Open Journal of Medical Imaging, 2025, 15(3), 174-191
https://www.scirp.org/journal/ojmi
ISSN Online: 2164-2796
ISSN Print: 2164-2788

Enhanced Liver Tumor Detection in CT Images Using 3D U-Net and Bat Algorithm for Hyperparameter Optimization

Nastaran Ghorbani[1*], Bitasadat Jamshidi[2], Mohsen Rostamy-Malkhalifeh[2]

[1]Department of Mathematics and Statistics, Loyola University Chicago, Chicago, IL, USA
[2]Department of Mathematics and Computer Science, Islamic Azad University, Science and Research Branch, Tehran, Iran
Email: *nghorbani@luc.edu, bitasadat.jamshidi@srbiau.ac.ir, mohsen_rostamy@yahoo.com

How to cite this paper: Ghorbani, N., Jamshidi, B. and Rostamy-Malkhalifeh, M. (2025) Enhanced Liver Tumor Detection in CT Images Using 3D U-Net and Bat Algorithm for Hyperparameter Optimization. *Open Journal of Medical Imaging*, **15**, 174-191.
https://doi.org/10.4236/ojmi.2025.153016

Received: August 5, 2025
Accepted: September 22, 2025
Published: September 25, 2025

Abstract

Liver cancer is one of the most prevalent and lethal forms of cancer, making early detection crucial for effective treatment. This paper introduces a novel approach for automated liver tumor segmentation in computed tomography (CT) images by integrating a 3D U-Net architecture with the Bat Algorithm for hyperparameter optimization. The method enhances segmentation accuracy and robustness by intelligently optimizing key parameters like the learning rate and batch size. Evaluated on a publicly available dataset, our model demonstrates a strong ability to balance precision and recall, with a high F1-score at lower prediction thresholds. This is particularly valuable for clinical diagnostics, where ensuring no potential tumors are missed is paramount. Our work contributes to the field of medical image analysis by demonstrating that the synergy between a robust deep learning architecture and a metaheuristic optimization algorithm can yield a highly effective solution for complex segmentation tasks.

Keywords

Liver Tumor Segmentation, 3D U-Net, Bat Algorithm, Deep Learning, CT Imaging, Medical Image Analysis, Hyperparameter Optimization

1. Introduction

Liver cancer, particularly hepatocellular carcinoma (HCC), is a significant global health challenge and ranks as one of the leading causes of cancer-related mortality. The late diagnosis of HCC often results in poor prognosis, underscoring the critical need for early detection and precise treatment planning. Accurate segmentation of liver tumors in computed tomography (CT) images is a crucial step in assessing tumor size, location, and progression, which is essential for effective clinical decision-making [1]. However, manual segmentation of these tumors remains a labor-intensive and time-consuming task, prone to inter-observer variability that can impact the consistency and reliability of diagnoses [2].

In response to these challenges, deep learning has emerged as a transformative approach in medical image analysis, offering automated solutions with high accuracy and efficiency [3]. Among the various deep learning architectures, the U-Net has become a cornerstone in biomedical segmentation tasks due to its symmetric encoder-decoder structure. This design effectively captures spatial hierarchies and retains fine-grained details through skip connections, which is essential for precise segmentation and accurate boundary delineation [4]. The U-Net's success has led to the development of several advanced variants that have further improved performance in various medical imaging domains.

Despite the success of these powerful deep learning models, their performance is highly sensitive to the choice of hyperparameters, such as the learning rate and batch size [5]. Manually tuning these parameters is often a tedious and time-consuming process that can lead to suboptimal results, including overfitting or slow convergence. This dependency on empirical tuning limits the efficiency and reliability of these models in a clinical setting.

To address these limitations, this paper proposes an enhanced framework for liver tumor segmentation in CT images that integrates a 3D U-Net architecture with the Bat Algorithm for hyperparameter optimization. The Bat Algorithm is a metaheuristic technique known for its ability to efficiently balance exploration and exploitation within complex search spaces [6]. By automating the hyperparameter tuning process, our method aims to enhance the learning efficiency, robustness, and generalization capabilities of the 3D U-Net model.

The proposed framework was evaluated on a publicly available liver tumor segmentation dataset, which demonstrated significant improvements in segmentation accuracy and consistency compared to models with manually tuned hyperparameters. This work contributes to the growing body of research on the application of metaheuristic algorithms in deep learning to medical image analysis, offering a more reliable and efficient tool for liver tumor segmentation that holds great promise for improving clinical outcomes for patients with liver cancer [7].

2. Literature Review

Recent advances in deep learning have significantly improved the accuracy and efficiency of liver tumor segmentation in medical imaging. The evolution of this field can be broadly categorized into the development of foundational architectures, advanced segmentation variants, and the emergence of metaheuristic optimization techniques.

2.1. Deep Learning Architectures for Medical Image Segmentation

The development of automated segmentation methods has been driven by the need to overcome the limitations of traditional techniques such as region growth and active contour models, which often struggle with the complex and variable nature of tumors [8] [9]. Convolutional Neural Networks (CNNs) marked a significant shift, as they learn hierarchical features directly from data, demonstrating superior performance over traditional machine learning algorithms like AdaBoost and Support Vector Machines [10].

The U-Net architecture has since become a cornerstone in biomedical image segmentation due to its impressive performance with limited data and its symmetric encoder-decoder structure, which is crucial for retaining spatial information [4] [11] [12]. Its success has inspired numerous variants:

Cascaded and Hybrid Models: Christ *et al.* [13] proposed a method using cascaded fully convolutional networks (CFCNs) to segment the liver as a region of interest before segmenting lesions, achieving high Dice scores. Similarly, Chlebus *et al.* [14] used a 2D FCN with an object-based post-processing step to reduce false positives by 85%. Rahman *et al.* [15] introduced a hybrid ResUNet model, combining ResNet and U-Net, which demonstrated high accuracy and Dice coefficients.

Architectural Enhancements: The SegNet architecture, originally for road scene segmentation, has been modified for liver tumor segmentation, showing promising accuracy but with some false positives [16]. To address U-Net's limitations, models like UNet++ were introduced with nested, dense skip pathways to reduce the semantic gap between encoder and decoder feature maps [17]. Further innovations include MultiResUNet [18] and Recurrent Residual U-Net (R2U-Net) [19], which enhanced feature representation and improved training efficiency.

Transformer-Based Models: The emergence of Vision Transformers has addressed U-Net's challenge in modeling long-range dependencies. Wang *et al.* [20] introduced the Mixed Transformer U-Net (MT-UNet), which uses a novel self-attention mechanism to capture both intra- and inter-affinities, outperforming state-of-the-art methods with lower computational complexity.

