Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

A Robustness-Aware Hybrid Deep Learning Framework With Blockchain-Based Auditability For Lung Cancer Detection

Hari Krishna Kalidindi[1], N Srinivasu[2]

[1]Department of CSE,K.L.University , Andhra Pradesh, India
[2]Department of CSE, K.L.University , Andhra Pradesh, India
Email: [1]harikrishna79.kalidindi@gmail.com , [2]srinivasu28@kluniversity.in

Abstract:
Accurate and reliable detection of lung cancer from Computed Tomography (CT) scans is essential for early diagnosis and effective clinical decision-making. Although hybrid deep learning models have demonstrated promising performance in medical image analysis, their robustness under realistic imaging degradations remains insufficiently explored. Moreover, safe storage and traceable reporting of diagnostic results are vital in, credible clinical implementation. The proposed paper is a robustness-conscious blockchain-based hybrid deep learning framework of secure lung cancer detection. The suggested system combines the HCNN-ALSTM and AHyNet systems in one experimental pipeline and explicitly simulates real-world CT degradations, such as Gaussian noise, motion blur, and partial occlusion. Relative accuracy degradation is used to develop a Robustness Index (RI) to quantitatively assess diagnostic stability. CT images are encrypted and stored off-chain, and diagnostic outputs and robustness measures are stored immutably on a block chain to provide integrity and auditing. Experimental performance on the publicly accessible LIDC-IDRI dataset , with an 80:20 traintest split with fivefold cross-validation, shows that AHyNet can perform at 98.08% accuracy on clean data, and that its performance is better under perturbed conditions, with an overall robustness index of 0.96. AHyNet, in comparison to the traditional CNN models, has a 6.7% higher classification accuracy and a 7-12% higher robustness in cases of noise, blur, and occlusions. Integration of blockchain adds little inference overhead, but maintains the possibility of real-time diagnostics. The results confirm that the proposed framework improves diagnostic accuracy and also significantly enhances robustness and security, thereby advancing reliable and clinically practical artificial intelligence systems for lung cancer diagnosis.

Keywords Lung cancer detection; Computed tomography (CT); Hybrid deep learning; Robustness Index (RI); Medical image analysis; Blockchain-based auditability; Attention mechanisms; Bidirectional recurrent neural network (Bi-RNN);

Introduction

Lung cancer is still one of the leading causes of cancer deaths worldwide, mainly because of late diagnosis and the subtle visual characteristics of early-stage pulmonary nodules on computed tomography (CT) scans.Early and accurate detection is therefore critical for improving patient survival rates. In recent years, computer-aided diagnosis (CAD) systems based on deep learning have emerged as powerful tools to assist radiologists in lung cancer screening and classification tasks [6], [7]. Convolutional Neural Networks (CNNs) in particular have demonstrated strong capability in extracting discriminative features from medical images, enabling significant improvements over traditional handcrafted approaches.

Despite these advances, conventional deep learning models often exhibit limited robustness when exposed to real-world imaging degradations. In clinical environments, CT images are frequently affected by acquisition noise, motion artifacts, and partial occlusion caused by patient movement or incomplete scans. Such perturbations can substantially degrade diagnostic performance, leading to unreliable predictions. Although hybrid architectures combining convolutional feature extraction with recurrent learning and attention mechanisms have been proposed to enhance representation capability [8], [9], most existing studies continue to evaluate performance primarily on clean datasets, thereby overlooking diagnostic stability under realistic

conditions.
Recently, robustness has been recognized as a critical requirement for trustworthy medical artificial intelligence. Models that perform well under controlled experimental settings may fail when deployed in practical healthcare environments, where imaging quality varies significantly [10]. While data augmentation and regularization strategies have been explored to improve generalization, systematic robustness quantification and stability-aware evaluation remain largely underrepresented in lung cancer diagnostic pipelines.
Separately, increasing concerns regarding patient privacy, data integrity, and accountability have driven growing interest in blockchain technology for healthcare applications. Blockchain provides decentralized, tamper-resistant record keeping and transparent audit trails, enabling secure storage and traceable access to sensitive medical data [11]. A number of blockchain-based medical imaging platforms have been suggested to safeguard patient data and enable secure information exchange between institutions. Nonetheless, the current blockchain-based models are mostly aimed at data management and access control, and do not include diagnostic reliability indicators or robustness metrics in the security infrastructure.
These limitations point to a severe research gap: existing lung cancer detection methods are mostly focused on accuracy when the conditions are theoretically ideal, but do not consider the robustness to the effects of realistic imaging degradations and the safe recording of diagnostic accuracy. No single frameworks yet exist that collectively tackle the diagnosis performance, robustness analysis, and blockchain-based integrity through a single clinically practical architecture.
In response to these issues, this paper will present a hybrid deep learning architecture that is blockchain-enabled and robust to detect lung cancer using CT images. In contrast to traditional accuracy-based methods, the proposed framework explicitly simulates realistic acquisition degradations, such as Gaussian noise, motion blur, and partial occlusion, and proposes a Robustness Index (RI) to measure diagnostic stability quantitatively. Our HCNN-ALSTM architecture, which is based on our earlier published study [19], is a benchmark hybrid classifier. Moreover, an Adaptive Hybrid Network (AHyNet) that uses the Pyramid Dilated Dense Attention learning[20] and Bidirectional Recurrent Neural Networks[21] is used to improve learning of multi-scale feature extraction and contextual representation. The logging with the help of blockchain provides the impossibility of diagnostic output records and robustness measures to be transparently verified.
Although conventional encrypted databases provide secure storage and role-based access control, they remain dependent on centralized administrative control and therefore cannot inherently guarantee immutable and independently verifiable audit trails. In contrast, blockchain records are cryptographically linked, making unauthorized modifications tamper-evident while enabling transparent verification of diagnostic history. Since the proposed framework aims to preserve not only diagnostic predictions but also the associated robustness information for future clinical auditing, blockchain serves as an auditability layer rather than merely a storage mechanism.
Although earlier research has indicated a high classification accuracy in detecting lung cancer, relatively less effort has been given to assessing diagnostic stability in the presence of realistic imaging degradations. Moreover, the current blockchain-based medical imaging systems are mainly oriented towards data security and access control, without including quantitative indicators of diagnostic reliability. As a result, there is still a necessity to have an integrated framework that will be able to provide simultaneously classification accuracy, robustness evaluation, and secure auditability in clinical decision-support systems.
Existing lung cancer diagnostic studies primarily focus on improving classification accuracy using hybrid deep learning models evaluated on clean benchmark datasets. However, they do not provide a unified quantitative measure for assessing the stability of diagnostic performance under realistic imaging degradations that frequently occur in clinical practice. Consequently, objective comparison of the robustness of different diagnostic models remains a significant challenge. To address this gap, the present work introduces a Robustness Index (RI) as the central methodological contribution of a robustness-aware evaluation framework. The proposed RI quantitatively measures the preservation of diagnostic performance across multiple clinically relevant perturbation scenarios, thereby complementing conventional classification metrics and enabling systematic robustness assessment for trustworthy AI-assisted lung cancer diagnosis.
The major contributions of this work are summarized as follows:
(i) A robustness-aware evaluation framework is proposed for assessing the reliability of hybrid deep learning models under clinically realistic CT image degradations.
(ii) A novel Robustness Index (RI) is introduced to quantitatively measure diagnostic stability across multiple perturbation scenarios using a normalized evaluation metric.

(iii) Blockchain-assisted auditability is incorporated to securely record both diagnostic predictions and robustness information, thereby improving transparency and traceability.
(iv) Extensive experiments on the LIDC-IDRI dataset demonstrate that the proposed framework improves diagnostic robustness while maintaining high classification accuracy and minimal blockchain overhead
In our previous publication [19], we have introduced the HCNN-ALSTM model and this was found to be effective in learning features because it combines the convolutional neural networks, latent representation learning with autoencoders and long short-term memory networks. The architecture has demonstrated encouraging results when it comes to the diagnosis of lung cancer and is used in the current study as a conventional hybrid classifier to determine the power of the classifier relative to each other.
The study will take a step forward to plausible and clinically viable artificial intelligence systems to diagnose lung cancer through association of diagnostic accuracy, robustness awareness and secure traceability of results.

