

ORIGINAL ARTICLE

Optimization of Neural Networks for the Diagnosis and Prediction of Lung Cancer Stages

Javad Tayebi* , Mohammadreza Rezaie

Department of Nuclear Engineering, Faculty of Modern Sciences and Technologies, Graduate University of Advanced Technology, Kerman, Iran

*Corresponding Author: Javad Tayebi

Received: 03 July 2024 / Accepted: 01 March 2025

Email: Javadtayebi.76@gmail.com

Abstract

Purpose: This study aims to optimize neural networks for diagnosing and predicting lung cancer stages, leveraging advanced recurrent neural network architectures such as Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM) networks.

Materials and Methods: The research utilizes a comprehensive dataset of clinical data and high-resolution imaging to develop novel Deep Neural Networks (DNNs). These networks are optimized using sophisticated algorithms like Stochastic Gradient Descent with Momentum (SGDM) and Adaptive Moment Estimation (Adam). Transfer Learning (TL) techniques and big data analytics are integrated to enhance model precision and reliability. Additionally, early stopping is employed to prevent overfitting, k-fold cross-validation (k=5) is used to ensure robust model evaluation, and L2 regularization is applied to improve generalization by penalizing large weights.

Results: The optimized GRU and LSTM networks demonstrate superior performance in detecting lung cancer stages, significantly improving diagnostic accuracy and prognostic predictions. The models effectively handle long-term dependencies and mitigate the vanishing gradient problem, common in simple RNNs. Using k-fold cross-validation (k=5), the optimized GRU network with SGDM achieved the following metrics in Fold 3: Test Loss of 0.0531, Test Accuracy of 0.97, Sensitivity (Recall) of 0.97, Precision of 0.98, and an F1 Score of 0.975.

Conclusion: The study's findings underscore the potential of integrating advanced neural network models into clinical practice, revolutionizing AI-driven medical diagnostics and elevating the standard of patient care.

Keywords: Deep Learning; Lung Cancer Diagnosis; Long Short-Term Memory; Gated Recurrent Unit; Optimization; Prediction.

1. Introduction

The profound transformation in medical science, particularly in disease diagnosis, is driven by the integration of Artificial Intelligence (AI) and Deep Learning (DL) [1]. These changes have occurred due to factors such as the rapid growth of computational power and the development of sophisticated algorithms. These advancements have created new opportunities to address fundamental challenges in healthcare, especially in disease diagnosis and management [2]. AI and DL technologies have significantly enhanced the capabilities of medical professionals [3]. By processing vast amounts of data with precision and speed, they can provide insights that were previously inaccessible [4]. In the context of lung cancer, AI-based tools enable earlier detection and more accurate staging, which are critical for effective treatment and improved patient outcomes [5].

1.1. Lung Cancer: A Global Health Challenge

Lung cancer is recognized as one of the leading causes of cancer-related mortality worldwide [6]. Despite significant advances in medical technology, diagnosing and treating this type of cancer remains challenging [7]. Traditional methods such as biopsies, imaging, and histopathological analysis serve as the primary means of diagnosing lung cancer, but these methods have limitations, including invasiveness, potential complications, and the need for expert interpretation [8-10]. Staging lung cancer, which determines the extent of disease spread, is crucial for developing an appropriate treatment plan. However, traditional staging techniques can be time-consuming and may not always provide the necessary details for personalized treatment strategies. This is where AI and DL can make a significant difference, offering non-invasive, rapid, and highly accurate assessments of the disease [11].

1.2. Revolutionizing Lung Cancer Diagnosis with DNNs

To enhance the diagnostic capabilities for lung cancer, this research explores Deep Neural Networks (DNNs), which act as AI structures simulating the functions of the human brain [12, 13]. These DNNs

are proficient in analyzing complex data patterns, a skill that is invaluable in examining medical images and clinical details to identify hidden signs of lung cancer [14-16]. The ability of DNNs in this research represents a significant advancement in cancer diagnosis. These models, trained on extensive and diverse clinical data sets, demonstrate an extraordinary ability to distinguish between the subtle stages of lung cancer. This precision not only improves the accuracy of disease staging but also includes the potential to predict patient survival, making it a valuable tool in shaping treatment pathways and prognostic consultations.

1.3. Importance and Capabilities of Recurrent Neural Networks (RNNs)

Long Short-Term Memory (LSTMs) networks: Due to their ability to manage long-term dependencies, LSTMs are key tools in medical diagnostics [17-19]. These networks, especially in lung cancer diagnosis, analyze sequences of medical images or clinical metrics to identify patterns indicative of disease progression [20, 21].

Gated Recurrent Units (GRUs): GRUs are designed to solve the vanishing gradient problem in traditional RNNs and, due to their simpler structure, provide faster training [22-24]. They strike a balance between computational speed and diagnostic accuracy [25].

1.4. Challenges of Neural Networks in Diagnosis

Despite the promising capabilities of neural networks in medical diagnostics, several technical challenges must be addressed to enhance their effectiveness and broader applicability. For instance, Long Short-Term Memory (LSTMs) networks and Gated Recurrent Units (GRUs) are powerful tools for managing long-term dependencies in sequential data, but they can be computationally intensive and require significant resources for training. Additionally, while LSTMs and GRUs help mitigate the vanishing gradient problem, they are not entirely immune to it, which can affect their performance in very deep networks.

Optimization algorithms such as Stochastic Gradient Descent with Momentum (SGDM) [26] and

Adaptive Moment Estimation (Adam) [27] play crucial roles in training deep neural networks. However, SGDM can sometimes lead to oscillations in the optimization process, requiring careful tuning of hyperparameters to achieve stable convergence. On the other hand, while Adam is known for its efficiency and adaptive learning rates, it can sometimes result in suboptimal solutions if not properly configured.

Moreover, the risk of overfitting remains a significant concern, as neural networks can easily memorize training data without generalizing well to new, unseen data. This necessitates the use of regularization techniques and cross-validation to ensure robust model performance. Additionally, the diversity of training datasets is crucial; models trained on homogeneous data may not perform well across different populations, highlighting the need for diverse and representative datasets.

1.5. Solutions to Overcome Challenges

To address the technical challenges of neural networks in medical diagnosis, several advanced and specialized solutions exist. One of these solutions is the use of k-fold cross-validation [28]. In this method, the data is divided into k parts, and the model is trained sequentially on k-1 parts and validated on the remaining part. This process is repeated k times, and the validation results are averaged to comprehensively evaluate the model's performance. This technique helps reduce the risk of overfitting and ensures that the model generalizes well to new data [29].

