Shen C, Han Z, Zhu T. Low-complexity channel estimation for circular and noncircular signals in virtual MIMO vehicle communication systems. *IEEE Trans Veh Technol.* 2020;69:3916-28.
- Natarajan V. Hybrid local prediction error-based difference expansion reversible watermarking for medical images. *Comput Electr Eng.* 2016;53:333-45.
- Gangadhar Y, Akula VSG, Reddy PC. An evolutionary programming approach for securing medical images using watermarking scheme in invariant discrete wavelet transformation. *Biomed Signal Process Control.* 2018;43:31-40.
- Sharma A, Singh AK, Ghrera SP. Secure hybrid robust watermarking technique for medical images. *Procedia Comput Sci.* 2015;70:778-84.
- Bouslimi D, Coatrieux G. A crypto-watermarking system for ensuring reliability control and traceability of medical images. *Signal Process Image Commun.* 2016;47:160-9.
<https://doi.org/10.1016/j.image.2016.05.021>.
- Liu C, Zhong D, Shao H. Data protection in palmprint recognition via dynamic random invisible watermark embedding. *IEEE Trans Circuits Syst Video Technol.* 2022;32:6927-40.
- Malayil MV, Vedhanayagam M. A novel image scaling based reversible watermarking scheme for secure medical image transmission. *ISA Trans.* 2021;108:269-81.
<https://doi.org/10.1016/j.isatra.2020.08.019>.
- Li XB, Qin J. Anonymizing and sharing medical text records. *Inf Syst Res.* 2017;28:332-52.
- Price WN, Cohen G. Privacy in the age of medical big data. *Nat Med.* 2019;25:37-43.
- Hua Z, Yi S, Zhou Y. Medical image encryption using high-speed scrambling and pixel adaptive diffusion. *Signal Process.* 2018;144:134-44.
<https://doi.org/10.1016/j.sigpro.2017.10.004>.
- Silva-García VM, Flores-Carapia R, Rentería-Márquez C, Luna-Benoso B, Aldape-Pérez M. Substitution box generation using chaos: An image encryption application. *Appl Math Comput.* 2018;332:123-35.
<https://doi.org/10.1016/j.amc.2018.03.019>.
- Liu Y, Tang S, Liu R, Zhang L, Ma Z. Secure and robust digital image watermarking scheme using logistic and RSA encryption. *Expert Syst Appl.* 2018;97:95-105.

Wang Z, Li M, Wang H, Jiang H, Yao Y, Zhang H, et al. Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features. *IEEE Access*. 2019;7:105146-58.
<https://doi.org/10.1109/ACCESS.2019.2892795>.

Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W. Deep learning to improve breast cancer detection on screening mammography. *Sci Rep*. 2019;9:12495.
<https://doi.org/10.1038/s41598-019-48995-4>.

Lee RS, Gimenez F, Rubin ADH. Curated breast imaging subset of DDSM. *Cancer Imaging Archive Technical Report*. 2016.

Zhang H, Wu C, Zhang Z, Zhu Y, Lin H, Zhang Z, et al. ResNeSt: Split-attention networks. *arXiv*. 2020;arXiv:2004.08955.

Zhang Y, Wang S, Wu H, Hu K, Ji S. Brain tumors classification for MR images based on attention guided deep learning model. In: *Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2021; Jalisco, Mexico*. p. 3233-6.

Jin Y, Dou Q, Chen H, Yu L, Heng PA. EndoRCN: Recurrent convolutional networks for recognition of surgical workflow in cholecystectomy procedure video. *IEEE Trans Med Imaging*. 2016. [Bibliographic details in the submitted list are incomplete; the original entry "53347671" is not a valid volume/page specification.]

Ghosh P, Azam S, Hasib KM, Karim A, Jonkman M, Anwar A. A performance based study on deep learning algorithms in the effective prediction of breast cancer. In: *Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN); 2021; Shenzhen, China*. p. 1-8.
<https://doi.org/10.1109/IJCNN52387.2021.9534293>.

Zheng J, Lin D, Gao Z, Wang S, He M, Fan J. Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis. *IEEE Access*. 2020;8:96946-54.
<https://doi.org/10.1109/ACCESS.2020.2993536>.

Yan Y, Zhao K, Cao J, Ma H. Prediction research of cervical cancer clinical events based on recurrent neural network. *Procedia Comput Sci*. 2021;183:221-9.
<https://doi.org/10.1016/j.procs.2021.02.052>.