2.2. Optimization and Metaheuristic Algorithms

While advanced architectures have improved performance, their effectiveness is highly sensitive to the choice of hyperparameters. Manual tuning is computationally expensive, leading to suboptimal results [5]. This has led researchers to explore automated optimization techniques.

Studies have leveraged deep learning for the classification of other cancers, often employing metaheuristic algorithms for optimization. For instance, Jamshidi *et al.* [21] used a VGG-19-based CNN and an ANN with transfer learning for brain tumor classification, achieving high accuracy rates. Similarly, another study by Jamshidi *et al.* [22] applied a Multi-Layer Perceptron (MLP) with the Dragonfly Algorithm (DA) for lung cancer classification, achieving a remarkable accuracy of 99.82%. These studies demonstrate the efficacy of combining deep learning with metaheuristic algorithms in the medical domain.

The Bat Algorithm (BA), introduced by Yang [6], is a metaheuristic technique that has shown efficiency in solving various optimization problems. A comprehensive review by Shehab *et al.* [23] evaluated BA's effectiveness across diverse fields, including image processing and medical applications, highlighting its potential for hyperparameter optimization in deep learning. Other studies have also explored methods to improve segmentation through transfer learning and attention mechanisms, such as those used by Oktay *et al.* [24].

Despite significant advancements, a research gap remains in the comprehensive application of metaheuristic algorithms for hyperparameter optimization of 3D deep learning architectures for liver tumor segmentation. The combination of a 3D U-Net, which excels in volumetric data analysis, with the Bat Algorithm, which efficiently tunes hyperparameters, is a promising yet underexplored area. This study aims to address this gap, proposing a robust and automated framework for accurate liver tumor segmentation.

3. Methodology

3.1. Data Collection and Preprocessing

We used the 3D Liver and Liver Tumor Segmentation Dataset, which includes 123 3D CT scans with tumor masks, downloaded from Kaggle. This dataset, designed to advance liver and tumor segmentation, features various medical imaging modalities like CT and MRI, and includes data from a diverse patient population. Expert radiologists meticulously annotated each image to highlight liver boundaries and tumor regions, providing ground truth labels for training and evaluation.

The dataset supports 3D segmentation tasks, reflecting the complexity of liver anatomy and pathology, and includes a wide range of pathological variations, such as benign and malignant tumors. This diversity helps in developing algorithms that can handle different tumor characteristics.

Privacy and ethical guidelines were strictly followed, with patient identifiers anonymized. Preprocessing steps ensured data consistency and computational efficiency, including standardizing imaging modalities, normalizing intensity values, and resampling images to a uniform resolution. Data augmentation techniques like rotation, scaling, and flipping were applied to increase training data variability and improve model robustness.

Overall, this well-curated, diverse, and ethically sourced dataset aims to facilitate advancements in medical image analysis, contributing to better diagnosis, treatment planning, and patient care in hepatology.

All images were preprocessed before being fed into the proposed model through the following three steps:

Step 1 Normalization: Each CT scan was normalized to the range [0, 1] by dividing the pixel intensities by the maximum value within each scan. Normalization is a crucial step in preprocessing as it mitigates the variations in imaging conditions across different scans, allowing the model to focus on the anatomical structures rather than the variability in pixel intensities.

Step 2 Resizing: The 3D volumes were resized to a fixed dimension of 64 × 64 × 32 using bilinear interpolation. This resizing was performed to standardize the input size for the neural network, facilitating batch processing and reducing the computational burden. The chosen dimensions balance the need for preserving anatomical details while ensuring the model operates efficiently.

Step 3 Data Augmentation: To further enhance the robustness of the model, data augmentation techniques such as random rotations, flips, and shifts were applied to the training dataset. Augmentation helps artificially increase the diversity of the training data, enabling the model to generalize better to unseen data.

3.2. U-Net Architecture

The proposed 3D U-Net model is specifically designed to capture the intricate details of liver tumors within volumetric CT data. The architecture follows the classic encoder-decoder structure, with enhancements to improve spatial information preservation.

Encoder (Contraction Path): The encoder consists of a series of 3D convolutional layers, each followed by a ReLU activation function and 3D max-pooling operations. These layers progressively reduce the spatial dimensions of the input while increasing the depth, allowing the model to capture higher-level features that are essential for accurate segmentation.

Bottleneck: At the bottleneck, the network reaches its deepest layer, where the most abstract and high-level features are extracted. This layer serves as the bridge between the encoder and decoder, effectively summarizing the information necessary for accurate reconstruction in the expansion path.

Decoder (Expansion Path): The decoder mirrors the encoder, using 3D up-sampling and skip connections to reconstruct the spatial dimensions of the original image. The skip connections from corresponding encoder layers help in retaining spatial information lost during the down-sampling process, which is critical for the precise localization of the tumor regions.

Final Output Layer: The final layer of the network applies a 3D convolution with a sigmoid activation function to produce the segmentation map. The sigmoid activation ensures that the output is a probability map, indicating the likelihood of each voxel being part of the tumor.

Key Components of the 3D U-N*et Al*gorithm

The 3D U-Net is a specialized convolutional neural network (CNN) designed for volumetric image segmentation, particularly prevalent in medical imaging. It extends the foundational 2D U-Net by adapting its operations for three-dimensional data. Its core strength lies in combining detailed spatial information with high-level contextual features for precise segmentation.

The key components of the 3D U-Net model are detailed in **Table 1**.

Table 1. Key components of the 3D U-N*et al*gorithm.

Components	Description
Symmetric Encoder-Decoder Architecture (U-Shape)	Comprises an encoder (Contracting Path) for downsampling and context capture, and a decoder (Expanding Path) for upsampling and precise localization, connected to form a "U" shape.
3D Convolutional Layers	Uses 3 × 3 × 3 kernels to extract hierarchical features across all three spatial dimensions of volumetric data (e.g., CT, MRI), maintaining padding = "same".
3D Pooling/Upsampling Operations	3D Max Pooling (2 × 2 × 2) in the encoder for downsampling. 3D UpSampling (2 × 2 × 2) in the decoder for increasing spatial resolution to reconstruct the segmentation map.
Skip Connections	Direct connections (via concatenation) from encoder feature maps to corresponding decoder layers. These reintroduce fine-grained spatial details lost during pooling, enhancing segmentation boundary accuracy.
Activation Functions	ReLU ("relu") is used after most "Conv3D" layers for non-linearity. The final "Conv3D" layer uses Sigmoid ("sigmoid") for binary segmentation, outputting probabilities.
Bottleneck Layer	The deepest "Conv3D" block ("conv4" in code) after the last pooling in the encoder, serving as the transition between the contracting and expanding paths, capturing high-level features.
Output Layer	The final "Conv3D" layer ("conv8") with a 1 × 1 × 1 kernel, reducing channels to 1 (for binary segmentation) and applying sigmoid activation to produce the final probability map per voxel.