2. Related Work

Deep learning has become the dominant paradigm for automated lung cancer detection from CT images. Initial convolutional neural network (CNN)-based models have shown encouraging results on pulmonary nodule detection and malignancy classification, but these methods were typically limited in terms of generalization and sensitivity to artifacts in images [12]. To overcome these shortcomings, it was proposed to use hybrid deep learning models that included convolutional feature generation with recurrent or densely connected networks to learn finer spatial and contextual features [8].
The accuracy of the diagnosis has also been improved by the mechanisms of attention-based learning since the models are able to focus on the salient portions of the lung and leave out the irrelevant background information. Recent research using attention modules and multi-scale feature extraction has documented large improvements in the classification of lung nodules, especially in high-complexity or low-contrast CT scans [9], [5], [13]. However, the majority of these methods assess performance only using clean datasets, which do not give a strong understanding of diagnostic stability to real imaging degradations.
Medical image analysis has recently become a key research direction, and robustness is one such critical research direction. A number of studies have shown that deep learning models are susceptible to noise, blur, and adversarial perturbations, and thus unreliable clinical predictions can be made [10], [17].Despite the research on robustness improvement methods like data augmentation and regularization, systematic robustness quantification and stability-conscious evaluation have been well-understood in lung cancer diagnostic pipelines [18].
Almost recent developments in hybrid deep learning have incorporated dense connectivity, dilated convolution, attention, and recurrent learning to enhance the ability to represent in analyzing medical images. These architectures allow the simultaneous derivation of multi-scale spatial features and contextual dependencies, and are thus especially effective in complex diagnostic tasks with heterogeneous imaging patterns[8][9][13].
Like the progress in algorithms, blockchain technology has found its way into the healthcare industry to ensure integrity of data, decentralized access control, and audit trails [11]. It has been proposed to use blockchain to support medical imaging systems to store sensitive patient data securely and to be shared across hospitals [14], [15]. Ensemble and hybrid learning platforms that incorporate blockchain have also shown enhanced diagnostic accuracy and privacy protection [16] more recently. Nevertheless, these studies are mostly concerned with the data security and accuracy improvement, without directly considering the diagnostic robustness in the presence of acquisition artifacts.
Table 1 summarizes the representative works in the detection of lung cancer and blockchain-enabled medical imaging to provide a systematic comparison of the current literature.
In the recent past, hybrid deep learning systems that combine various mechanisms of learning features have shown better diagnostics of lung cancer. Kalidindi and Srinivasu [19] suggested a blockchain-aided deep ensemble architecture that includes Hybrid Convolutional Neural Network with Autoencoder and Long Short-Term Memory (HCNN-ALSTM) architecture. The framework enhanced the classification accuracy by incorporating spatial feature extraction, latent feature compression, and sequential learning, and provided a safe way of sharing medical data with blockchain technology. Nonetheless, the strength to realistic imaging degradations was not explored.

Table 1. Comparative review of existing lung cancer detection and secure medical imaging approaches

Author	Method	Dataset	Accuracy%	Robustness	Blockchain	Limitation
Setio et al. [12]	Multi-view CNN	LIDC-IDRI	90.1	No	No	Sensitive to noise and artifacts
Ardila et al. [7]	3D CNN	NLST	94.4	No	No	Evaluated only on clean scans
Wang et al. [13]	Multi-scale attention network	LIDC-IDRI	95.3	Limited	No	No robustness quantification
Finlayson et al. [10]	Adversarial analysis	Multiple	—	Yes	No	No diagnostic framework
Zhang et al. [14]	Blockchain-based image sharing	Private CT	—	No	Yes	Focuses only on data security
Islam et al. [16]	Blockchain ensemble DL	Medical datasets	96.2	No	Yes	Lacks degradation modeling
	HCNN-ALSTM + Blockchain	LIDC-IDRI	97+	No	Yes	No robustness evaluation

Accuracy improvement is the focus of most of the current research, whereas the quantification of robustness and the secure diagnostic traceability are not explored thoroughly. Moreover, existing blockchain-based solutions are mostly focused on data storage and accessibility, without the addition of diagnostic reliability measures. This leads to a shortage of integrated solutions that can all take into account diagnostic accuracy, resilience to realistic imaging degradations, and blockchain-based integrity in the same system-level design.
The current work is driven by these constraints and proposes a resiliency-conscious blockchain-based hybrid deep learning model that can collectively tackle performance, stability, and safe result recording, which will take a step forward toward dependable and clinically viable lung cancer diagnostic systems.

2.1 Research Gaps

Although there have been significant advances in deep learning-based lung cancer detection, there are still a number of significant limitations in the current literature. To begin with, the majority of the diagnostic models are tested using clean CT data, and the effect of realistic acquisition corruptions, including noise, motion blur, and partial occlusions, which are frequent in clinical settings, are not taken into account [17]. As a result, the reported performance measures are mostly not representative of diagnostic reliability in the real world.

Second, despite the demonstrated superior classification accuracy of attention-based and hybrid architectures, systematic quantification of robustness is mostly absent in diagnostic pipelines of lung cancer [10], [18]. The existing methods are mainly focused on improving the accuracy aspect, whereas they neglect the stability of diagnosis in disturbed imaging conditions.

Third, secure data storage and access control are the primary concerns of the blockchain-based healthcare systems with little consideration of the indicators of diagnostic reliability or robustness metrics being considered as part of the security framework [11], [14]. Therefore, current medical imaging systems powered by blockchains do not have ways of delivering traceable confidence or stability data and prediction results.

Lastly, although ensemble and hybrid deep learning models have demonstrated good results, there is no cohesive framework that can simultaneously focus on diagnostic accuracy, resilience to realistic imaging degradations, and the ability to log results in a single system-level framework [15], [16].These gaps motivate the need for reliability-aware diagnostic systems that jointly consider performance, robustness, and security for practical clinical deployment.

2.2 Objectives of This Paper

Motivated by the above research gaps, this study aims to achieve the following objectives:

1. To develop a unified hybrid deep learning framework for lung cancer detection by integrating HCNN-ALSTM and AHyNet architectures within a common experimental pipeline.
2. To introduce robustness-aware evaluation by simulating realistic CT image degradations, including Gaussian noise, motion blur, and occlusion, and by defining a Robustness Index to quantitatively assess diagnostic stability.
3. To incorporate blockchain-assisted secure storage for encrypted CT images while immutably logging diagnostic outputs and robustness metrics to enhance integrity and auditability.
4. To perform comprehensive experimental validation, assessing diagnostic accuracy, robustness, and security–performance trade-offs for clinically feasible deployment.

3. Proposed Robustness-Aware Blockchain-Enabled Methodology

In this section, the authors introduce the suggested robustness-conscientious blockchain-based hybrid deep learning model of lung cancer detection using CT images. Unlike the conventional diagnostic pipelines where the performance of the tests is only tested with clean data, the proposed methodology explicitly models realistic acquisition degradations and quantitatively establishes diagnostic stability. In addition, blockchain also supports logging to offer safe and traceable records of diagnostic outcomes and robustness indicators.

As shown in Fig. 1, the framework comprises of six key steps, namely: (i) CT image preprocessing, (ii) robustness simulation, (iii) secure off-chain storage with blockchain access control, (iv) hybrid deep learning-based classification, (v) robustness evaluation, and (vi) blockchain-based integrity logging.

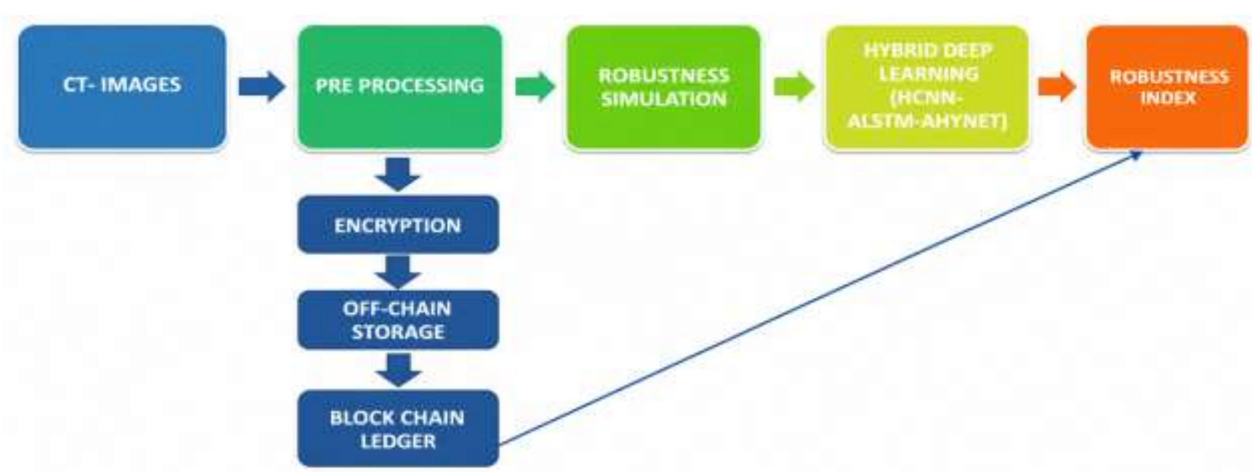

Figure 1. Overall architecture of the proposed robustness-aware blockchain-enabled lung cancer detection framework.