Zhang X, Li R, Dai H, Liu Y, Zhou B, Wang Z. Localization of myocardial infarction with multi-lead bidirectional gated recurrent unit neural network. *IEEE Access*. 2019;7:161152-66.
<https://doi.org/10.1109/ACCESS.2019.2946932>.

Fan J, Vercauteren F. Somewhat practical fully homomorphic encryption. *IACR Cryptol ePrint Arch*. 2012;2012:144.

Samardzic N, Feldmann A, Krastev A, Manohar N, Genise N, Devadas S, et al. CraterLake: A hardware accelerator for efficient unbounded computation on encrypted data. In: *Proceedings of the 49th Annual International Symposium on Computer Architecture (ISCA '22); 2022; New York, NY*. p. 173-87.

Mert AC, Öztürk E, Savas E. Design and implementation of encryption/decryption architectures for BFV homomorphic encryption scheme. *IEEE Trans Very Large Scale Integr VLSI Syst*. 2020;28(2):353-62.
<https://doi.org/10.1109/TVLSI.2019.2943127>.

Ibarrondo A, Chabanne H, Despiegel V, Önen M. Colmade: Collaborative masking in auditable decryption for BFV-based homomorphic encryption. In: *Proceedings of the 2022 ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '22); 2022; New York, NY*. p. 129-39.

Yang H, Liang S, Zhang Y, Li X. Cloud-based privacy-and integrity-protecting density peaks clustering. *Future Gener Comput Syst*. 2021;125:758-69.

Zhang X, Chen C, Xie Y, Chen X, Zhang J, Xiang Y. A survey on privacy inference attacks and defenses in cloud-based deep neural network. *Comput Stand Interfaces*. 2023;83:103672.
<https://doi.org/10.1016/j.csi.2022.103672>.

Natsheh Q, Sălăgean A, Zhou D, Edirisinghe E. Automatic selective encryption of DICOM images. *Appl Sci*. 2023;13(8):4779.
<https://doi.org/10.3390/app13084779>.

Kanso A, Ghebleh M. An efficient and robust image encryption scheme for medical applications. *Commun Nonlinear Sci Numer Simul*. 2015;24:98-116.

Song W, Fu C, Zheng Y, Tie M, Liu J, Chen J. A parallel image encryption algorithm using intra bitplane scrambling. *Math Comput Simul*. 2023;204:71-88.

Ding Y, Wu G, Chen D, Zhang N, Gong L, Cao M, et al. DeepEDN: A deep-learning-based image encryption and decryption network for Internet of Medical Things. *IEEE Internet Things J*. 2021;8(3):1504-18.
<https://doi.org/10.1109/JIOT.2020.3012452>.

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS 2015)*; 2015; Montreal, QC. p. 2672-80.

Liu W, Liu X, Ma H, Cheng P. Beyond human-level license plate super-resolution with progressive vehicle search and domain priori GAN. In: *Proceedings of the 25th ACM International Conference on Multimedia*; 2017; Mountain View, CA. p. 1618-26.

Yi Z, Zhang H, Tan P, Gong M. DualGAN: Unsupervised dual learning for image-to-image translation. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*; 2017; Venice, Italy. p. 2868-76.

Radanliev P, De Roure D. Epistemological and bibliometric analysis of ethics and shared responsibility—health policy and IoT systems. *Sustainability*. 2021;13(15):8355.

Jain D. Regulation of digital healthcare in India: Ethical and legal challenges. *Healthcare (Basel)*. 2023;11(6):911.
<https://doi.org/10.3390/healthcare11060911>.

Zhang Z, Gao Q, Liu L, He Y. A high-quality rice leaf disease image data augmentation method based on a dual GAN. *IEEE Access*. 2023;11:21176-91.

Liu X, Zhang T, Zhang J. Toward visual quality enhancement of dehazing effect with improved Cycle-GAN. *Neural Comput Appl*. 2023;35:5277-90.

Panzade P, Takabi D. FENet: Privacy-preserving neural network training with functional encryption. In: *Proceedings of the 9th ACM International Workshop on Security and Privacy Analytics (IWSPA '23)*; 2023; New York, NY. p. 33-43.

Zhao D. Communication-efficient search under fully homomorphic encryption for federated machine learning. *arXiv*. 2023;arXiv:2308.04648.

Li Q, Lai Y, Adamu MJ, Qu L, Nie J, Nie W. Multi-level residual feature fusion network for thoracic disease classification in chest X-ray images. *IEEE Access*. 2023;11:40988-1002.
<https://doi.org/10.1109/ACCESS.2023.3269068>.