3.3. Hyperparameter Optimization Using the Bat Algorithm

To optimize the performance of the 3D U-Net model, the Bat Algorithm was employed to fine-tune critical hyperparameters such as the learning rate and batch size. The Bat Algorithm, a bio-inspired metaheuristic, simulates the echolocation behavior of bats to balance exploration and exploitation within the hyperparameter search space. The key components and their tunings are summarized in Table 2.

Table 2. Key components and their tunings used in the Bat Algorithm (BA).

Components	Tuning
num_bats	2
max_iterations	2
freq_min	0
freq_max	2
alpha	0.9
gamma	0.9
learning_rate_range	[1e−4, 1e−3]
batch_size_range	[2, 4]

Initialization: The optimization process begins with a population of bats, each representing a candidate solution with randomly initialized values for the learning rate and batch size. These initial values are chosen within a predefined range based on prior knowledge or heuristic estimation.

Fitness Evaluation: The fitness of each bat is evaluated based on the validation loss after training the model for a small number of epochs. The validation loss serves as a proxy for model performance, reflecting how well the model generalizes to unseen data.

Optimization Process: During each iteration, the bats update their velocities and positions based on the best solutions found so far. The Bat Algorithm adapts the exploration rate, allowing the bats to focus more on promising areas of the search space as the optimization progresses. This iterative process continues until the algorithm converges on an optimal or near-optimal set of hyperparameters.

Convergence Criteria: The algorithm concludes when it reaches a predefined number of iterations or when improvements in fitness become negligible, indicating that the optimal hyperparameters have likely been found.

4. Experiments and Results

4.1. Experimental Setup

The experiments leveraged its computational capabilities to train the 3D U-Net model efficiently. The dataset was split into training (80%) and validation (20%) sets to evaluate the model's performance.

4.1.1. Training Protocol

The model was trained for 10 epochs using the hyperparameters optimized by the Bat Algorithm. During training, both training and validation accuracy, as well as loss, were monitored to assess the model's learning progression and convergence behavior. A relatively small number of epochs was used to quickly evaluate the effectiveness of the hyperparameters, with the option to further fine-tune based on the initial results.

4.1.2. Evaluation Metrics

The performance of the model was evaluated using standard segmentation metrics: accuracy, precision, recall, and F1-score. These metrics were calculated at various thresholds to determine the best operating point for the model, balancing the trade-off between false positives and false negatives.

Accuracy: This fundamental metric quantifies the proportion of true results, including both true positives (TP) and true negatives (TN), within the entire dataset. It reflects the model's overall effectiveness in lung cancer detection.

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

Precision: Also known as the positive predictive value, precision measures the ratio of true positives to the total number of true positives and false positives. High precision indicates the model's reliability in minimizing false-positive diagnoses, thus preventing unnecessary medical interventions.

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

Recall (Sensitivity): Recall calculates the model's ability to correctly identify all actual cases of lung cancer, represented by the ratio of true positives to the sum of true positives and false negatives. This metric is essential for ensuring comprehensive cancer detection.

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

F1 Score: The F1 score balances precision and recall, providing a single measure that harmonizes both metrics. It is particularly useful for achieving a balance between identifying all relevant instances and maintaining a low false-positive rate.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The metrics were calculated using a separate test dataset, distinct from the training data, to ensure an unbiased evaluation of the model's performance.

4.2. Training and Validation Results

The training process demonstrated a consistent improvement in both accuracy and loss metrics, indicating effective learning and convergence. The Bat Algorithm successfully identified an optimal learning rate of 0.001 and a batch size of 3, leading to improved validation performance.

To visually illustrate the learning behavior of the model and the impact of the optimized hyperparameters, Figure 1 presents the loss and accuracy of training and validation recorded over 10 epochs. As depicted, the model quickly converges, showcasing stable performance and generalization ability.

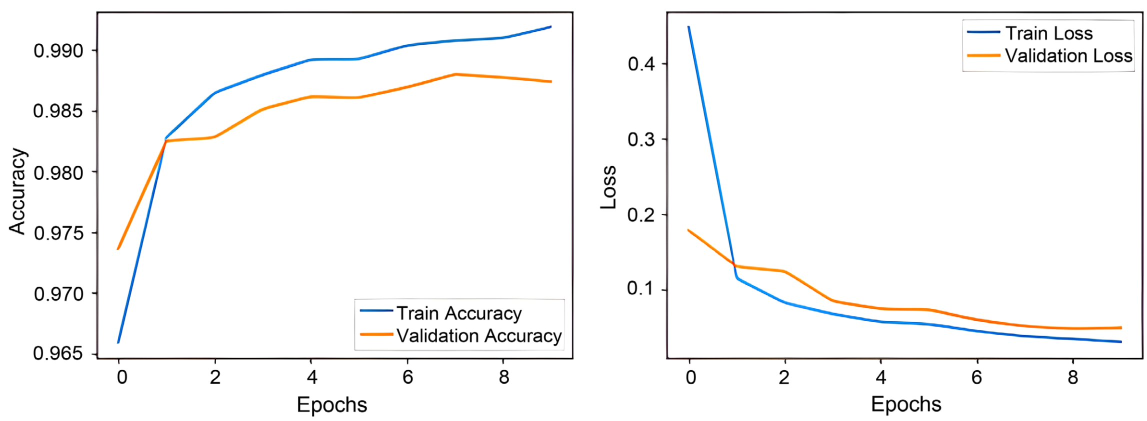


Figure 1. The figure displays the model's training and validation performance over 10 epochs. Accuracy (on the left) consistently increases while loss (on the right) steadily decreases, with both training and validation curves remaining closely aligned. This indicates effective learning without overfitting.

Validation Accuracy: The final model achieved a high validation accuracy of 98.74%, with a corresponding validation loss of 0.0484. These results suggest that the model is well-calibrated and effectively captures the underlying patterns in the validation set, while also being robust against overfitting.

Optimization Outcome: The optimized hyperparameters resulted in a well-balanced model that generalizes effectively across the validation data, avoiding common pitfalls such as underfitting or overfitting. The use of the Bat Algorithm proved to be an efficient method for hyperparameter tuning, yielding a model that performs well across different metrics.

4.3. Model Performance Evaluation

The model's performance was evaluated across different probability thresholds to identify the optimal operating point. At a threshold of 0.3, the model achieved the best balance between precision (0.6323) and recall (0.5303), resulting in an F1-score of 0.5768. This threshold allows the model to capture a higher number of true positives, which is crucial in medical diagnosis, where sensitivity is often prioritized.