3.1 Image Preprocessing

Effective preprocessing is essential for improving image quality, reducing noise, and ensuring consistent input representation for deep learning models. The LIDC-IDRI CT scans undergo a sequence of preprocessing operations before classification.

Initially, raw CT volumes are converted into axial slices and lung windowing is applied to enhance pulmonary structures and suppress irrelevant anatomical information. Subsequently, median filtering is employed to reduce acquisition noise while preserving edge information. The median-filtered image is expressed as:

$I_m(x,y)=\text{Median}\{I(i,j)\}$ ……..(1)

where I(x,y) denotes the original CT image and $I_m(x,y)$ represents the filtered image obtained from the neighborhood pixels (i,j).

To improve contrast and enhance nodule visibility, histogram equalization is applied. The transformed intensity value is computed as:

$H(v)=\text{round}((L-1) \times CDF(v))$ ………(2)

where L represents the number of gray levels and CDF(v) denotes the cumulative distribution function of intensity value v.

Lung regions are then segmented using threshold-based segmentation followed by morphological refinement to eliminate non-lung structures. The segmented binary mask M(x,y) is obtained as:

$M(x,y)=1$, if $I(x,y)>T$

$=0$, otherwise …….(3)

where T denotes the segmentation threshold.

Finally, all segmented slices are resized to 224×224 pixels and normalized to ensure uniform intensity distribution across samples. Min-max normalization is performed as:

$$I_{norm}=\frac{I-Imin}{Imax-Imin} \qquad …….(4)$$

where Imin and Imax denote the minimum and maximum pixel intensities, respectively.

The preprocessing pipeline standardizes CT images, enhances feature visibility, and improves the robustness of subsequent HCNN-ALSTM and AHyNet classification stages[4][6][12].

3.2 Robustness Simulation Module

To evaluate diagnostic reliability under realistic clinical conditions, a robustness simulation module is incorporated into the proposed framework. Unlike conventional evaluation strategies that rely solely on clean CT images, the proposed methodology systematically introduces controlled perturbations to emulate acquisition artifacts commonly encountered in medical imaging environments.

Three categories of perturbations are considered: Gaussian noise, motion blur, and partial occlusion. These perturbations represent scanner noise, patient movement during image acquisition, and incomplete or corrupted image regions, respectively.

Gaussian Noise Simulation

Gaussian noise is introduced to simulate random intensity fluctuations generated during CT image acquisition. The noisy image is defined as:

$$I_{noise}(x,y) = I(x,y) + N(0, \sigma^2) \qquad (5)$$

where I(x,y) denotes the original image, $N(0,\sigma^2)$ represents Gaussian noise with zero mean and variance σ^2, and $I_{noise}(x,y)$ is the resulting noisy image.

Three noise levels are considered:

- $\sigma = 0.01$
- $\sigma = 0.05$
- $\sigma = 0.10$

to evaluate model stability under increasing acquisition uncertainty.

Motion Blur Simulation

Motion blur is applied to emulate patient movement during CT scanning. The blurred image is obtained through convolution:

$$I_{blur} = I*H \qquad (6)$$

where H represents the motion blur kernel and * denotes convolution.
Kernel sizes of:

- 3×3
- 7×7
- 11×11

are employed to represent varying degrees of motion-induced degradation.

Partial Occlusion Simulation

Partial occlusion is introduced to model missing or corrupted image regions that may occur during image acquisition or transmission. The occluded image is generated as:

$$I_{occ}(x, y) = I(x, y) \times M(x, y) \qquad (8)$$

where M(x,y) denotes a binary masking function.
Random masking levels of:

- 5%
- 10%
- 20%

of the image area are considered.

Robustness-Aware Dataset Generation

For each original CT image, multiple perturbed versions are generated across different degradation levels. The resulting robustness-aware dataset enables systematic analysis of classification stability and performance degradation. By evaluating HCNN-ALSTM and AHyNet under controlled perturbations, the proposed framework provides a realistic assessment of diagnostic reliability beyond conventional clean-image testing scenarios.

3.3 Secure Off-Chain Storage with Blockchain Access Control

The medical imaging systems should possess secure storage, limited access and diagnostic records which can be traced. To meet these needs, the suggested framework uses a blockchain-based storage design that integrates off-chain image repositories with on-chain verification of integrity.

CT image datasets are extremely large and therefore having complete datasets of images stored on the blockchain would create a large storage overhead and transaction delay. Therefore, the ciphertext CTs are saved in off-chain distributed repositories and the cryptographic hash values and diagnostic metadata are saved in the blockchain.

For each CT image, a SHA_{256} hash is generated as:

$$H = SHA_{256}(I) \qquad (7)$$

where I denotes the encrypted CT image and H represents the corresponding unique hash value.

The generated hash serves as a digital fingerprint for verifying image integrity. Any modification to the stored image produces a different hash value, enabling tamper detection.

A private Ethereum-based blockchain network is employed to maintain immutable records of:

- Image hash values
- Diagnostic predictions
- Robustness Index (RI) scores
- Timestamp information
- Access logs

Smart contracts control access control and authorization. Authenticated healthcare personnel are the only ones who can access diagnostic information based on established contract rules.

A Proof-of-Authority (PoA) consensus is adopted to reduce the computation load and increase the transaction throughput. Unlike Proof-of-Work systems, PoA does not require intensive mining, and can be used to validate transactions with low latency, which is suitable in the healthcare context where real-time responsiveness is required.

The deep learning inference engine is not relied on in the blockchain layer. This results in image classification being done off-chain, and integrity verification and audit logging being done on-chain. This hybrid design is safe, can be traced and is efficient to implement clinically without impacting the diagnostic performance[11], [14], [15].

Although conventional secure databases can provide encryption, authentication, and role-based access control, they remain centrally managed and rely on trusted administrators to maintain data integrity and audit records. In contrast, the proposed blockchain layer stores only cryptographic hashes, diagnostic metadata, and robustness scores in an immutable distributed ledger, making any unauthorized modification immediately detectable. By combining off-chain encrypted CT image storage with on-chain integrity verification, the proposed framework complements conventional database security while providing enhanced auditability and tamper resistance with minimal computational overhead.

3.4 Hybrid Deep Learning Classification

In the proposed framework, HCNN-ALSTM and AHyNet are two hybrid deep learning architectures to achieve accurate and robust lung cancer diagnosis. These architectures are developed to extract spatial and contextual information of CT images and be resilient to imaging degradations.

3.4.1 HCNN-ALSTM Architecture

The HCNN-ALSTM model that is used in this research is founded on the architecture used in our previous research [19]. The model combines Convolutional Neural Networks (CNN), Autoencoders (AE), and Long Short-Term Memory (LSTM) networks to concurrently utilize spatial, compressed latent, and sequential feature representations.

The first stage is convolutional layers, which learn discriminative spatial features of preprocessed CTs. Convolution operation can be written as:

$$F_k = f\,(X*K_k+b_k) \qquad (9)$$

where X denotes the input image, K_k represents the kth convolution kernel, and b_k denotes the corresponding bias term.

The extracted features are subsequently compressed through an autoencoder to obtain compact latent representations while minimizing redundant information. The encoding process is defined as:

$$Z = f(W_e\,X+b_e) \qquad (10)$$

where Z represents the latent feature vector, We denotes the encoder weight matrix, and be denotes the encoder bias.

The compressed feature maps are then flattened into an ordered feature vector sequence, which serves as the input to the LSTM. The LSTM models contextual dependencies among these learned feature representations rather than processing adjacent CT slices.

Finally, LSTM units learn contextual dependencies among extracted feature sequences. The hidden state update is represented as:

$$H_t=LSTM(X_t, h_{t-1}) \qquad (11)$$

where xt denotes the current feature input and ht−1 represents the previous hidden state.

The combination of CNN, autoencoder, and LSTM components enables HCNN-ALSTM to effectively learn complex lung nodule characteristics while reducing feature redundancy[19].