Early stopping is another effective solution. In this method, the training process is halted when the model's performance on the validation set no longer improves [30]. This technique prevents overtraining the model and helps maintain a balance between accuracy and generalization. Additionally, the use of regularization techniques such as Dropout and L2 Regularization can help reduce the risk of overfitting [31]. Dropout randomly deactivates some of the network units during training, which helps prevent the model from becoming overly dependent on specific units [32]. L2 Regularization adds a penalty to the cost function to prevent the network weights from becoming excessively large [33].

2. Materials and Methods

In this article, clinical data from patients with Non-Small Cell Lung Cancer (NSCLC) were meticulously collected. The sample selection process was based on precise criteria, including clinical stages of the disease (T, N, M), age, gender, and patient's medical history. The collected data included vital information such as histopathological stage, gender, survival, time of death, and event status in patients. These data were extracted from the database available on The Cancer Imaging Archive (TCIA) website and used for radiomic analyses. These analyses help predict treatment response and patient survival, making this dataset a valuable resource for medical research and the development of AI algorithms in medical imaging. These data have been extracted from a database available on The Cancer Imaging Archive website and have been used for radiometric analyses that assist in predicting treatment response and patient survival.

2.1. Steps for Collecting Clinical Data from NSCLC Patients

Clinical data from NSCLC patients were collected using precise protocols. These protocols included the following steps:

1- Sample Selection: Patients were selected based on specific criteria, including age, gender, clinical stage of the disease, and medical history, to ensure diversity and appropriate representation of the data.

2- Collection of Clinical Data: Clinical information, including medical history, disease stages (T, N, M), and survival status, was collected. This information allows researchers to identify clinical patterns and make accurate predictions about survival and treatment response (Table 1).

2.2. Key Clinical Variables for Neural Network Analysis in Lung Cancer Research

In advanced medical research, precise inputs for neural networks like GRU and LSTM are crucial. These networks are capable of processing a set of clinical variables that provide important information about the disease. For example, the patient's age (age) can influence the prediction of disease progression and response to treatment. The clinical stage of the tumor

Table 1. Patient Characteristics and Survival Prediction in Lung Cancer

Patient ID	age	clinical. T.Stage	Clinical. N.Stage	Clinical. M.Stage	Overall. Stage	Histology	gender	Survival. time	deadstatus. event
1	78.7515	2	3	0	IIIb	large cell	male	2165	1
2	83.8001	2	0	0	I	squamous cell carcinoma	male	155	1
3	68.1807	2	3	0	IIIb	large cell	male	256	1
4	70.8802	2	1	0	II	squamous cell carcinoma	male	141	1
5	80.4819	4	2	0	IIIb	squamous cell carcinoma	male	353	1
6	73.8864	3	1	0	IIIa	squamous cell carcinoma	male	173	1
7	81.5288	2	2	0	IIIa	squamous cell carcinoma	male	137	1
8	71.666	2	2	0	IIIa	adenocarcinoma	male	77	1
9	56.1342	2	2	0	IIIa	squamous cell carcinoma	male	131	1
10	71.0554	4	3	0	IIIb	squamous cell carcinoma	female	2119	0
11	64.3313	4	0	0	IIIb	squamous cell carcinoma	male	515	1
413	60.5315	4	2	0	IIIb	squamous cell carcinoma	female	246	1
414	50.8274	3	2	0	IIIa	adenocarcinoma	female	292	1
415	79.8712	1	0	0	I	adenocarcinoma	female	2517	1
416	69.9918	2	1	0	II	squamous cell carcinoma	male	409	1
417	65.5616	4	2	0	IIIb	squamous cell carcinoma	male	648	1
418	53.6712	2	0	0	I	adenocarcinoma	male	346	1
419	66.5096	4	1	0	IIIb	squamous cell carcinoma	male	2772	0
420	73.3808	2	1	0	II	squamous cell carcinoma	male	2429	1
421	61.7041	2	2	0	IIIa	squamous cell carcinoma	female	369	1
422	68.126	2	0	0	I	NA	female	1590	1

(Clinical.T.Stage), the clinical stage of lymph node involvement (Clinical.N.Stage), and the clinical stage of metastasis (Clinical.M.Stage) provide vital information about the severity and spread of the disease. The overall stage of the disease (Overall.Stage) and the type of tumor histology (Histology) assist in more accurate diagnosis and the selection of an appropriate treatment strategy. Finally, the patient's gender (gender) may be considered an important variable in statistical and clinical analyses.

2.3. Neural Network Architectures for Sequential Data Processing

The neural network architecture encompasses two distinct types of networks: LSTM, and GRU. Each of

these networks plays a crucial role in processing sequential data, and they are particularly adept at tasks that involve time-series analysis, natural language processing, and other applications where data points are interdependent.

LSTM networks are specifically designed to address challenges associated with training recurrent neural networks (RNNs), particularly issues related to maintaining long-term dependencies in data sequences. They incorporate mechanisms known as gates, which regulate the flow of information within the network. These gates enable the model to discern which elements of the input sequence are significant for retention or omission, thereby allowing the network to preserve relevant information over extended periods.

LSTMs are widely utilized in complex sequence modeling tasks, such as machine translation and event prediction. The following formulas illustrate the functioning of an LSTM cell, highlighting how it processes and retains information effectively [17, 34].

Forget Gate:

$$f_t = \sigma(w_f \cdot [h_t - 1, x_t] + b_f) \quad (1)$$

It plays a vital role in determining the extent to which information from the previous cell state (c_{t-1}) should be retained or forgotten. In Equation 1, (w_f) is the weight matrix for the forget gate, and (b_f) is the bias term. The combination of $[h_{t-1}, x_t]$ represents the previous hidden state (h_{t-1}) and the current input (x_t). The sigmoid activation function (σ) compresses the result between 0 and 1, where 0 means 'completely forget' and 1 means 'completely retain'. The output of the forget gate (f_t) is a vector of values between 0 and 1, which decides which information should be removed from or remain in the cell memory.