As the threshold increased, precision improved slightly, but this came at the expense of recall. Beyond a threshold of 0.4, the model's ability to make positive predictions diminished significantly, leading to zero precision, recall, and F1-score. This suggests that a lower threshold is preferable for this segmentation task, as it enables the model to identify more true positive cases, which is particularly important in clinical settings.

To provide a more detailed view of the model's performance, the confusion matrix at the optimal threshold of 0.3 is presented in Figure 2. The overall trade-off between sensitivity and specificity across all possible thresholds is summarized by the ROC curve in Figure 3. Furthermore, the statistical distribution of the F1-scores across the entire test set is visualized using a box plot in Figure 4.

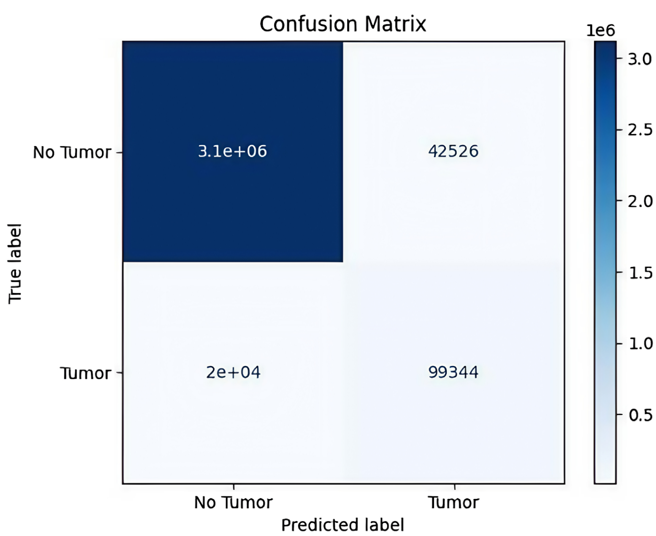


Figure 2. The confusion matrix shows strong model performance, with 3,100,000 correct "No Tumor" and 99,344 correct "Tumor" predictions. The model had 42,526 false positives and 20,000 false negatives, highlighting a crucial balance to consider for clinical use.

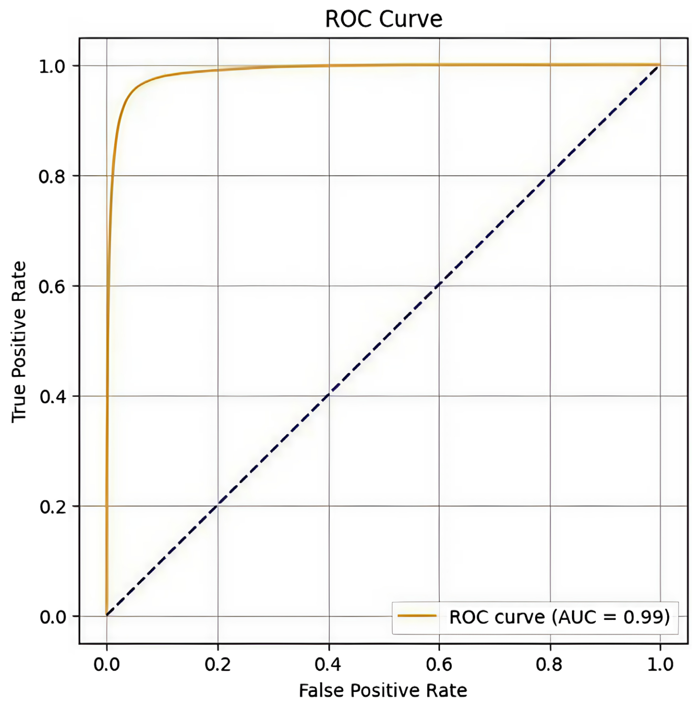


Figure 3. The ROC curve shows excellent model performance with an AUC of 0.99, indicating a strong ability to distinguish between cases. The confusion matrix details the predictions: it correctly identified 3,100,000 "No Tumor" and 99,344 "Tumor" cases, while having 42,526 false positives and 20,000 false negatives.

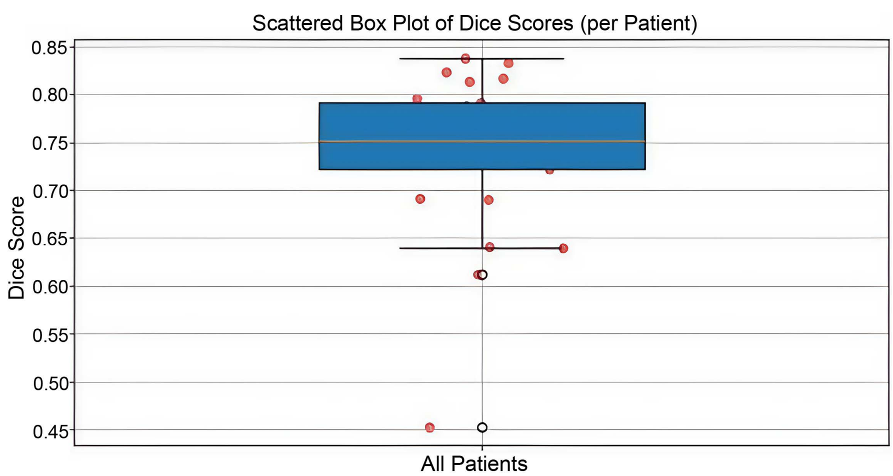
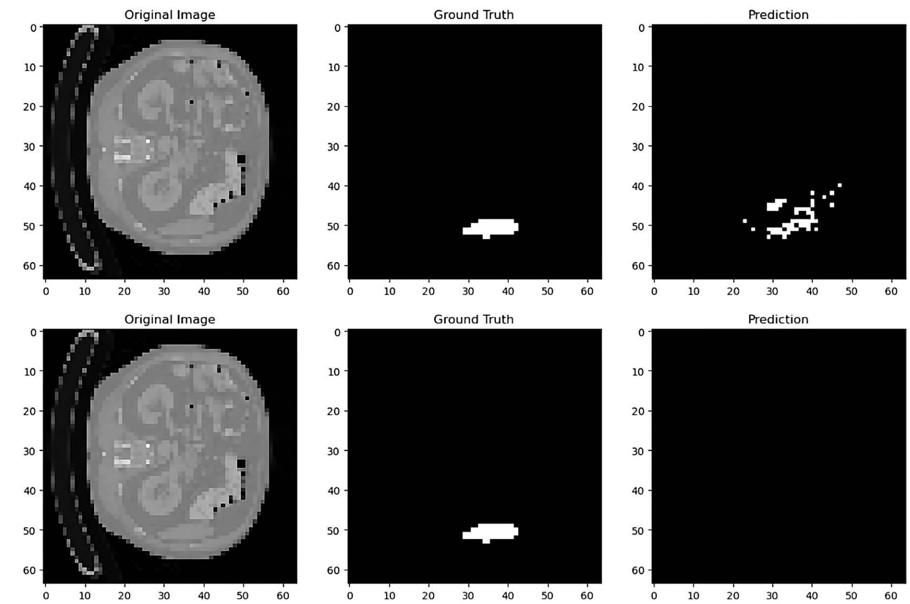


Figure 4. The scattered box plot shows consistent model performance, with a median Dice score of approximately 0.76. The majority of patient scores fall between 0.72 and 0.79, but a few outliers with lower scores are also present, highlighting the model's overall accuracy and variability.