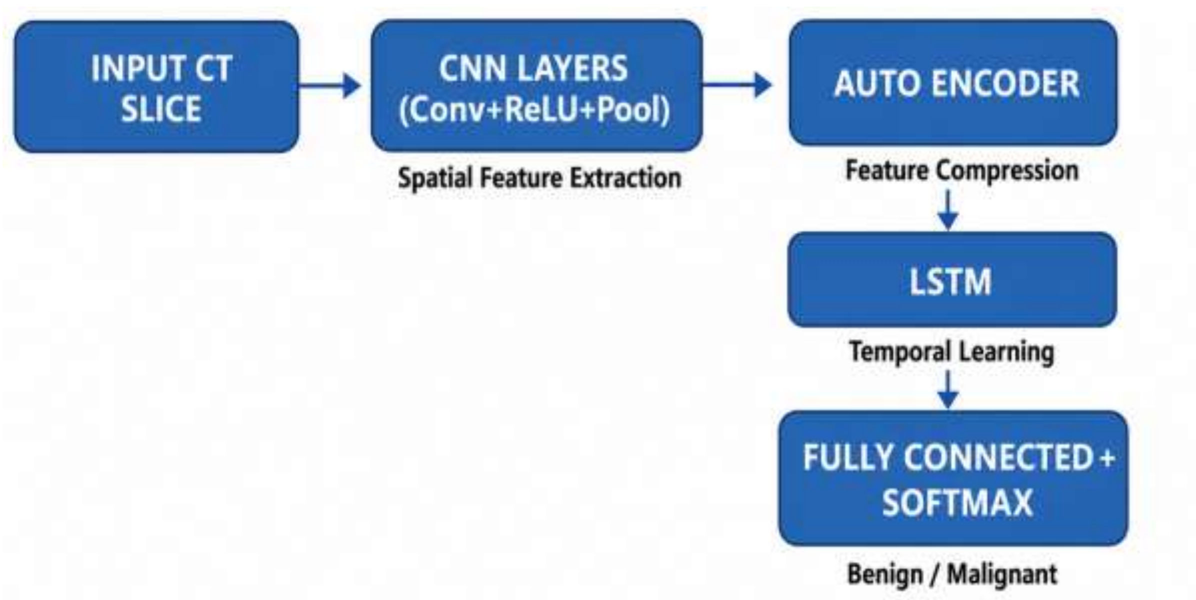

Figure 2. Architecture of the HCNN-ALSTM hybrid deep learning model.

3.4.2 AHyNet Architecture

The proposed framework is supplemented by an Adaptive Hybrid Network (AHyNet) that increases the quality of the diagnosis and its resilience. AHyNet is based on a Pyramid Dilated Dense Attention Network (PDDANet) that encodes multi-scale spatial and contextual information of CT images with a Bidirectional Recurrent Neural Network (Bi-RNN).

Unlike older CNN systems, which employ fixed receptive fields, AHyNet employs dilated convolutions to dilate the receptive field without increasing the amount of computational effort. Dilated convolution can be defined as: $y(i) = \sum_{k=1}^{k} x(i + r \cdot k)w(k)$ (12)

where r denotes the dilation rate, w(k) represents the convolution kernel, and y(i) is the output feature map.

The Pyramid Dilated Dense Attention Network extracts features at multiple scales through parallel dilated convolution paths. Dense connectivity enables feature reuse across layers, thereby improving information propagation and reducing feature loss.

To emphasize diagnostically relevant pulmonary regions, an attention mechanism is incorporated. Attention weights are computed using:

$$\alpha_i = \frac{\exp(s_i)}{\sum_j \exp(s_i)} \quad (13)$$

where α_i denotes the attention weight corresponding to feature score s_i.

The attention-enhanced feature maps are flattened into an ordered sequence of feature vectors before being supplied to the Bidirectional Recurrent Neural Network (Bi-RNN). The Bi-RNN captures contextual relationships among these feature vectors by processing the sequence in both forward and backward directions.The resulting attention-enhanced feature maps are subsequently processed by a Bidirectional Recurrent Neural Network. Bi-RNN is used to utilize both forward and backward contextual information at the same time, unlike unidirectional recurrent learning. The bidirectional hidden representation can be defined as:

$$H_t = \left\lfloor \overrightarrow{h_t} ; \overleftarrow{h_t} \right\rfloor \quad (14)$$

where forward and backward hidden states are concatenated to generate a comprehensive contextual representation.

The last layer of classification uses the activation function of Softmax to classify CTs as benign and malignant. AHyNet provides a better classification accuracy and a greater ability to resist image degradations by combining dense feature propagation, multi-scale attention learning, and bidirectional contextual modeling.

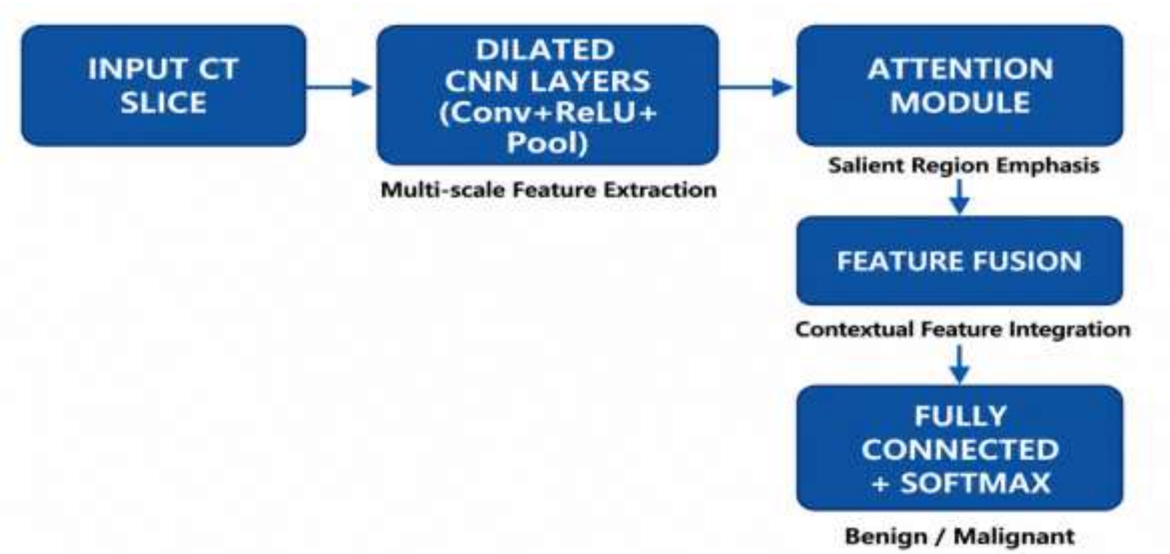

Figure 3. Architecture of the AHyNet model incorporating attention mechanisms and dilated convolutions.

AHyNet accepts normalized CT images of size 224 × 224 × 1 as input. The network employs successive convolutional layers with 3 × 3 kernels and multi-scale dilated convolutions using dilation rates of 1, 2, and 4 to capture local and global features. Dense feature concatenation enhances feature reuse, while an attention module emphasizes diagnostically relevant regions. The resulting feature maps are converted into an ordered feature sequence and processed by a Bi-RNN with 128 hidden units in each direction. A dropout layer (0.5) is applied to reduce overfitting, and the fused feature representation is finally classified using a fully connected Softmax layer for benign and malignant prediction.

The use of both HCNN-ALSTM and AHyNet enables a comprehensive evaluation of robustness across different hybrid learning paradigms. While HCNN-ALSTM emphasizes compressed latent feature learning and temporal dependency modeling, AHyNet focuses on multi-scale attention-guided feature extraction and bidirectional contextual representation. Consequently, comparative robustness analysis can reveal the relative strengths of distinct hybrid architectures under realistic imaging degradations.

The proposed AHyNet consists of four Pyramid Dilated Dense Attention (PDDA) blocks with dilation rates of {1, 2, 4}. Each convolutional layer employs 3 × 3 kernels with 64, 128, 256, and 512 filters, respectively, followed by batch normalization and ReLU activation. Dense skip connections are used within each PDDA block to facilitate feature reuse. A channel-wise attention module is applied after multi-scale feature extraction, and the resulting feature maps are processed by a Bi-RNN containing 128 hidden units in each direction. A dropout rate of 0.5 is applied before the final fully connected Softmax classifier to reduce overfitting. The network is trained using the Adam optimizer with a learning rate of 0.0001 and binary cross-entropy loss.

3.5 Robustness Evaluation

Diagnostic accuracy in itself is not usually adequate to determine the practical functionality of medical artificial intelligence systems. A model can be very high-performing on clean data and then significantly degrade when presented with realistic imaging artifacts. Thus, along with the traditional metrics of classification, the strategy of evaluation based on robustness is also included in the proposed framework.

Accuracy, Precision, Recall, F1-score, Matthews Correlation Coefficient (MCC) and Area Under the Receiver Operating Characteristic Curve (AUC) are used to evaluate model performance. These are measures that are employed to gauge diagnostic performance in clean and perturbed imaging.