Input Gate:

$$i_t = \sigma(w_i \cdot [h_t - 1, x_t] + b_i) \quad (2)$$

The input gate (i_t), which is calculated using Equation 2, determines the amount of new information ($\hat{c}_t$) to be added to the cell state. Here, (w_i) is the weight matrix for the input gate, and (b_i) is the bias term. The rest of the equation is similar to the forget gate, but it uses a different weight matrix and bias term. The output of the input gate (i_t) is also a vector of values between 0 and 1, which decides how much of the new information should be added to the cell state, where 0 means 'adding nothing' and 1 means 'full addition'.

Output Gate:

$$o_t = \sigma(w_o \cdot [h_t - 1, x_t] + b_o) \quad (3)$$

The output gate (O_t), which is calculated using Equation 3, determines the extent to which the current cell state (c_t) is utilized to produce the current hidden state (h_t). Here, (w_o) is the weight matrix for the output gate, and (b_o) is the bias term. The rest of the equation is similar to the forget and input gates, but it employs a different weight matrix and bias term. The output of the output gate (O_t) is also a vector of values

between (0) and (1), which decides how much of the current cell state should be used in generating the hidden state, where (0) means 'no use at all' and (1) means 'full use'.

Candidate Cell State

$$\hat{c}_t = \tanh(w_c \cdot [h_t - 1, x_t] + b_c) \quad (4)$$

The candidate cell state, which is calculated using Equation 4, reveals the new information that could potentially be added to the cell state. Here, (w_c) is the weight matrix for the candidate cell state, and (b_c) is the bias term. The hyperbolic tangent activation function ($\tanh$) compresses the result between -1 and 1. The rest of the equation is similar to the gates, but it uses a different weight matrix and bias term. This part of the LSTM architecture allows the model to consider new information and add it to the cell state if necessary, which helps the LSTM to learn more effectively from the input data.

Cell State Update:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \hat{c}_t \quad (5)$$

The cell state update within an LSTM network is a crucial process that involves a careful blend of the past and the present. It starts with the previous cell state, denoted as (c_{t-1}), which holds the information up to that point. The forget gate, (f_t), applies an element-wise multiplication to this previous state, effectively deciding which information is no longer necessary and can be discarded (Equation 5). Concurrently, the input gate, (i_t), interact with the candidate cell state, ($\hat{c}_t$), which contains potential new information. This interaction determines what new data should be incorporated into the cell state. Finally, the outcomes of these interactions—what to forget and what to add—are combined to form the updated cell state, (c_t). This updated state is a fusion of past knowledge and new insights, ready to be carried forward to the next step in the sequence.

Hidden State Update:

$$h_t = O_t \cdot \tanh(C_t) \quad (6)$$

The hidden state (h_t) is computed by applying the hyperbolic tangent function ($\tanh$) to the cell state (C_t) and multiplying it with the output gate (O_t)

(Equation 6). This allows the LSTM to selectively output information from the cell state based on the current input and the learned weights (Figure 1).

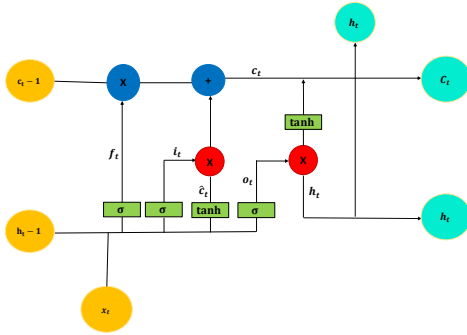


Figure 1. LSTM Neural Network Structure

Gated Recurrent Units (GRU) are similar to LSTMs as they also aim to solve the vanishing gradient problem but with a simpler structure. GRUs combine the forget and input gates into a single update gate and merge the cell state and hidden state, resulting in a more efficient and less complex model. GRUs are effective in tasks where LSTMs might be considered too heavy, such as smaller datasets or when computational efficiency is a priority.

The architecture of Gated Recurrent Unit (GRU) neural networks includes three main gates: the update gate, the reset gate, and the gate related to the content of the hidden state. The formulas related to these gates are as follows [35, 36]:

Update Gate:

$$z_t = \sigma(W_z \cdot [h_t - 1, x_t] + b_z) \quad (7)$$

$$r_t = \sigma(W_r \cdot [h_t - 1, x_t] + b_r) \quad (8)$$

The reset gate decides how much of the previous hidden state h_{t-1} should be forgotten when computing the candidate hidden state h_t . Similar to the update gate, the sigmoid activation function is applied to the weighted sum of the concatenated previous hidden state and current input, along with the bias term b_r . The resulting values in r_t range from 0 to 1, indicating the proportion of the previous hidden state to be forgotten (Equations 7, 8).

New Content for Hidden State:

$$\hat{h}_t = \tanh(W \cdot [r_t * h_t - 1, x_t] + b) \quad (9)$$

The new content for the hidden state $\hat{h}_t$ is computed based on the reset gate r_t , the previous hidden state h_{t-1} , and the current input x_t . The reset gate is used to selectively forget parts of the previous hidden state, and the result is concatenated with the current input. The weighted sum of this concatenation, along with the bias term b , is passed through the hyperbolic tangent activation function ($\tanh$) to obtain the new content for the hidden state (Equation 9).

Final Hidden State:

$$h_t = (1 - z_t) * h_t - 1 + z_t * \hat{h}_t \quad (10)$$

The update gate determines how much of the previous hidden state $h_t - 1$ and the current input x_t should be combined to form the new hidden state h_t . The sigmoid activation function σ is applied to the weighted sum of the concatenated previous hidden state and current input, along with the bias term b_z . The resulting values in z_t range from (0) to (1), indicating the proportion of the previous hidden state to be retained (Equation 10).

Reset Gate:

The final hidden state h_t is computed as a weighted sum of the previous hidden state h_{t-1} and the new content for the hidden state h_t . The update gate z_t determines the proportion of the previous hidden state to be retained and the proportion of the new content to be added. This allows the GRU to selectively remember or forget information from the previous hidden state based on the current input and the learned weights (Figure 2).