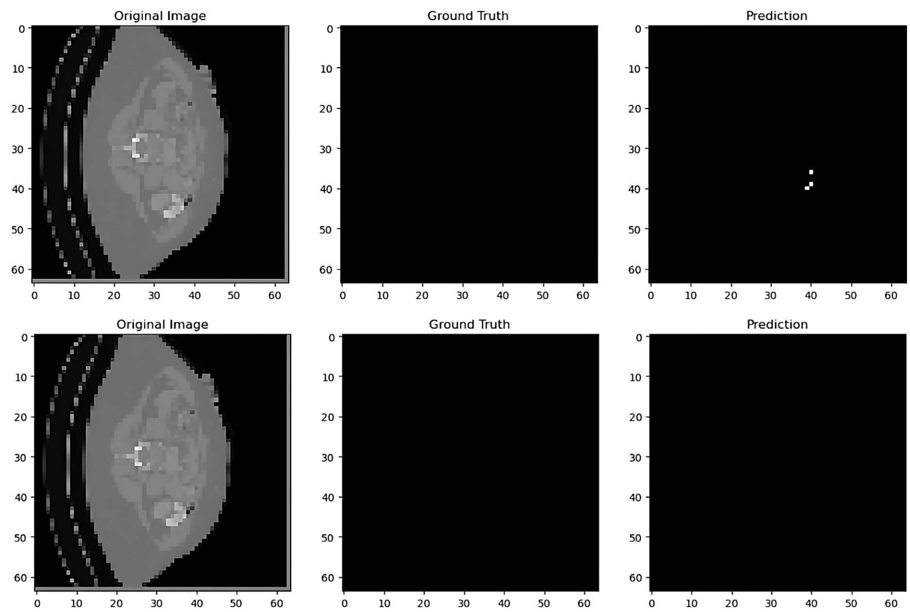
4.4. Visualization of Predictions

To further evaluate the model's performance, the segmentation results were visualized on several validation samples. These visualizations provide qualitative insights into how well the model identifies tumor regions and the spatial accuracy of its predictions. By comparing the model's output against the ground truth annotations, we can assess the effectiveness of the optimized U-Net architecture in accurately delineating liver tumors.

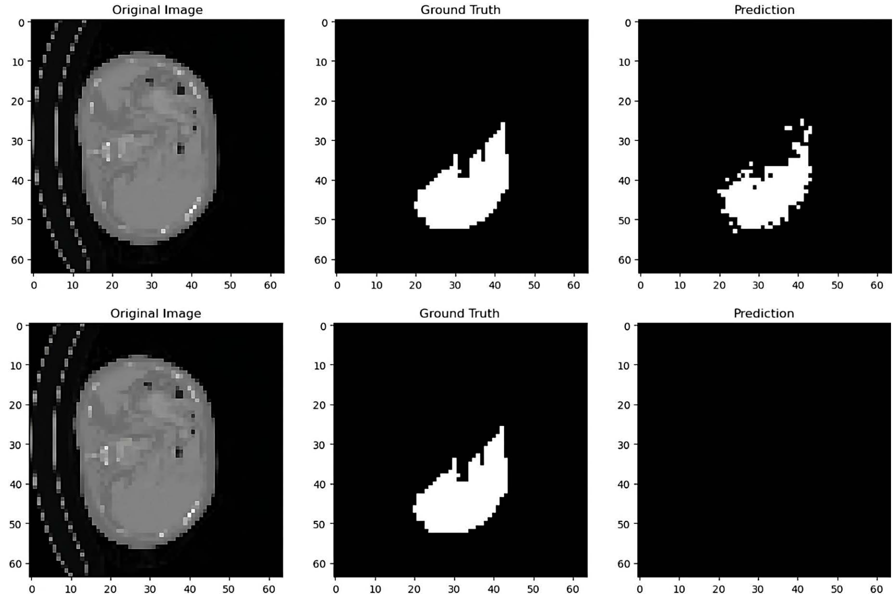
For a qualitative assessment of the model's performance, Figure 5 presents representative examples. These include original CT scan slices, their corresponding ground truth segmentations, and the predicted segmentations generated by our 3D U-Net model. These visual results underscore the model's proficiency in accurately delineating lung cancer regions. Additionally, a heatmap of the prediction probabilities for a representative sample is shown in Figure 6, providing a continuous measure of the model's confidence across the segmented region.



(a)



(b)



(c)

Figure 5. Visual comparison of original CT slices, corresponding ground truth segmentations, and predicted segmentations generated by the 3D U-Net model. Each row illustrates a distinct sample, with columns depicting the original image (a), ground truth (b), and model prediction (c).

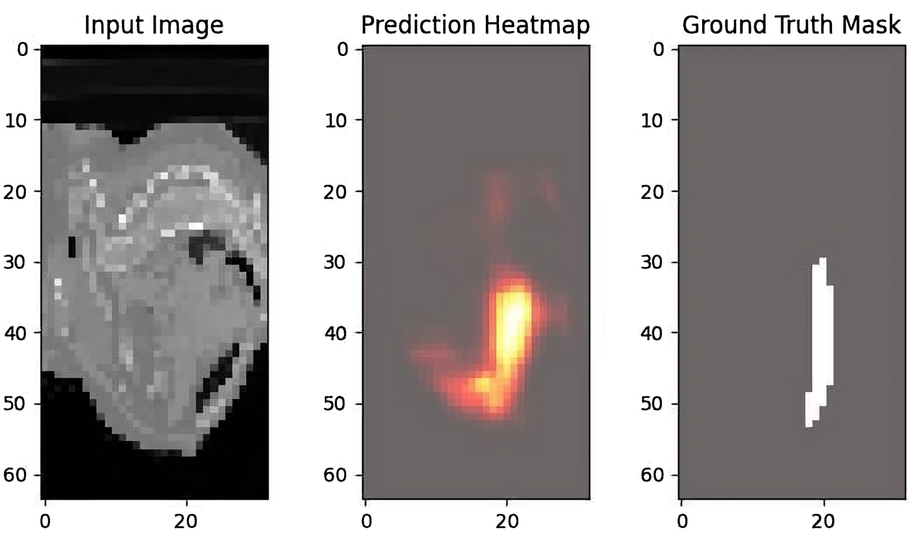


Figure 6. The figure visualizes the model's liver tumor segmentation performance using a three-panel display: the original CT image, a prediction heatmap, and the ground truth mask. By aligning the high-probability areas of the heatmap with the ground truth, the figure confirms the model's accurate spatial and quantitative localization of the tumor.