Conventional performance metrics such as Accuracy, Precision, Recall, F1-score, MCC, and AUC evaluate predictive performance under specific testing conditions but do not quantify the stability of diagnostic performance when CT images are affected by realistic acquisition degradations. Since reliable clinical deployment requires models that maintain consistent performance under varying imaging conditions, this work introduces a Robustness Index (RI) as a normalized measure of diagnostic stability. Unlike conventional evaluation metrics that independently report performance for each perturbation scenario, RI summarizes the overall robustness of a diagnostic model using a single interpretable score, thereby enabling objective comparison of different models under identical degradation conditions.

A Robustness Index (RI) is introduced to measure the stability against image degradations. The RI measures how well classification accuracy is preserved in several perturbation cases and is measured as:

$$\mathrm{RI} = 1 - \frac{1}{N}\sum_{i=1}^{N}\frac{A_c - A_i}{A_c} \qquad (15)$$

where:

- A_c denotes classification accuracy obtained on clean CT images,
- A_i represents accuracy under the ith perturbation condition,
- N denotes the total number of perturbation scenarios.

The RI value lies within the interval: $0 \le \mathrm{RI} \le 1$

A value of RI close to 1 means a high robustness and low performance degradation and a smaller value means that the performance is more sensitive to image perturbation.

The proposed Robustness Index (RI) is based on the principle of normalized relative performance degradation, which has been widely adopted in robustness evaluation for machine learning models. Instead of considering absolute accuracy alone, RI measures the average proportional reduction in classification performance across multiple perturbation scenarios relative to the clean-image baseline. The normalization by the clean accuracy enables fair comparison among models with different baseline performances while

constraining RI to the interval [0,1], where values closer to one indicate greater robustness. Consequently, RI provides an intuitive quantitative measure of diagnostic stability under realistic imaging degradations.
The proposed RI possesses several desirable characteristics. It is normalized within the interval [0,1], where values closer to one indicate greater robustness. It aggregates robustness across multiple perturbation scenarios into a single quantitative measure, is independent of the underlying deep learning architecture, and provides an interpretable basis for comparing the diagnostic stability of different classification models.
The proposed robustness assessment framework can compare the diagnostic stability of different hybrid architectures in a more realistic fashion, and provides a more realistic picture of the clinical deployment reliability than the conventional accuracy-only methods of evaluation.
Existing robustness evaluation methods commonly report perturbed accuracy, corruption error, adversarial robustness, or expected corruption error independently for each perturbation. Unlike these measures, the proposed RI summarizes the average normalized performance degradation over multiple clinically relevant perturbations into a single interpretable score. This provides a compact quantitative indicator for comparing diagnostic stability across different hybrid architectures while complementing conventional performance metrics such as Accuracy, AUC, and F1-score.

3.6 Blockchain-Based Robustness Logging

In conventional AI-assisted diagnostic systems, the prediction outcomes are usually placed in centralized repositories where any other person can easily modify, delete, or tamper with them. Moreover, the information about robustness is not usually maintained along with diagnostic results, which restricts the ability to trace and evaluate the reliability of automated clinical judgments.
The proposed framework will overcome these limitations by including blockchain-based robustness logging. After the classification, the diagnostic results of HCNN-ALSTM and AHyNet, along with their respective robustness measures, are safely stored on a private blockchain network.For each diagnostic instance, the following information is stored:

• Patient-independent image identifier
• Classification result (Benign/Malignant)
• Prediction confidence score
• Robustness Index (RI)
• Timestamp
• SHA-256 hash value
• Access transaction information

The blockchain transaction record can be represented as:

$$T = \{ID, C, RI, H, TS\} \quad (16)$$

where:

- ID denotes the image identifier,
- C represents the classification result,
- RI denotes the Robustness Index,
- H represents the SHA-256 hash value,
- TS denotes the transaction timestamp.

Smart contracts automatically check the correctness of the received transactions and apply access control policies before the records are added to the blockchain ledger forever.In contrast to the conventional healthcare information systems, blockchain-enhanced robustness logging offers unchangeable records of diagnostic results and reliability measures. As a result, healthcare practitioners are not only able to verify the prediction itself, but also the stability of the underlying model to realistic imaging perturbations.The blockchain technology integration consequently improves the levels of transparency, auditability and trustworthiness, and imposes little computational load as the inference processing and blockchain operations are decoupled.
The purpose of incorporating blockchain is not to replace conventional database systems for storing CT images. Instead, blockchain is used to provide immutable logging of diagnostic predictions and robustness metrics. Encrypted CT images remain stored off-chain, whereas only cryptographic hashes and diagnostic metadata are recorded on-chain. This hybrid architecture combines the storage efficiency of conventional databases with the tamper-evident auditability provided by blockchain.
Overall, the workflow of the proposed framework can be summarized in Algorithm 1.

Computational Complexity

The complexity of inference is about O(n d), where nnn is the number of CT slices and ddd is the depth of the network. As blockchain processes are confined to hash logging and metadata storage, model inference is conducted off-chain, maintaining real-time viability

Algorithm 1: Robustness-Aware Blockchain Hybrid Deep Learning Framework

Input: CT images from LIDC-IDRI dataset
Output: Robust lung cancer classification (Benign / Malignant) with robustness score
Step 1:Obtain raw CT scans in the dataset.
Step 2: Convert volumes to axial slices and lung windowing.
Step 3: Run median filter and histogram equalization.
Step 4:Division of the lung areas with thresholding and morphological operations.
Step 5:Resize slices to 224×224 and normalize pixel intensities.
Step 6: Create perturbed datasets through the addition of Gaussian noise, motion blur, and occlusion on a variety of intensity levels.
Step 7: Encrypt preprocessed images and store them in off-chain distributed storage.
Step 8: Register image metadata and access policy encrypted on the blockchain ledger.
Step 9: Retrieve encrypted images, which are authorized with an authorized request and decrypt them.
Step 10: Extract spatial features using convolutional layers.
Step 11: Classify with the HCNN-ALSTM and AHyNet hybrid models.
Step 12: Fine-tune model hyperparameters with adaptive optimization techniques.
Step 13: Compare the performance on clean and perturbed datasets with accuracy, precision, recall, F1-score, MCC and Robustness Index.
Step 14: Store diagnostic results and robustness measurements on the blockchain to be audited.
Step 15: Get final diagnosis and robustness score.

4. Experimental Setup

4.1 Dataset Description

The publicly available Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset [4] were used in the experiments obtained via The Cancer Imaging Archive (TCIA). It consists of 1,018 thoracic CT scans acquired in various medical institutions and marked by up to four experienced radiologists.
Every pulmonary nodule in the data set is given a malignancy score between 1 and 5 with:
• Rating 1: Highly unlikely for malignancy
• Rating 2: Moderately unlikely for malignancy
• Rating 3: Indeterminate
• Rating 4: Moderately suspicious for malignancy
• Rating 5: Highly suspicious for malignancy
Based on generally accepted annotation guidelines in studies of lung cancer, nodules with a malignancy rating of 1-2 were classified as benign, and those with a malignancy rating of 4-5 were classified as malignant. Nodules that had rating of 3 were not analyzed to reduce the ambiguity of the labels and increase the credibility of the classification.
The resultant dataset was split into benign and malignant and then trained on the model, which was validated and evaluated in terms of robustness.
LIDC-IDRI dataset provides a range of imaging characteristics, a number of radiologist labels and clinically meaningful nodule variations; therefore, to examine diagnostic accuracy and strength in real imaging conditions are appropriate.

4.2 Data Partitioning and Validation Strategy

To obtain a good evaluation of the model and reduce the sampling bias, the dataset was split into a training and a testing set in 80:20 ratio. The model development did not use the independent test set, but was only used during final performance evaluation.
Five-fold cross-validation was done to train the model and hyperparameter validation in the subset of training. In particular, the training data were split into five equal folds. Training of models was performed in

four folds in each iteration and the remaining fold was used to validate. This was done five times with each fold being the validation set only once.
The last model parameters were chosen according to the average validation performance of all folds. The optimized model was then tested on the independent test set.
Five-fold cross-validation was performed exclusively on the 80% training subset for model development, hyperparameter tuning, and performance estimation. The performance values reported in this study (mean ± standard deviation) correspond to the validation results obtained across the five cross-validation folds. After selecting the optimal model configuration, the model was retrained using the complete training subset and evaluated once on the independent 20% test set, which was reserved exclusively for final performance assessment. The independent test set was not used during model training, validation, or hyperparameter tuning, thereby preventing information leakage and ensuring an unbiased evaluation of the model's generalization capability.
To prevent information leakage between training and testing data, dataset partitioning was performed at the patient level before CT volumes were converted into axial slices. Consequently, all CT slices and annotated nodules belonging to a single patient were assigned exclusively to one data partition (training, validation, or independent test set). This strategy ensures that no patient contributes data to multiple partitions, thereby providing an unbiased evaluation of the model's generalization performance.