Model parameters play a significant role in the performance of these networks. Adjusting the number of layers adds depth to the network, allowing it to learn more complex features at the cost of increased computational load. The number of neurons per layer determines the capacity of the network to capture information; more neurons can capture finer details but also increase the risk of overfitting. The learning rate controls how quickly the network updates its weights during training; a smaller learning rate might lead to more precise convergence, while a larger learning rate can speed up training but might overshoot the optimal solution. To evaluate the model's performance and prevent overfitting, cross-validation was used. This method, by dividing the data into

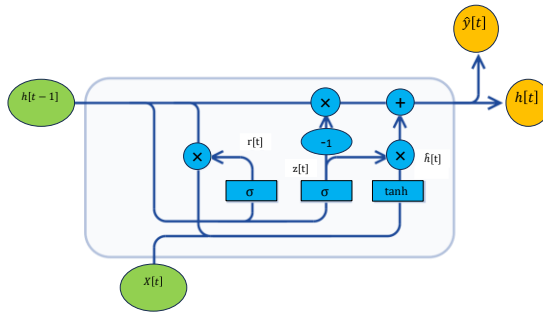


Figure 2. GRU Neural Network Structure

several subsets and testing the model on each of these subsets, examines the accuracy and generalizability of the model. The use of cross-validation as a standard method in machine learning ensures that the LSTM and GRU models are well-tuned and ready for precise and reliable predictions.

2.4. Design and Evaluation of LSTM and GRU Models Using Various Optimizers

An advanced model has been designed using the Keras library. The input shape of the model is defined by the number of features per sample and the number of channels (in this case, 1). The model consists of several layers, starting with an LSTM or GRU layer with 50 units. These layers have been chosen for their high capability in learning long-term temporal dependencies. To prevent overfitting, a Dropout layer with a rate of 0.2 has been added, which helps reduce the correlation between neural units. Additionally, L2 Regularization is applied to improve generalization by penalizing large weights. The final Dense layer has 1 unit and a sigmoid activation function for binary classification. The model is compiled using various optimizers such as SGD and Adam, with accuracy monitored as the primary metric during training.

To ensure the model's generalizability and prevent overfitting, cross-validation using K-Fold and Early-Stopping has been employed. The model is trained for 2000 epochs with a batch size of 32, using 20% of the training data for validation. Early stopping is utilized to monitor validation loss and adjust the learning rate to achieve the best possible model. These methods help effectively train and evaluate the LSTM and GRU models, ensuring optimal performance under various conditions. The results obtained from these models using different optimizers are detailed in the article to precisely examine the superiority and efficiency of

each method. In this study, LSTM and GRU models have been compiled using various optimizers such as SGD and Adam. The SGD optimizer is configured with a learning rate of 0.01 and momentum of 0.9 to provide faster convergence. The Adam optimizer has been chosen for its high capability to adaptively adjust the learning rate.

2.5. Model Evaluation

In the evaluation of neural network performance for diagnosis and prediction tasks, a variety of metrics are employed, including accuracy, sensitivity [37], and the F1 score [38]. Accuracy reflects the overall correctness of the model's diagnoses, while sensitivity measures the model's ability to identify true positive cases. The F1 score, which is the harmonic mean of accuracy and sensitivity, provides a balance between precision and recall, useful in situations with imbalanced class distributions.

2.6. Accuracy

The accuracy metric evaluates the overall correctness of the model's predictions. It calculates the ratio of correctly predicted instances (both true positives (TP) and true negatives (TN)) to the total number of instances. FP (False Positives): The number of negative samples that have been incorrectly classified as positive. FN (False Negatives): The number of positive samples that have been incorrectly classified as negative (Equation 11).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

2.7. Sensitivity (Recall or True Positive Rate):

Sensitivity, also known as recall or true positive rate, measures the proportion of actual positive instances that are correctly identified by the model. It represents the model's ability to detect true positives (Equation 12).

$$Sensitivity = \frac{TP}{TP + FN} \quad (12)$$

F1 Score:

The F1 score is the harmonic mean of precision and sensitivity. It provides a balance between the two metrics, especially in cases where the class distribution is imbalanced. The F1 score ranges from (0) to (1), with (1) indicating perfect precision and sensitivity (Equations 13, 14).

$$F1\ Score = 2 \times \frac{Precision \times Sensitivity}{TP + FP} \quad (13)$$

Where:

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

3. Results

In this section, using two distinct optimization methods, Adam and SGDM, for the LSTM and GRU neural networks, the models were trained on a diverse dataset including demographic and clinical features such as PatientID, age, clinical T Stage, Clinical N Stage, Clinical M Stage, Overall Stage, Histology, gender, Survival time, and dead status event. The k-fold cross-validation results, presented in Figures 3 to 6, demonstrate the performance.

4. Discussion

This article investigates and optimizes two prevalent neural network architectures, LSTM and GRU, to enhance the accuracy of diagnosing and predicting lung cancer stages. By utilizing two distinct optimization methods, Adam and SGDM, models have been trained that not only exhibit high diagnostic accuracy but also can predict patient survival using clinical data. Based on the outcomes of numerous evaluations, the GRU neural network optimized with SGDM has demonstrated remarkable performance. This model, with a Test Loss of 0.0531, Test Accuracy of 0.97, Sensitivity (Recall) of 0.97, Precision of 0.98, and an F1 Score of 0.975, has been identified as the best model for predicting and diagnosing clinical patient data of lung cancer. In Fold 3, the mean training loss was 0.1981, the last training loss was 0.0029, the mean validation loss was 0.2488, and the last validation loss was 0.0533. These results suggest that the GRU model optimized with SGDM can act as