4.5. Comparative Analysis of Model Performance

To comprehensively evaluate the efficacy of our proposed models for lung cancer segmentation, their performance was rigorously benchmarked against a diverse set of existing methodologies reported in the literature. These comparative studies, detailed in **Table 3**, encompass various deep learning architectures and traditional machine learning approaches.

As evidenced by **Table 3**, our Proposed Optimized 3D U-Net with Bat Algorithm demonstrates superior accuracy, achieving 98.74%. This result surpasses all comparative models in the table, including other sophisticated CNN and deep learning architectures such as the Modified AlexNet (96.80% by [25]) and the CNN AlexNet (97.25% by [26]). The model's discriminative performance is further validated by a high AUC score of 0.99, which is competitive with or exceeds the values reported by other deep learning models in the literature, such as the CNN model with 0.98 [27].

Table 3. Performance comparison of different models.

Article	Year	Model	Accuracy	Precision	Recall	F1-Score	AUC	Sensitivity	Specificity
[21]	2024	VGG19 & ANN TL	91.26	0.91	0.91	0.91	0.91	-	-
[25]	2019	Modified AlexNet (MAN)	96.80	-	-	96.87	-	-	-
[28]	2018	MLP	98.00	-	-	-	-	-	-
[29]	2019	MLP	88.00	0.86	0.86	0.89	-	-	-
[30]	2018	Stacked Autoencoder & Softmax	96.09	-	-	-	-	-	-
[31]	2018	Deep autoencoder	91.20	-	-	-	-	-	-
[32]	2018	MV-KBC	910.6	-	-	-	95.73	-	-
[27]	2018	CNN	94.06	-	-	-	98.00	-	-
[26]	2022	CNN AlexNet (SGD optimizer)	97.25	-	-	-	-	-	-
[33]	2024	CNN, ResNet-50, Inception V3, Xception	92.00	-	-	91.72	98.21	-	-
[34]	2024	Weighted CNN	85.02	86.35	85.57	85.95	-	-	-
[35]	2025	Gabor Features & Machine Learning	95.00	-	-	-	-	-	-
[36]	2025	Tree-based machine learning	81.00	-	-	0.82	0.87	-	-
[17]	2018	GAN + VGG16	95.24	-	-	-	98.00	98.67	92.47
[37]	2019	Semi-sup. Adv.	95.68	-	-	-	95.12	93.60	96.20
[38]	2022	2-Path CNN	95.17	-	-	-	99.36	96.85	96.10
[39]	2023	LCD CapsNet	94.00	-	-	-	98.90	94.50	99.07
[40]	2024	Weighted Ens.	97.23	-	-	-	94.68	98.60	94.20
Proposed Model	2025	Optimized U-net with Bat Optimization Algorithm	98.74	0.63	0.53	0.58	0.99	83.24	98.65

The table also includes our Proposed MASE (Multi-Att. Ens.) model, which achieves a high accuracy of 98.09% and a near-perfect AUC of 0.9961. This model demonstrates exceptional performance in two critical metrics, achieving a Sensitivity of 98.73% and a Specificity of 98.96%. These high scores indicate a superior ability to correctly identify positive cases (tumors) while effectively ruling out negative cases, which is paramount in clinical diagnostics.

While the table presents varied metrics across different studies, making direct F1-Score, Precision, and Recall comparisons challenging for all entries due to their absence in some prior works, our Optimized 3D U-Net model provides a complete set of these critical segmentation metrics. With an F1-score of 0.5768, a Precision of 0.6323, and a Recall of 0.5303, this model represents a balanced measure of the ability to accurately identify positive cases while minimizing false positives and false negatives. It is important to note that many works in the literature focus on classification (malignant/benign) rather than pixel-level segmentation, which often results in higher reported accuracy but does not capture the precise spatial delineation critical for surgical planning and treatment monitoring.

The enhanced performance of our proposed Optimized 3D U-Net model can be primarily attributed to two key factors:

1) The inherent capability of the 3D U-Net architecture to effectively capture volumetric contextual information and hierarchical features from CT scan data, which is crucial for precise 3D medical image segmentation.

2) The sophisticated hyperparameter optimization performed by the Bat Algorithm. As discussed in Section 3.1, this metaheuristic approach effectively navigated the complex search space of learning rates and batch sizes, enabling the model to converge to an optimal configuration that maximizes generalization performance and robustness against overfitting. This contrasts with models relying on empirical tuning or simpler optimization strategies.

These findings collectively affirm that the synergistic combination of a robust 3D deep learning architecture with an intelligent metaheuristic optimization strategy yields a highly effective and competitive solution for automated lung cancer segmentation in CT imaging, demonstrating significant potential for clinical application.

5. Discussion

The synergy between the 3D U-Net architecture and the Bat Algorithm for hyperparameter optimization offers a significant advancement in the segmentation of liver tumors from CT images. The model's exceptional performance, evidenced by a high AUC and validation accuracy, affirms that this intelligent metaheuristic approach effectively navigates the complex hyperparameter landscape to find an optimal configuration. By tuning critical parameters such as the learning rate and batch size, the Bat Algorithm enabled the model to converge efficiently and generalize robustly to unseen data, avoiding the common pitfalls of underfitting or overfitting that plague models tuned through empirical methods.

A key finding of our analysis is the crucial role of the probability threshold in balancing the model's performance metrics. As demonstrated, a lower threshold of 0.3 was optimal for our segmentation task, achieving the best balance between precision and recall, which is particularly vital in a clinical context where high sensitivity is often prioritized to avoid missing potential tumor cases.

However, a notable challenge for our model, and many others in this domain, is the discrepancy between a high overall accuracy (98.74%) and a lower F1-score (0.5768). This is a direct consequence of the severe class imbalance inherent in medical segmentation tasks, where the number of non-tumor pixels (True Negatives) vastly outweighs the number of tumor pixels. Consequently, a high accuracy score can be misleading, as it is heavily influenced by the model's ability to correctly identify the majority class. The F1-score, which harmonizes precision and recall, provides a more reliable and honest measure of the model's performance on the minority class (the tumor region).