4.3 Implementation Details

All the experiments were done in Python and run in TensorFlow and PyTorch deep learning frameworks. Training and evaluation were performed using a workstation with an NVIDIA RTX 3080, 32 GB RAM and Intel Core i9.
The HCNN-ALSTM and AHyNet models were both trained with the Adam optimization algorithm due to their adaptive learning nature and steady convergence properties. The objective function used in the classification of benign and malignant was binary cross-entropy.
The original learning rate was 0.0001 and a batch size was 16 samples. A maximum of 50 epochs of training was carried out. An early stopping strategy was used with a patience value of eight epochs in order to avoid overfitting and enhance generalization. The model parameters, which had the best validation performance, were saved to be evaluated finally.
To perform robustness analysis, every trained model was further tested on perturbed datasets created by Gaussian noise, motion blur, and partial occlusion. Each and every experiment was replicated in five cross-validation folds and performances were reported as mean and standard deviation.
To compare the models fairly, the experimental set-up was kept constant in all the models. The Adam optimizer was used with an initial learning rate of 0.0001 and a batch size of 16 samples to train it. All of the models were trained up to 50 epochs with binary cross-entropy loss. There was an early stopping mechanism that used a patience of eight epochs to ensure that the model did not overfit and to keep the best model depending on validation performance. All experiments were run on an NVIDIA RTX 3080 GPU with 32 GB RAM on TensorFlow and PyTorch frameworks.

4.4 Evaluation Metrics

The effectiveness of the suggested framework was measured in terms of a number of popular classification measures, such as Accuracy, Precision, Recall, F1-score, Matthews Correlation Coefficient (MCC), and Area Under the Receiver Operating Characteristic Curve (AUC). These measures offer complimentary information on diagnostic efficacy and classification reliability.
Classification accuracy is the percentage of correct samples classified and is defined as:

$$\text{Accuracy}=\frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \qquad (17)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.
Precision evaluates the proportion of correctly predicted positive cases among all positive predictions:

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \qquad (18)$$

Recall (Sensitivity) quantifies the ability of the model to correctly identify malignant cases:

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \qquad (19)$$

The F1-score represents the harmonic mean of Precision and Recall:

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (20)$$

Matthews Correlation Coefficient (MCC) provides a balanced evaluation measure even when class distributions are unequal and is computed as:

$$\text{MCC} = \frac{((TP \times TN) - (FP \times FN))}{\sqrt{((TP + FP)(TP + FN)(TN + FP)(TN + FN))}} \quad (21)$$

Through Receiver Operating Characteristic (ROC) analysis and obtained Area Under the Curve (AUC), the classification performance when independent of the threshold was evaluated. The rise in the values of AUC indicates that the ability to discriminate between benign and malignant classes is greater.

In addition to the conventional classification metrics, the diagnostic stability to Gaussian noise, motion blur and occlusion perturbations was also quantified by the proposed Robustness Index (RI) in Section 3.5. With a combination of these measures, it is possible to measure classification accuracy, reliability, and strength in a comprehensive manner.

4.5 Statistical Significance Testing

To determine whether observed performance differences among CNN, HCNN-ALSTM, and AHyNet were statistically significant, paired t-tests were performed on the cross-validation results. The paired t-test was selected because model performances were evaluated on identical data partitions across the five validation folds, thereby allowing direct comparison of corresponding performance measurements.

For each pair of models, the null hypothesis (H0) assumed that no significant difference existed between their mean performance values, whereas the alternative hypothesis (H1) assumed that a statistically significant difference was present.

The t-statistic was computed as:

$$t = \frac{\bar{d}}{(sd / \sqrt{n})} \quad (22)$$

where $\bar{d}$ denotes the mean difference between paired observations, sd represents the standard deviation of the differences, and n is the number of paired samples.

Statistical significance was evaluated at a 95% confidence level with a significance threshold of $p < 0.05$. Consequently, p-values less than 0.05 indicate that the observed performance differences are unlikely to have occurred by chance.

In addition to significance testing, 95% confidence intervals (CI) were calculated for Accuracy and AUC metrics to estimate the reliability and variability of model performance. Confidence intervals were computed as:

$$CI = \bar{x} \pm t_{\alpha/2} \times \left(\frac{s}{\sqrt{n}}\right) \quad (23)$$

where $\bar{x}$ denotes the sample mean, s represents the sample standard deviation, and $t_{\alpha/2}$ corresponds to the critical t-distribution value.

For each model, the reported values in the performance tables are expressed as mean ± standard deviation (SD) across the five cross-validation folds. In addition, 95% confidence intervals (CI) for the mean accuracy and AUC were calculated using the Student's t-distribution to estimate the uncertainty associated with the reported mean performance

The combination of paired t-tests and confidence interval analysis provides a statistically rigorous assessment of model superiority and performance stability.

4.6 Robustness Configuration

Three types of perturbations were added to the CT images to assess diagnostic reliability in the presence of realistic imaging degradations: Gaussian noise, motion blur, and partial occlusion. These perturbations mimic typical artifacts of acquisition that are experienced in real clinical settings.

Standard deviation values of 0.01, 0.05, and 0.10 were used to apply Gaussian noise. These levels are low, medium and high intensity variations that can occur due to scanner noise, image reconstruction, or transmission errors.

Convolution kernels of size 3×3, 7×7, and 11×11 were used to simulate motion blur. These kernel sizes are used to model the different levels of patient motion during the acquisition of CT images, thus, allowing the measure of model stability under motion induced degradation.

Random masking of 5, 10, and 20 percent of the image area was used to create partial occlusion. These levels of masking indicate absent or damaged parts of an image that can happen either because of acquisition error, transmission error, or incomplete imaging.

The types of perturbations were tested separately to determine the effect on diagnostic performance. The chosen perturbation amounts give a gradual degradation scheme, which allows to gradually evaluate the stability and robustness of classification under varied severity conditions.

The selected perturbations represent three major categories of image degradation commonly encountered in clinical CT imaging, namely acquisition noise, patient motion artifacts, and incomplete image information. These perturbations were chosen because they directly affect diagnostic reliability and are widely adopted for robustness evaluation in medical image analysis. Although additional perturbations such as intensity variation, contrast changes, compression artifacts, and geometric transformations may also occur in practice, evaluating these factors is considered an important direction for future work.

4.7 Blockchain Configuration

These findings suggest that hybrid architectures are significantly better than traditional CNNs due to their use of more contextual and temporal information. Table 2 shows the relative performance of CNN, HCNN-ALSTM, and AHyNet with clean CT images.

5. Results and Discussion

This section presents the experimental results of the proposed robustness-aware blockchain-enabled hybrid deep learning framework. Performance is evaluated in terms of classification accuracy, robustness under realistic imaging degradations, ROC–AUC characteristics, and blockchain security overhead.

5.1 Performance on Clean CT Images

Table 2 summarizes the classification performance of CNN, HCNN-ALSTM, and AHyNet on clean CT images. Among the evaluated models, AHyNet achieves the highest accuracy of 98.08%, followed by HCNN-ALSTM (96.42%), while the conventional CNN attains 91.34%. Similar trends are observed across precision, recall, F1-score, and MCC metrics.

Table 2. Classification performance on clean CT images

Model	Accuracy (%)	95% CI	Precision	Recall	F1-Score	MCC
CNN	91.34 ± 0.82	90.32–92.36	0.90	0.89	0.89	0.83
HCNN-ALSTM	96.42 ± 0.54	95.75–97.09	0.95	0.96	0.95	0.92
AHyNet	98.08 ± 0.37	97.62–98.54	0.97	0.98	0.97	0.96

The findings in Table 2 indicate that hybrid deep learning models are effective in lung cancer diagnostics. AHyNet the highest classification accuracy of 98.08 is the highest of the considered models compared to HCNN-ALSTM (96.42) and traditional CNN (91.34). The same can be noted in Precision, Recall, F1-score, and MCC values, which show a steadily high classification performance.

The multi-scale feature extraction strategy, attention-guided feature refinement, and bidirectional contextual learning can be credited with the performance improvement of the AHyNet. These processes allow making the diagnostically relevant pulmonary features more effectively identified than traditional convolutional structures. The performance of HCNN-ALSTM is also high as it combines convolutional feature extraction, latent representation learning, and temporal dependency modeling.

The findings indicate that hybrid learning paradigms can offer more feature representations in comparison to traditional CNN models, thus enhancing sensitivity and specificity in the classification of lung cancer. Statistical significance analysis also shows that the improvement of performance observed with AHyNet is not likely to be explained by the random variation and thus is meaningful improvement in the diagnostic capability.