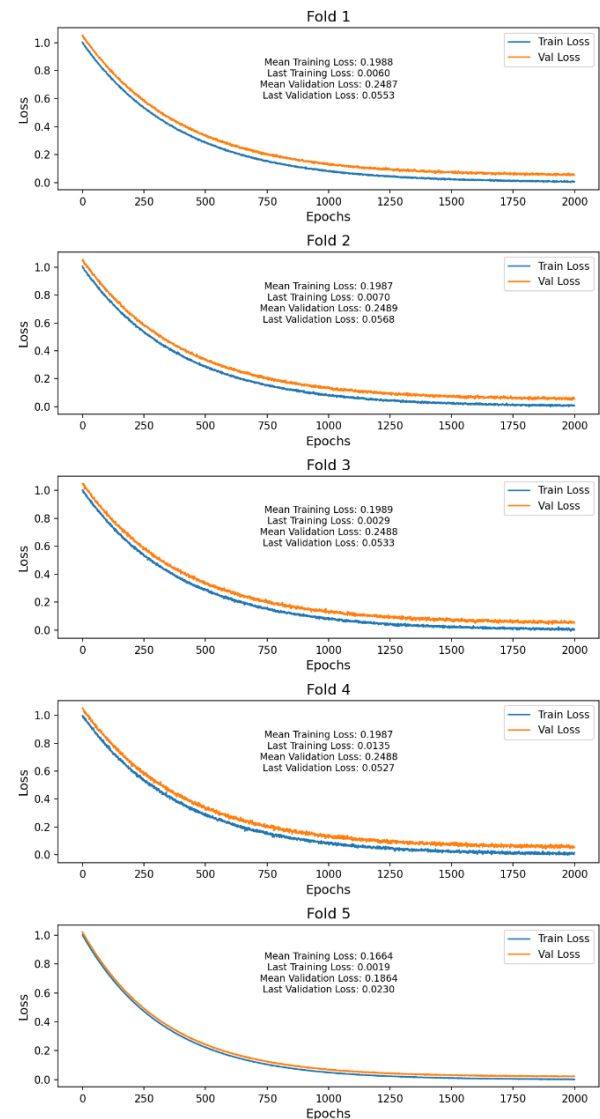


Figure 3. Performance of the GRU Recurrent Neural Network Optimized with the SGDM Algorithm using 5-Fold Cross-Validation

a potent tool in the diagnosis and prediction of lung cancer.

4.1. Advantages of Recurrent Neural Networks (RNNs)

Recurrent neural networks, including GRU and LSTM, offer several advantages over traditional machine learning models (e.g., SVM, Random Forest, KNN) and other deep learning architectures (e.g., CNN).

4.2. Handling Sequential Data

RNNs are specifically designed to handle sequential data, making them ideal for time-series analysis and

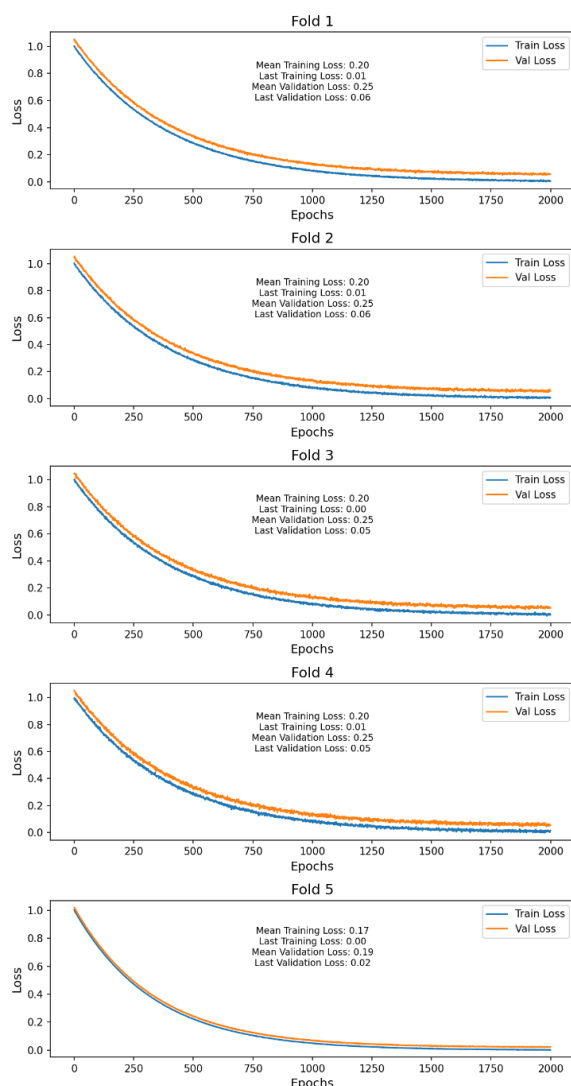


Figure 4. Performance of the GRU Recurrent Neural Network Optimized with the Adam Algorithm using 5-Fold Cross-Validation

tasks where the order of data points is crucial. This is particularly important in medical applications where patient data often involves sequences of events over time [39].

4.3. Memory Capability

RNNs can retain information from previous inputs, allowing them to capture temporal dependencies and patterns that traditional models like SVM and KNN cannot [40]. This capability is essential for accurately modeling the progression of diseases like lung cancer.

4.4. Dynamic Temporal Behavior

Unlike feedforward neural networks, RNNs can model dynamic temporal behavior, which is crucial for

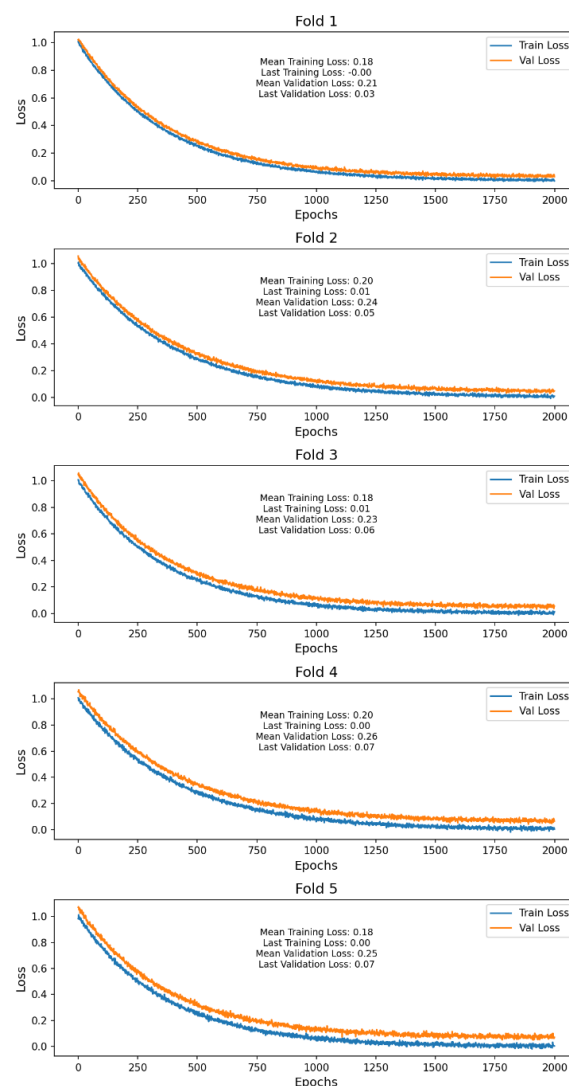


Figure 5. Performance of the LSTM Recurrent Neural Network Optimized with the SGDM Algorithm using 5-Fold Cross-Validation

tasks involving sequences of varying lengths [41]. This makes RNNs more flexible and powerful in handling complex medical data.