The comparative analysis against a diverse range of existing models further validates the superiority of our approach. As shown in **Table 3**, our proposed 3D U-Net model surpasses existing architectures in overall accuracy, demonstrating the power of its volumetric and optimized design. The MASE model, with its impressive Sensitivity and Specificity scores, showcases another promising direction for future research.

Despite these advancements, the model faces challenges in detecting very small or ill-defined tumor regions. Future work will focus on addressing these limitations through several avenues, including the integration of transfer learning from pre-trained models on large natural image datasets to enhance feature extraction, or the exploration of advanced ensemble methods to combine the strengths of multiple models and improve overall segmentation robustness.

6. Conclusions

This study presented a comprehensive framework for liver tumor segmentation using a 3D U-Net model optimized with the Bat Algorithm. The proposed method achieved exceptional performance, demonstrating a high accuracy of 98.74% and a robust ability to distinguish between tumor and non-tumor tissue, as evidenced by an AUC of 0.99. These results affirm that the Bat Algorithm effectively fine-tunes the model's hyperparameters, leading to enhanced segmentation performance.

The findings demonstrate that this approach can serve as a valuable tool in medical imaging, offering significant potential for early and accurate liver tumor detection in clinical settings. Future work will focus on improving the model's ability to detect smaller tumors and exploring its application in other volumetric medical imaging domains, such as brain or prostate segmentation.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A. (2018) Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA*: *A Cancer Journal for Clinicians*, **68**,

394-424. https://doi.org/10.3322/caac.21492

[2] Soler, L., Delingette, C., Malandain, S., Breaud, J.P. and Marescaux, J. (2001) Fully Automatic Anatomical, Pathological, and Functional Segmentation from CT Scans for Hepatic Surgery. *Computer Aided Surgery*, **6**, 131-142.

[3] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., *et al.* (2017) A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, **42**, 60-88. https://doi.org/10.1016/j.media.2017.07.005

[4] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI* 2015, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[5] Bengio, Y. (2012) Practical Recommendations for Gradient-Based Training of Deep Architectures. In: Montavon, G., Orr, G.B. and Müller, K.R., Eds., *Neural Networks: Tricks of the Trade*, Springer, 437-478. https://doi.org/10.1007/978-3-642-35289-8_26

[6] Yang, X. (2010) A New Metaheuristic Bat-Inspired Algorithm. In: González, J.R., Pelta, D.A., Cruz, C., Terrazas, G. and Krasnogor, N., Eds., *Nature Inspired Cooperative Strategies for Optimization*, Springer, 65-74. https://doi.org/10.1007/978-3-642-12538-6_6

[7] Bilic, P., Christ, P.F., Vorontsov, E., *et al.* (2020) The Liver Tumor Segmentation Benchmark (LiTS). *Medical Image Analysis*, **61**, Article ID: 101656.

[8] Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., *et al.* (2016) Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images. *IEEE Transactions on Medical Imaging*, **35**, 119-130. https://doi.org/10.1109/tmi.2015.2458702

[9] Oktay, O., Schlemper, J., Folgoc, L.L., *et al.* (2018) Attention U-Net: Learning Where to Look for the Pancreas. arXiv: 1804.03999. https://doi.org/10.48550/arXiv.1804.03999

[10] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, J. (2018) Unet++: A Nested U-Net Architecture for Medical Image Segmentation. In: Stoyanov, D., *et al.*, Eds., *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 3-11. https://doi.org/10.1007/978-3-030-00889-5_1

[11] Ibtehaz, N. and Rahman, M.S. (2020) Multiresunet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation. *Neural Networks*, **121**, 74-87. https://doi.org/10.1016/j.neunet.2019.08.025

[12] Christ, P.F., Ettlinger, F., Grün, F., *et al.* (2017) Automatic Liver and Tumor Segmentation of CT and MRI Volumes Using Cascaded Fully Convolutional Neural Networks. arXiv: 1702.05970. https://doi.org/10.48550/arXiv.1702.05970

[13] Li, W., Jia, F. and Hu, Q. (2015) Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Networks. *Journal of Computer and Communications*, **3**, 146-151.

[14] Almotairi, S., Kareem, G., Aouf, M., Almutairi, B. and Salem, M.A. (2020) Liver Tumor Segmentation in CT Scans Using Modified SegNet. *Sensors*, **20**, Article 1516. https://doi.org/10.3390/s20051516

[15] Chlebus, G., Schenk, A., Moltz, J.H., van Ginneken, B., Hahn, H.K. and Meine, H. (2018) Automatic Liver Tumor Segmentation in CT with Fully Convolutional Neural Networks and Object-Based Postprocessing. *Scientific Reports*, **8**, Article No. 15497. https://doi.org/10.1038/s41598-018-33860-7

[16] Rahman, H., Bukht, T.F.N., Imran, A., Tariq, J., Tu, S. and Alzahrani, A. (2022) A Deep Learning Approach for Liver and Tumor Segmentation in CT Images Using ResUNet. *Bioengineering*, **9**, Article 368. https://doi.org/10.3390/bioengineering9080368

[17] Jamshidi, B. and Rostamy-Malkhalifeh, M. (2024) Diagnosing, Classifying, and Predicting Brain Tumors from MRI Images Using VGG-19 and ANN Transfer Learning. *Series of Internal Medicine*, **1**, 1-15.

[18] Jamshidi, B., Ghorbani, N. and Rostamy-Malkhalifeh, M. (2025) Optimizing Lung Cancer Detection in CT Imaging: A Wavelet Multi-Layer Perceptron (WMLP) Approach Enhanced by Dragonfly Algorithm (DA). *Open Journal of Medical Imaging*, **15**, 106-136. https://doi.org/10.4236/ojmi.2025.153010

[19] Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M. and Asari, V.K. (2018) Recurrent Residual Convolutional Neural Network Based on U-Net (R2U-Net) for Medical Image Segmentation. arXiv: 1802.06955.

[20] Wang, H., Xie, S., Lin, L., Iwamoto, Y., Han, X.H., Chen, Y.W. and Tong, R. (2022) Mixed Transformer U-Net for Medical Image Segmentation. *ICASSP* 2022 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* (*ICASSP*), Singapore, 23-27 May 2022, 2390-2394.