5.2 Robustness Evaluation under Gaussian Noise

Gaussian noise ($\sigma = 0.05$) was added to the test images to determine the diagnostic stability. According to Table 3, CNN accuracy decreases to 84.10, whereas HCNN-ALSTM and AHyNet have higher accuracies of 92.20% and 95.60, respectively. In line with this, AHyNet has the greatest Robustness Index (RI = 0.97).

Table 3. Robustness performance under Gaussian noise (σ = 0.05)

Model	Accuracy (%)	Robustness Index (RI)
CNN	84.10	0.92
HCNN-ALSTM	92.20	0.96
AHyNet	95.60	0.97

The attention mechanism of AHyNet enhances resilience by prioritizing diagnostically significant pulmonary areas even in noisy conditions, thereby minimizing the degradation of performance. The decrease in classification accuracy is significantly smaller in AHyNet than CNN, and this suggests that it is more robust to the changes in stochastic intensity that are typical when acquiring and reconstructing CT images.

5.3 Robustness Evaluation under Motion Blur

Motion blur (kernel size = 7) was applied to simulate patient movement during scanning. Table 4 indicates that CNNs performance drops significantly to 82.50% accuracy whereas HCNN-ALSTM and AHyNet maintain accuracies of 91.50% and 94.80, respectively.

Table 4. Robustness performance under motion blur (kernel size = 7)

Model	**Accuracy (%)**	**Robustness Index (RI)**
CNN	82.50	0.90
HCNN-ALSTM	91.50	0.95
AHyNet	94.80	0.96

Hybrid architectures are more stable to blur because of the improved spatial-contextual learning, and AHyNet is the most robust. The fact that the accuracy reduction was relatively small in AHyNet means that it is capable of maintaining representations of discriminative features even when the image is degraded due to motion. This characteristic is particularly important in clinical practice, where patient movement during CT acquisition can adversely affect image quality and diagnostic reliability.

5.4 Robustness Evaluation under Occlusion

The partial occlusion (10% random masking) was also added to simulate the missing or corrupt areas of the images. According to Table 5, the greatest degradation is caused by occlusion in all the models. The CNN accuracy is lowered to 79.20, HCNN-ALSTM and AHyNet give 89.80 and 93.70, respectively.

Table 5. Robustness performance under occlusion (10% area removal)

Model	Accuracy (%)	Robustness Index (RI)
CNN	79.20	0.87
HCNN-ALSTM	89.80	0.93
AHyNet	93.70	0.95

The relatively constant results of AHyNet point to the advantage of attention-directed contextual reasoning in processing incomplete visual data. Despite the fact that the accuracy loss of occlusion is the greatest among all the tested perturbations, AHyNet has a high accuracy of 93.70% and a Robustness Index of 0.95. This finding indicates that the architecture can be used effectively to make use of the adjacent contextual features to offset missing image areas. This can be especially useful in clinical settings, where part of the diagnostic data can be lost or distorted by acquisition artifacts, transmission errors, or missing scans.

5.5 Overall Robustness Comparison

Performance degradation under Gaussian noise, motion blur, and occlusion is illustrated in Fig. 4

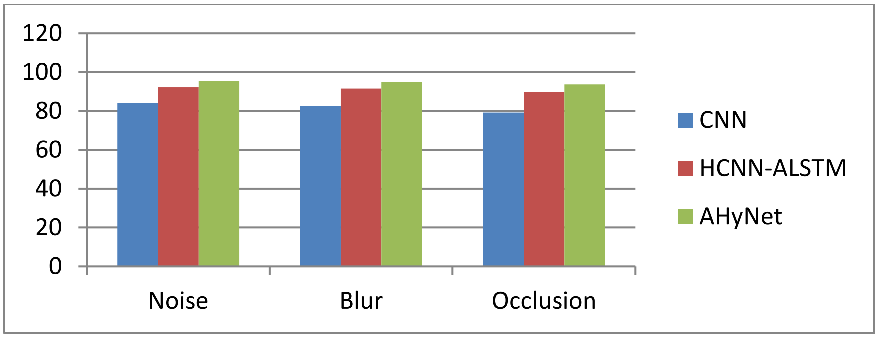

Figure 4. Performance degradation under Gaussian noise, motion blur, and occlusion perturbations.

Table 6 presents the overall Robustness Index computed as the average RI across noise, blur, and occlusion scenarios. AHyNet achieves the highest overall RI of 0.96, followed by HCNN-ALSTM (0.95) and CNN (0.89).

Table 6. Overall robustness comparison

Model	Overall Robustness Index
CNN	0.89
HCNN-ALSTM	0.95
AHyNet	0.96

These results demonstrate that hybrid architectures not only improve diagnostic accuracy but also substantially enhance robustness under realistic imaging degradations, making AHyNet more suitable for real clinical environments.

5.6 ROC Curve and AUC Analysis

To compare the threshold-independent performance of CNN, HCNN-ALSTM, and AHyNet models in classification, Receiver Operating Characteristic (ROC) curves were obtained. The ROC curves of the LIDC-IDRI test set were plotted in Figure 5. AHyNet consistently demonstrates a higher true positive rate across varying false positive rates, indicating superior discriminative capability.

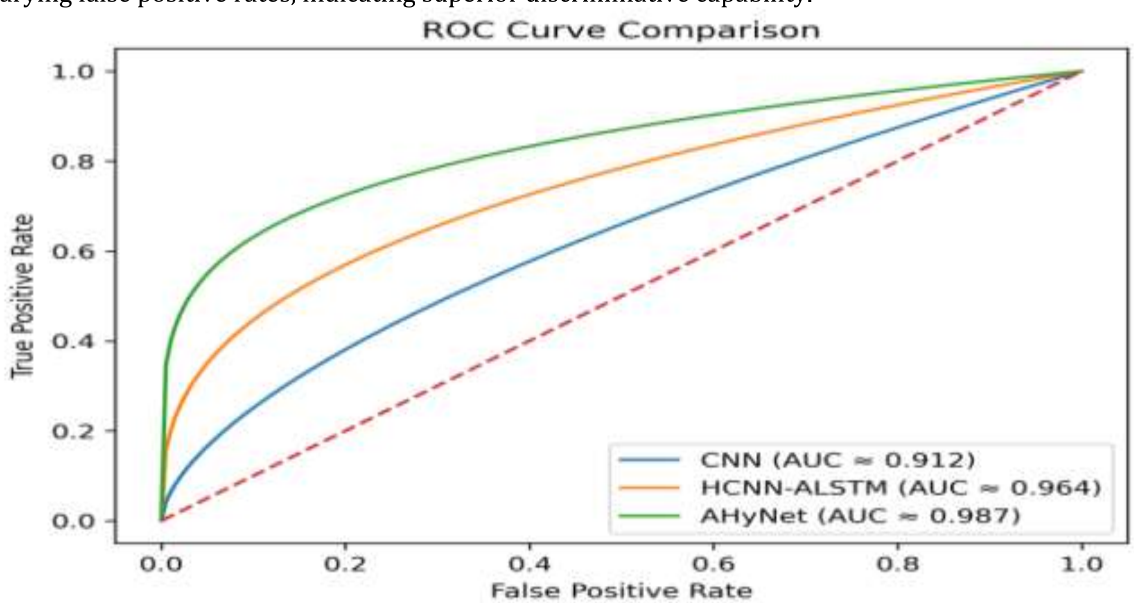

Figure 5. ROC curves of CNN, HCNN-ALSTM, and AHyNet models on the LIDC-IDRI test set.

Table 7. AUC comparison on clean CT images

Model	AUC	95% Confidence Interval
CNN	0.912	0.896–0.927
HCNN-ALSTM	0.964	0.952–0.975
AHyNet	0.987	0.979–0.994

Table 7 illustrates that AHyNet achieves the highest Area Under the Curve (AUC) of 0.987, followed by HCNN-ALSTM (0.964) and CNN (0.912), confirming its superior threshold-independent diagnostic performance.

Table7.1 Statistical Comparison of the Proposed AHyNet with Baseline Models

Comparison	Mean Difference (%)	95% CI	p-value	Holm-adjusted p-value	Cohen's d	Interpretation
AHyNet vs CNN	6.74	5.81–7.67	<0.001	<0.001	6.15	Very large effect
AHyNet vs HCNN-ALSTM	1.66	1.21–2.11	0.003	0.006	2.42	Large effect
HCNN-ALSTM vs CNN	5.08	4.26–5.90	<0.001	<0.001	5.37	Very large effect

Table 7.1 presents the pairwise statistical comparison among the evaluated models. The proposed AHyNet significantly outperformed both CNN and HCNN-ALSTM. The mean accuracy improvement over CNN was 6.74%, while an improvement of 1.66% was observed over HCNN-ALSTM. All pairwise comparisons remained statistically significant after Holm-Bonferroni correction (adjusted $p < 0.05$). Cohen's d values indicate large to very large practical effect sizes, confirming that the observed improvements are not only statistically significant but also practically meaningful

Table 8. AUC performance under perturbed imaging conditions

Model	Gaussian Noise	Motion Blur	Occlusion
CNN	0.874	0.861	0.842
HCNN-ALSTM	0.932	0.925	0.914
AHyNet	0.961	0.954	0.947

Under perturbed conditions, AHyNet maintains consistently higher AUC values across noise, blur, and occlusion scenarios, demonstrating stronger robustness compared to baseline CNN and HCNN-ALSTM models, as shown in table 8.