4.5. Improved Performance

In experiments, RNNs, particularly GRUs, have shown better performance in terms of accuracy and sensitivity compared to other models. The GRU model optimized with SGDM, delivering exceptional results in all evaluation metrics, demonstrates how advanced machine learning techniques can be applied in the medical field to achieve more accurate and reliable outcomes.

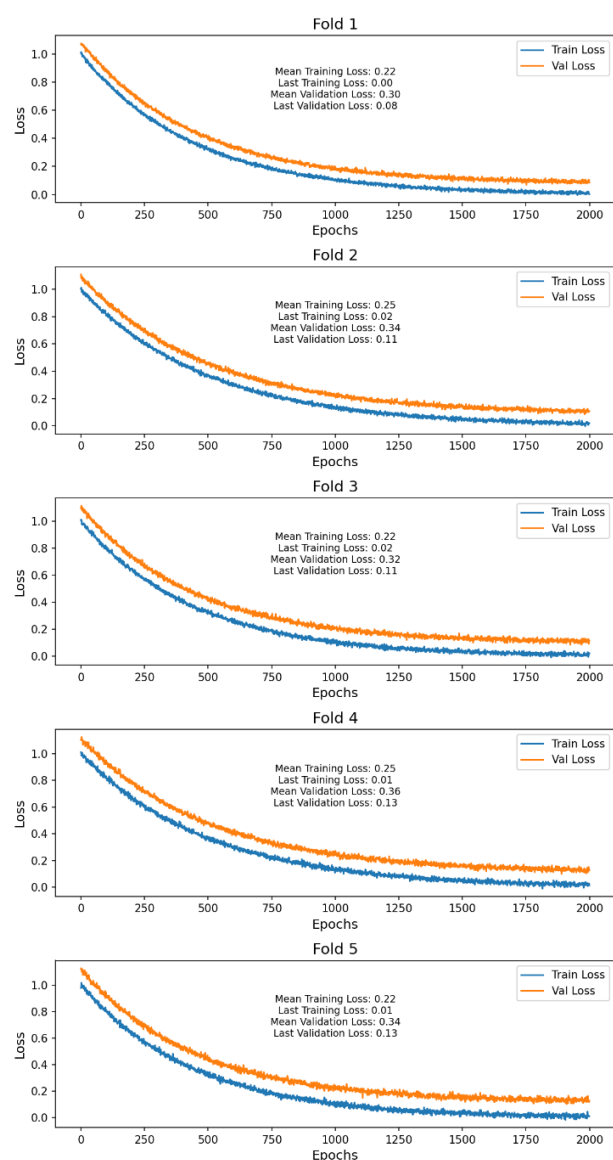


Figure 6. Performance of the LSTM Recurrent Neural Network Optimized with the Adam Algorithm using 5-Fold Cross-Validation

4.6. Advantages over CNNs

While CNNs are excellent for processing image and spatial data, RNNs are more suitable for sequential and temporal data. RNNs can identify long-term dependencies in data, which is crucial for tasks involving temporal data, such as predicting disease progression [42].

4.7. Non-Invasive and Efficient

The GRU model optimized with SGDM provides a non-invasive, fast, and highly accurate assessment of lung cancer stages. The use of advanced optimization methods like SGDM significantly reduces the training

time of the neural network, making the process more efficient. In this study, the optimized GRU model was able to provide results in just 50 seconds, demonstrating its potential for real-time clinical applications. This efficiency not only saves time but also reduces the overall cost of the diagnostic process.

4.8. Timely Diagnosis

One of the key reasons for using optimization methods like SGDM is to reduce the time required for the learning and testing phases of the neural network. The ability to quickly achieve accurate results in a clinical setting is crucial, where timely diagnosis can significantly impact patient outcomes. The reduced processing time also leads to lower operational costs, making the technology more accessible and sustainable for healthcare providers.

4.9. Limitations and Future Work

While the results are promising, it is important to acknowledge the limitations related to the diversity of the dataset used. The dataset may not fully represent the wide range of clinical scenarios encountered in real-world settings. Future work should focus on expanding the dataset to include more diverse patient populations and clinical conditions to ensure the model's robustness and generalizability. Additionally, the novelty of the approach could be further enhanced by integrating multi-modal data sources, such as imaging and genetic data, to provide a more comprehensive diagnostic tool. Exploring the combination of GRU models with other advanced techniques, such as attention mechanisms, could also lead to further improvements in performance. From Tables 2 to 5, the performance metrics of the recurrent neural network models, GRU and LSTM, optimized with Adam and SGDM, across 5-fold Cross-Validation are presented.

5. Conclusion

This study meticulously optimized and evaluated two prevalent neural network architectures, LSTM and GRU, for the diagnosis and prediction of lung cancer stages. By employing advanced optimization techniques, specifically Adam and SGDM, the models

Table 2. Performance Metrics of the GRU Model Optimized with SGDM Across 5-Fold Cross-Validation

Fold	Test Loss	Test Accuracy	Sensitivity (Recall)	Precision	F1 Score
1	0.0549	0.96	0.96	0.97	0.96
2	0.0571	0.94	0.95	0.96	0.955
3	0.0531	0.97	0.97	0.98	0.975
4	0.0541	0.95	0.96	0.97	0.965
5	0.0225	0.96	0.95	0.96	0.955

Table 3. Performance Metrics of the GRU Model Optimized with Adam Across 5-Fold Cross-Validation

Fold	Test Loss	Test Accuracy	Sensitivity (Recall)	Precision	F1 Score
1	0.05	0.90	0.90	0.91	0.90
2	0.06	0.89	0.89	0.90	0.89
3	0.05	0.88	0.88	0.89	0.88
4	0.05	0.87	0.87	0.88	0.87
5	0.02	0.86	0.86	0.87	0.86

Table 4. Performance Metrics of the LSTM Model Optimized with SGDM Across 5-Fold Cross-Validation

Fold	Test Loss	Test Accuracy	Sensitivity (Recall)	Precision	F1 Score
1	0.03	0.93	0.93	0.94	0.93
2	0.05	0.92	0.92	0.93	0.92
3	0.06	0.91	0.91	0.92	0.91
4	0.07	0.90	0.90	0.91	0.90
5	0.07	0.89	0.89	0.90	0.89

Table 5. Performance Metrics of the LSTM Model Optimized with Adam Across 5-Fold Cross-Validation