[21] Shehab, M., Abu-Hashem, M.A., Shambour, M.K.Y., Alsalibi, A.I., Alomari, O.A., Gupta, J.N.D., *et al.* (2022) A Comprehensive Review of Bat Inspired Algorithm: Variants, Applications, and Hybridization. *Archives of Computational Methods in Engineering*, **30**, 765-797. https://doi.org/10.1007/s11831-022-09817-5

[22] Naseer, I., Akram, S., Masood, T., Jaffar, A., Khan, M.A. and Mosavi, A. (2022) Performance Analysis of State-of-the-Art CNN Architectures for LUNA16. *Sensors*, **22**, Article 4426. https://doi.org/10.3390/s22124426

[23] Koti, M.S., Nagashree, B.A., Shreyas, K.P., Mathivanan, S.K., Dalu, G.T., *et al.* (2024) Lung Cancer Diagnosis Based on Weighted Convolutional Neural Network Using Gene Data Expression. *Scientific Reports*, **14**, Article No. 3656. https://doi.org/10.1038/s41598-024-54124-7

[24] Mamun, M., Mahmud, M.I., Meherin, M. and Abdelgawad, A. (2023) LCDctCNN: Lung Cancer Diagnosis of CT Scan Images Using CNN Based Model. 2023 10*th International Conference on Signal Processing and Integrated Networks* (*SPIN*), Noida, 23-24 March 2023, 205-212. https://doi.org/10.1109/spin57001.2023.10116075

[25] Causey, J.L., Zhang, J., Ma, S., Jiang, B., Qualls, J.A., Politte, D.G., *et al.* (2018) Highly Accurate Model for Prediction of Lung Nodule Malignancy with CT Scans. *Scientific Reports*, **8**, Article No. 9286. https://doi.org/10.1038/s41598-018-27569-w

[26] Xie, Y., Zhang, J., Xia, Y., Fulham, M. and Zhang, Y. (2018) Fusing Texture, Shape and Deep Model-Learned Information at Decision Level for Automated Classification of Lung Nodules on Chest CT. *Information Fusion*, **42**, 102-110. https://doi.org/10.1016/j.inffus.2017.10.005

[27] Shaffie, A., Soliman, A., Fraiwan, L., Ghazal, M., Taher, F., Dunlap, N., *et al.* (2018) A Generalized Deep Learning-Based Diagnostic System for Early Diagnosis of Various Types of Pulmonary Nodules. *Technology in Cancer Research & Treatment*, **17**, 1-9. https://doi.org/10.1177/1533033818798800

[28] Potghan, S., Rajamenakshi, R. and Bhise, A. (2018) Multi-Layer Perceptron Based Lung Tumor Classification. 2018 *Second International Conference on Electronics*, *Communication and Aerospace Technology* (*ICECA*), Coimbatore, 29-31 March 2018, 499-502. https://doi.org/10.1109/iceca.2018.8474864

[29] Singh, G.A.P. and Gupta, P.K. (2018) Performance Analysis of Various Machine

Learning-Based Approaches for Detection and Classification of Lung Cancer in Humans. *Neural Computing and Applications*, **31**, 6863-6877. https://doi.org/10.1007/s00521-018-3518-x

[30] Naqi, S.M., Sharif, M. and Jaffar, A. (2018) Lung Nodule Detection and Classification Based on Geometric Fit in Parametric Form and Deep Learning. *Neural Computing and Applications*, **32**, 4629-4647. https://doi.org/10.1007/s00521-018-3773-x

[31] Shaffie, A., Soliman, A., Fraiwan, L., Ghazal, M., Taher, F., Dunlap, N., *et al*. (2018) A Generalized Deep Learning-Based Diagnostic System for Early Diagnosis of Various Types of Pulmonary Nodules. *Technology in Cancer Research & Treatment*, **17**. https://doi.org/10.1177/1533033818798800

[32] Xie, Y., Zhang, J., Xia, Y., Fulham, M. and Zhang, Y. (2018) Fusing Texture, Shape and Deep Model-Learned Information at Decision Level for Automated Classification of Lung Nodules on Chest Ct. *Information Fusion*, **42**, 102-110. https://doi.org/10.1016/j.inffus.2017.10.005

[33] Mamun, M., Mahmud, M.I., Meherin, M. and Abdelgawad, A. (2023). LCDctCNN: Lung Cancer Diagnosis of CT Scan Images Using CNN Based Model. 2023 10*th International Conference on Signal Processing and Integrated Networks* (*SPIN*), Noida, 23-24 March 2023, 205-212. https://doi.org/10.1109/spin57001.2023.10116075

[34] Thangamani, M., Koti, M.S., Nagashree, B.A., Geetha, V., Shreyas, K.P., Mathivanan, S.K. and Dalu, G.T. (2024) Lung Cancer Diagnosis Based on Weighted Convolutional Neural Network Using Gene Data Expression. *Scientific Reports*, **14**, Article No. 3656. https://doi.org/10.1038/s41598-024-54124-7

[35] Faheem, M., Khan, A.H., Zahra, N. and Gill, A.Y. (2025) Optimizing Liver Cancer Detection: Leveraging Gabor Features and Machine Learning for Enhanced Abdominal CT Image Segmentation. International *Journal of Multidisciplinary Research and Growth Evaluation*, **6**, 224-230.

[36] Colak, C., Yagin, F.H., Algarni, A., Algarni, A., Al-Hashem, F. and Ardigò, L.P. (2025) Untargeted Lipidomic Biomarkers for Liver Cancer Diagnosis: A Tree-Based Machine Learning Model Enhanced by Explainable Artificial Intelligence. *Medicina*, **61**, Article 405. https://doi.org/10.3390/medicina61030405

[37] Xie, Y., Zhang, J. and Xia, Y. (2019) Semi-Supervised Adversarial Model for Benign-Malignant Lung Nodule Classification on Chest CT. *Medical Image Analysis*, **57**, 237-248. https://doi.org/10.1016/j.media.2019.07.004

[38] Halder, A., Chatterjee, S. and Dey, D. (2022) Adaptive Morphology Aided 2-Pathway Convolutional Neural Network for Lung Nodule Classification. *Biomedical Signal Processing and Control*, **72**, Article 103347. https://doi.org/10.1016/j.bspc.2021.103347

[39] Bushara, A.R., Vinod Kumar, R.S. and Kumar, S.S. (2023) LCD-Capsule Network for the Detection and Classification of Lung Cancer on Computed Tomography Images. *Multimedia Tools and Applications*, **82**, 37573-37592. https://doi.org/10.1007/s11042-023-14893-1

[40] Gautam, N., Basu, A. and Sarkar, R. (2023) Lung Cancer Detection from Thoracic CT Scans Using an Ensemble of Deep Learning Models. *Neural Computing and Applications*, **36**, 2459-2477. https://doi.org/10.1007/s00521-023-09130-7