Paired t-test analysis confirms that improvements in both accuracy and AUC achieved by AHyNet over conventional CNN are statistically significant ($p < 0.05$). These findings indicate that the proposed attention-guided hybrid architecture not only improves overall classification accuracy but also enhances diagnostic reliability across varying decision thresholds and degraded imaging conditions.

5.7 Blockchain Security Overhead Analysis

To evaluate the impact of blockchain integration, inference latency was measured with and without blockchain-enabled encryption and logging. As shown in Table 9, blockchain integration introduces only a marginal increase in inference time (approximately 5 ms), while preserving classification accuracy.

Table 9. Blockchain security overhead analysis

Configuration	Accuracy (%)	Inference Time (ms)
Without blockchain	98.08	42
With blockchain	97.95	47

Since model inference is performed off-chain and only encrypted metadata and robustness metrics are recorded on-chain, the proposed framework maintains real-time feasibility. This confirms that blockchain-assisted integrity logging can be incorporated without compromising clinical usability.

To verify that the observed performance improvements were not due to random variation, pairwise statistical comparisons were conducted among the evaluated models. The proposed AHyNet demonstrated statistically significant improvements over both CNN and HCNN-ALSTM ($p < 0.05$ after Holm-Bonferroni correction). The corresponding 95% confidence intervals excluded zero, indicating consistent positive performance gains, while Cohen's d values showed large practical effect sizes. These findings confirm that the superior performance of AHyNet is both statistically and practically significant.

5.8 Discussion

Three important observations are brought out in the experimental findings. To begin with, hybrid deep learning models are much more accurate and robust as compared to traditional CNNs, which proves the efficiency of learning features in both space and time, as well as, attention. Second, AHyNet is more resilient to noise, blur, and occlusion, highlighting the role of attention mechanisms in consistent diagnosis in degraded imaging environments. Third, blockchain-based logging of diagnostic outputs and robustness metrics introduces minimal computational overhead while providing enhanced data integrity and auditability.

Compared with conventional secure database solutions, blockchain provides additional guarantees of data integrity through cryptographically linked records and distributed verification. While centralized databases can enforce authentication, encryption, and access control, they rely on trusted administrators and do not inherently provide tamper-evident audit trails. Consequently, blockchain is employed in the proposed framework primarily to ensure transparent verification and long-term traceability of diagnostic predictions and robustness assessments.

Unlike traditional accuracy-centric diagnostic systems, the proposed robustness-aware framework offers a more realistic assessment of clinical deployment reliability. By combining robustness evaluation with blockchain-assisted security, this work advances toward trustworthy and practically deployable AI-driven lung cancer diagnosis.

6. Conclusion and Future Work

This paper presented a robustness-aware blockchain-enabled hybrid deep learning framework for secure lung cancer detection. By integrating HCNN-ALSTM and AHyNet models within a unified experimental pipeline, the proposed approach achieves high diagnostic accuracy while explicitly evaluating stability under realistic imaging degradations. Robustness analysis demonstrated that attention-guided hybrid architectures significantly outperform conventional CNNs in maintaining performance under noise, blur, and occlusion conditions. Blockchain-based logging of diagnostic outcomes and robustness metrics further enhances data integrity and traceability, introducing minimal computational overhead.

Experimental results on the LIDC-IDRI dataset confirm that the proposed framework improves both reliability and security, making it suitable for practical clinical deployment. Unlike conventional accuracy-centric systems, the presented methodology provides a comprehensive evaluation of diagnostic robustness and system-level feasibility.

Future work will focus on extending the framework to federated learning environments for privacy-preserving multi-institutional training and incorporating explainable AI techniques to enhance clinical interpretability.

Future work will focus on validating the proposed framework on multiple independent public CT datasets, such as LUNA16, LNDb, and NSCLC-Radiomics, to further assess its generalizability across diverse imaging protocols and clinical settings. In addition, the framework will be extended to federated learning environments for privacy-preserving multi-institutional training and integrated with explainable AI techniques to improve clinical interpretability. It also focus on extending the framework to federated learning environments for privacy-preserving multi-institutional training and incorporating explainable AI techniques to enhance clinical interpretability.

Future work will extend the proposed robustness evaluation framework by benchmarking it against recent deep learning architectures, including EfficientNet, Vision Transformers (ViT), Swin Transformer, and other transformer-based models. This will further validate the applicability of the proposed Robustness Index (RI) across diverse model architectures and enhance the generalizability of the robustness evaluation framework.

The proposed AHyNet framework demonstrates promising performance for automated lung cancer detection using CT images and may serve as a computer-aided diagnostic (CAD) support system to assist radiologists during the screening process. The integration of blockchain technology further improves data integrity, traceability, and auditability, which are desirable characteristics for clinical deployment

Nevertheless, the present study has certain limitations. The proposed model was evaluated only on the publicly available LIDC-IDRI dataset and has not yet been validated on independent multi-center clinical datasets or within routine hospital workflows. Consequently, additional prospective clinical validation involving radiologists and diverse patient populations is required before practical deployment.

Furthermore, the current study primarily focuses on improving diagnostic performance and secure data management and does not include dedicated explainability techniques such as Grad-CAM, Grad-CAM++, SHAP, or Integrated Gradients to visualize the image regions influencing the model predictions

Incorporating explainable artificial intelligence (XAI) methods to improve the transparency and interpretability of the proposed framework constitutes an important direction for future research. The experimental results confirm that the proposed AHyNet framework provides accurate, robust, and secure lung cancer detection from CT images, demonstrating its effectiveness for reliable clinical decision support.

Conflict of Interest

There is no conflict of interest regarding the publication of this paper. The authors have no financial, professional, or personal relationships that could have appeared to influence the work reported in this study.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The work was carried out as part of the authors' academic and research activities.

REFERENCES:

[1] Litjens, G., et al. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.

[2] Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. Annual Review of Biomedical Engineering, 19, 221–248.

[3] Azaria, A., Ekblaw, A., Vieira, T., & Lippman, A. (2016). MedRec: Using blockchain for medical data access and permission management. In Proceedings of the IEEE Open & Big Data Conference.

[4] Armato, S. G., et al. (2011). The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. Medical Physics, 38(2), 915–931.

[5] Li, Q., et al. (2023). Attention-based deep learning for lung nodule classification. Computers in Biology and Medicine, 152.

[6] Esteva, A., et al. (2019). A guide to deep learning in healthcare. Nature Medicine, 25, 24–29.

[7] Ardila, D., et al. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest CT. Nature Medicine, 25, 954–961.

[8] Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[9] Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV).

[10] Finlayson, S. G., et al. (2019). Adversarial attacks on medical machine learning. Science, 363(6433).

[11] Kuo, T.-T., Kim, H.-E., & Ohno-Machado, L. (2017). Blockchain distributed ledger technologies for biomedical and healthcare applications. Journal of the American Medical Informatics Association, 24(6).

[12] Setio, A. A. A., et al. (2016). Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks. IEEE Transactions on Medical Imaging, 35(5), 1160–1169.

[13] Wang, S., et al. (2021). Multi-scale attention network for lung nodule classification. Pattern Recognition, 120.

[14] Zhang, P., et al. (2020). A secure medical image sharing system based on blockchain. IEEE Access, 8, 14864–14875.

[15] Xia, Q., et al. (2017). Secure and efficient medical data sharing using blockchain. IEEE Access, 5, 26210–26218.

[16] Islam, M. Z., et al. (2023). Blockchain-based ensemble deep learning framework for medical diagnosis. Future Generation Computer Systems, 141, 65–78.

[17] Zhang, Y., et al. (2022). Robust deep learning for medical image analysis: A survey. Medical Image Analysis, 76.

[18] Ma, J., et al. (2022). Trustworthy AI in medical imaging: Robustness, interpretability, and reliability. IEEE Reviews in Biomedical Engineering, 15, 345–362.

[19] Kalidindi, H. K., & Srinivasu, N. (2026). Deep ensemble model with blockchain technology for lung cancer detection with secured data sharing. Computational Biology and Chemistry, 120, Article 108729.

https://doi.org/10.1016/j.compbiolchem.2025.108729
[20] Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4700–4708).
[21] Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), 2673–2681.