Fold	Test Loss	Test Accuracy	Sensitivity (Recall)	Precision	F1 Score
1	0.08	0.87	0.87	0.88	0.87
2	0.11	0.86	0.86	0.87	0.86
3	0.11	0.85	0.85	0.86	0.85
4	0.13	0.84	0.84	0.85	0.84
5	0.12	0.83	0.83	0.84	0.83

were rigorously trained and validated on a diverse dataset encompassing critical demographic and clinical features. The GRU model optimized with SGDM demonstrated superior performance, achieving exceptional accuracy and evaluation metrics, including high sensitivity, precision, and F1-score. These findings underscore the profound impact that the precise selection of optimization methods and neural network architectures can have on the performance of models in diagnosing and predicting complex diseases such as lung cancer. The success of the GRU model optimized with SGDM highlights its potential as a powerful tool in clinical settings, offering non-invasive, rapid, and highly accurate assessments of lung cancer stages. Moreover, this research emphasizes the importance of integrating robust data augmentation techniques and comprehensive cross-validation to enhance model

generalizability and reliability. The practical implications of these advanced machine learning techniques extend beyond lung cancer diagnosis, paving the way for their application in other medical domains. In conclusion, the advancements presented in this study mark a significant step towards improving the diagnosis and treatment of lung cancer. The demonstrated efficacy of the GRU model optimized with SGDM not only contributes to the field of medical diagnostics but also sets the stage for future innovations and improvements in healthcare through the application of cutting-edge neural network architectures and optimization strategies.

Acknowledgment

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

- 1- Abhishek Padhi, Ashwini Agarwal, Shailendra K Saxena, and CDS Katoch, "Transforming clinical virology with AI, machine learning and deep learning: a comprehensive review and outlook." *VirusDisease*, Vol. 34 (No. 3), pp. 345–55, (2023).
- 2- Nishant Upadhyay and Anamika Gulati, "Applications of Artificial Intelligence and Machine Learning for Diagnosis, Prediction, and Smart Health Care." in *Artificial Intelligence and Machine Learning in Healthcare: Springer*, (2023), pp. 79–94.
- 3- Dhir Gala, Haditya Behl, Mili Shah, and Amgad N Makaryus, "The role of artificial intelligence in improving patient outcomes and future of healthcare delivery in cardiology: a narrative review of the literature." in *Healthcare*, (2024), Vol. 12 (No. 4): MDPI, p. 481.
- 4- Balázs Acs, Mattias Rantalainen, and Johan Hartman, "Artificial intelligence as the next step towards precision pathology." *Journal of internal medicine*, Vol. 288 (No. 1), pp. 62–81, (2020).
- 5- Mohammed Kanan *et al.*, "AI-Driven Models for Diagnosing and Predicting Outcomes in Lung Cancer: A Systematic Review and Meta-Analysis." *Cancers*, Vol. 16 (No. 3), p. 674, (2024).
- 6- Amanda Leiter, Rajwanth R Veluswamy, and Juan P Wisnivesky, "The global burden of lung cancer: current status and future trends." *Nature reviews Clinical oncology*, Vol. 20 (No. 9), pp. 624–39, (2023).
- 7- Saife N Lone *et al.*, "Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments." *Molecular cancer*, Vol. 21 (No. 1), p. 79, (2022).
- 8- Robert Dziedzic, Tomasz Marjański, and Witold Rzyman, "A narrative review of invasive diagnostics and treatment of early lung cancer." *Translational Lung Cancer Research*, Vol. 10 (No. 2), p. 1110, (2021).
- 9- Yongzhong Li, Donglai Chen, Xuejie Wu, Wentao Yang, and Yongbing Chen, "A narrative review of artificial intelligence-assisted histopathologic diagnosis and decision-making for non-small cell lung cancer: achievements and limitations." *Journal of Thoracic Disease*, Vol. 13 (No. 12), p. 7006, (2021).
- 10- Daniel Di Capua, Dara Bracken-Clarke, Karine Ronan, Anne-Marie Baird, and Stephen Finn, "The liquid biopsy for lung cancer: state of the art, limitations and future developments." *Cancers*, Vol. 13 (No. 16), p. 3923, (2021).
- 11- Aristidis G Vrahatis, Konstantina Skolariki, Marios G Krokidis, Konstantinos Lazaros, Themis P Exarchos, and Panagiotis Vlamos, "Revolutionizing the early detection of Alzheimer's disease through non-invasive biomarkers: the role of artificial intelligence and deep learning." *Sensors*, Vol. 23 (No. 9), p. 4184, (2023).
- 12- Damien Gruson, Sergio Bernardini, Pradeep Kumar Dabla, Bernard Gouget, and Sanja Stankovic, "Collaborative AI and laboratory medicine integration in precision cardiovascular medicine." *Clinica Chimica Acta*, Vol. 509pp. 67–71, (2020).
- 13- SKB Sangeetha *et al.*, "An enhanced multimodal fusion deep learning neural network for lung cancer classification." *Systems and Soft Computing*, Vol. 6p. 200068, (2024).
- 14- Shane Dawson, "The concept of personalized and adaptive learning has long been touted but seldom enacted in education at scale. Data Analytics and Adaptive Learning brings together a compelling set of experts that provide novel and research-informed insights into contemporary education spaces."
- 15- Anirudh Atmakuru *et al.*, "Deep learning in radiology for lung cancer diagnostics: A systematic review of classification, segmentation, and predictive modeling techniques." *Expert Systems with Applications*, p. 124665, (2024).
- 16- Shriniket Dixit, Anant Kumar, and Kathiravan Srinivasan, "A current review of machine learning and deep learning models in oral cancer diagnosis: Recent technologies, open challenges, and future research directions." *Diagnostics*, Vol. 13 (No. 7), p. 1353, (2023).
- 17- Federico Landi, Lorenzo Baraldi, Marcella Cornia, and Rita Cucchiara, "Working memory connections for LSTM." *Neural Networks*, Vol. 144pp. 334–41, (2021).
- 18- Benjamin Lindemann, Benjamin Maschler, Nada Sahlab, and Michael Weyrich, "A survey on anomaly detection for technical systems using LSTM networks." *Computers in Industry*, Vol. 131p. 103498, (2021).
- 19- Rusul L Abduljabbar, Hussein Dia, and Pei-Wei Tsai, "Unidirectional and bidirectional LSTM models for short-term traffic prediction." *Journal of Advanced Transportation*, Vol. 2021 (No. 1), p. 5589075, (2021).
- 20- Shalini Wankhade and S Vigneshwari, "A novel hybrid deep learning method for early detection of lung cancer using neural networks." *Healthcare Analytics*, Vol. 3p. 100195, (2023).
- 21- Rijwan Khan, Indrajeet Kumar, and Pushkar Praveen, "Applications of Parallel Data Processing for Biomedical Imaging." *IGI Global*, (2024).
- 22- Apeksha Shewalkar, Deepika Nyavanandi, and Simone A Ludwig, "Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and

- GRU." *Journal of Artificial Intelligence and Soft Computing Research*, Vol. 9 (No. 4), pp. 235–45, (2019).
- 23- Thanh Bui-Tien, Thanh Nguyen-Chi, Thang Le-Xuan, and Hoa Tran-Ngoc, "Enhancing bridge damage assessment: Adaptive cell and deep learning approaches in time-series analysis." *Construction and Building Materials*, Vol. 439p. 137240, (2024).
 - 24- Hossein Shahinzadeh, Arezou Mahmoudi, Amirhosein Asilian, Hamidreza Sadrarhami, Mohammadreza Hemmati, and Yaghoub Saberi, "Deep Learning: A Overview of Theory and Architectures." in *2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)*, (2024): IEEE, pp. 1–11.
 - 25- Sixian Chan, Menghui Zhang, Yu-Yi Zhi, Saeid Razmjoo, Ahmed M El-Sherbeeney, and Lingzhi Lin, "Improved anterior cruciate ligament tear diagnosis using gated recurrent unit networks and Hybrid Tasmanian Devil Optimization." *Biomedical Signal Processing and Control*, Vol. 95p. 106309, (2024).
 - 26- TL Spandana, D Tejomai, D Malleswara Rao, A Vijaya Shyam, and B Sushanth, "Tumor Identification and Classification of MRI Brain Images Using Deep Learning Optimizers (Adam vs Sgdm)." in *2024 IEEE Wireless Antenna and Microwave Symposium (WAMS)*, (2024): IEEE, pp. 1–5.
 - 27- Yakub Kayode Saheed, Adekunle Isaac Omole, and Musa Odunayo Sabit, "GA-mADAM-IIoT: A new lightweight threats detection in the industrial IoT via genetic algorithm with attention mechanism and LSTM on multivariate time series sensor data." *Sensors International*, Vol. 6p. 100297, (2025).
 - 28- Mahmoud Y Shams, Zahraa Tarek, and Ahmed M Elshewey, "A novel RFE-GRU model for diabetes classification using PIMA Indian dataset." *Scientific Reports*, Vol. 15 (No. 1), p. 982, (2025).
 - 29- Harun Özbay, İlyas Özer, Adem Dalcı, Onursal Çetin, and Feyzullah Temurtaş, "Sensorless Position Estimation in Electromagnetic Launchers Using Recurrent Neural Networks with Repeated k-Fold Cross-Validation." *Arabian Journal for Science and Engineering*, pp. 1–19, (2025).
 - 30- Tang Shi and Kazuya Shide, "A comparative analysis of LSTM, GRU, and Transformer models for construction cost prediction with multidimensional feature integration." *Journal of Asian Architecture and Building Engineering*, pp. 1–16, (2025).
 - 31- Bara Masalmeh, Rashmi Prasad, Vaibhav Nougain, and Aleksandra Lekić, "Neural Networks in RSCAD: Enhancing MMC-based HVDC Simulation with Advanced Machine Learning." *IEEE Transactions on Industry Applications*, (2025).
 - 32- Abdullah Kaviani Rad, Mohammad Javad Nematollahi, Abbas Pak, and Mohammadreza Mahmoudi, "Predictive modeling of air quality in the Tehran megacity via deep learning techniques." *Scientific Reports*, Vol. 15 (No. 1), p. 1367, (2025).
 - 33- Diyin Zhu, Yihang Wei, Jiali Cai, Jingwen Wang, and Zhongshan Chen, "A security data detection and management method in digital library network based on deep learning." *Frontiers in Physics*, Vol. 12p. 1492114, (2025).
 - 34- Minsi Wang, Li Song, Xiaokang Yang, and Chuanfei Luo, "A parallel-fusion RNN-LSTM architecture for image caption generation." in *2016 IEEE international conference on image processing (ICIP)*, (2016): IEEE, pp. 4448–52.
 - 35- Meng Jiao, Dongqing Wang, and Jianlong Qiu, "A GRU-RNN based momentum optimized algorithm for SOC estimation." *Journal of Power Sources*, Vol. 459p. 228051, (2020).
 - 36- Rohaifa Khaldi, Abdellatif El Afia, Raddouane Chiheb, and Siham Tabik, "What is the best RNN-cell structure to forecast each time series behavior?" *Expert Systems with Applications*, Vol. 215p. 119140, (2023).
 - 37- Ruan Luzia, Lihki Rubio, and Carlos E Velasquez, "Sensitivity analysis for forecasting Brazilian electricity demand using artificial neural networks and hybrid models based on Autoregressive Integrated Moving Average." *Energy*, Vol. 274p. 127365, (2023).
 - 38- Shagun Sharma and Kalpna Guleria, "A deep learning based model for the detection of pneumonia from chest X-ray images using VGG-16 and neural networks." *Procedia Computer Science*, Vol. 218pp. 357–66, (2023).
 - 39- Mohd Sakib, Suhel Mustajab, and Mahfooz Alam, "Ensemble deep learning techniques for time series analysis: a comprehensive review, applications, open issues, challenges, and future directions." *Cluster Computing*, Vol. 28 (No. 1), pp. 1–44, (2025).
 - 40- Zhi-Qiang Dai, Jie Li, Yang-Jie Cao, and Yong-Xiang Zhang, "SALSTM: segmented self-attention long short-term memory for long-term forecasting." *The Journal of Supercomputing*, Vol. 81 (No. 1), p. 115, (2025).
 - 41- Sai Babu Veeram et al., "Design of an integrated model with temporal graph attention and transformer-augmented RNNs for enhanced anomaly detection." *Scientific Reports*, Vol. 15 (No. 1), p. 2692, (2025).
 - 42- Shamsul Haq and Prabal Verma, "Edge-Cloud-Assisted Multivariate Time Series Data-Based VAR and Sequential Encoder–Decoder Framework for Multi-Disease Prediction." *Arabian Journal for Science and Engineering*, pp. 1–21, (2025).