Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 12s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Enhanced Cancer Classification Using Optimized Deep Learning Approaches: A Novel Framework Integrating RNA Sequence Analysis and AI Techniques

Sivasubramaniam S[1], Dr. S. Prabhu[2]

[1]Research Scholar, PG and Research Department, Department of Computer Science Park's

[2]Research Supervisor, PG and Research Department, Department of Computer Science Park's College, Chinnakarai, Tirupur, India.

Cite this paper as: Sivasubramaniam S, Dr. S. Prabhu, (2025) Enhanced Cancer Classification Using Optimized Deep Learning Approaches: A Novel Framework Integrating RNA Sequence Analysis and AI Techniques. *Journal of Neonatal Surgery*, 14 (11s), 17-29.

ABSTRACT

Lung cancer is one of the most aggressive diseases globally, responsible for over 9 million deaths annually. Early staging is critical for improving recovery rates, and RNA sequence analysis has emerged as a vital technique in this process. Recent advancements in AI have significantly enhanced the efficiency and accuracy of human genomics analysis. This study proposes Lung **Cancer Classification Using Optimized Deep Learning Approaches(CC-ODLA)** to classify various lung cancer typesspecifically kidney renal clear cell carcinoma (KIRC), breast invasive carcinoma (BRCA), lung squamous cell carcinoma (LUSC), lung adenocarcinoma (LUAD), and uterine corpus endometrial carcinoma (UCEC)—using deep learning methods. The first approach combines BPSO-DT and CNN to analyzetumor RNA-seq gene expression data. The proposed methodology consists of three phases: (1) preprocessing, which optimizes high-dimensional RNA-seq data to select optimal features and converts them into 2D images; (2) data augmentation, which increases the dataset size from 2,086 samples to five times larger to mitigate overfitting; and (3) deep CNN architecture, employing a two-layer convolutional framework for feature extraction and classification, achieving a testing accuracy of 96.90%. The second approach introduces an AFOECNN, designed to enhance classification accuracy in lung cancer datasets. This approach also involves three phases: (1) preprocessing using K-Means Clustering (KMC) to reduce noise and handle missing data; (2) feature subset selection via the AFO algorithm to identify significant features based on fitness values; and (3) classification using the Enhanced CNN (ECNN) algorithm, which demonstrates superior precision, recall, F-measure, and accuracy compared to existing algorithms while maintaining lower time complexity. The comparative results indicate that the proposed methods not only improve classification performance but also offer efficient computational resource management, underscoring the transformative potential of AI in smart healthcare solutions.Using Adaptive Firefly Optimisation, the Enhanced CNN improves classification performance with 98% testing accuracy, 93.21% precision, 98% recall, and 91.82% F1-score. However, the combination BPSO-DT and CNN method achieves 96.90% testing accuracy.

Keywords: *Lung Cancer Classification, Optimized Deep Learning, Convolutional Neural Network, K-Means Clustering, BRCA, UCEC, particle swarm optimization*

1. INTRODUCTION

Still one of the most horrible global health hazard with over 9 million deaths per year is lung cancer[1]. Early identification and correct classification of lung cancer's complexity and aggressiveness is necessary for better treatment and prognosis[2]. Among the innovative diagnostic technologies, RNA sequencing (RNA-seq) has been more crucial for understanding tumor biology and lung cancer type diagnosis depending on gene expression patterns[3]. Nevertheless, processing and interpretation of the high dimension of RNA-seq data presents challenges[4].

Recent artificial intelligence (AI) developments especially deep learning offer promising approaches to handle such difficult material[5]. Deep learning models especially CNNs have shown astonishing performance in feature extraction and classification issues across many disciplines, including healthcare[6]. Optimizing these models to gain higher accuracy, lower processing costs, and more generalizing power is very essential in lung cancer classification[7].

Lung cancer Classification Using Optimized Deep Learning Approaches (CC-ODLA) is a new framework integrating AI techniques with RNA-seq data presented in this paper[8]. Two first approaches are proposed: a hybrid model combining BPSO-DT with CNN and an AFOECNN[8]. These methods strive to increase classification accuracy, efficiency, and resource management for various lung cancer types, thereby offering a robust AI-based tool for tailored lung cancer diagnosis[9].

Motivation: Given the critical need of accurate lung cancer classification and early diagnosis, this study intends to employ advanced deep learning techniques to address the limits of current lung cancer classification methods[10]. Early intervention assisted, computational complexity reduced, and classification accuracy much raised by application of optimum AI algorithms helps to save lives and improve patient outcomes. Combining deep learning with RNA-seq advances the field of customized lung cancer treatment even further.

Problem statement: Correct classification of lung cancer types based on RNA-seq data remains challenging given the tremendous dimensionality and complexity of gene expression patterns. Current machine learning methods can suffer in multi-class lung cancer classification tasks from overfitting, inefficiencies, and low accuracy[11]. This paper aims to address these challenges thus improving early diagnosis and treatment choices by means of a novel deep learning framework maximizing feature selection, reduces computational overhead, and improves classification performance for various lung cancer types.

The objectives of this paper

- To provide a framework suitable for lung cancer classification With RNA sequence analysis combined with innovative artificial intelligence techniques, optimistically to design and apply the CC-ODLA framework for exact categorization of numerous lung cancer kinds.
- Use of this helps to enhance data processing and feature selection. Using BPSO-DT and K-Means Clustering (KMC), for effective feature selection and data augmentation, hence improving the quality and quantity of RNA-seq data utilized in classification applications.
- By means of deep learning models including CNN and Enhanced CNN (ECNN) to classify lung cancer types with high accuracy, precision, recall, and F-measure, so highlighting the possibilities of these approaches in enhancing early-stage lung cancer detection and so improving patient outcomes.

The upcoming section is as follows: section 2 deliberates the related works, section 3 examines the proposed methodology, section 4 describes the results and discussion and section 5 concludes the overall paper work.

2. RELATED WORK

This section will review several advanced optimization techniques used to increase the accuracy of lung cancer classification by means of machine learning approaches. BPSO-DT, AFOECNN, RNN, SDAE, and the BFA are discussed in this study. Every method emphasizes, if one aims to provide a proper diagnosis, the need of data processing and feature selection.

Binary Particle Swarm Optimization with Decision Tree (BPSO-DT)

An evolutionary strategy is used to optimise the classification of lung nodules using CNNs. An GA is used to create CNN designs and optimise hyperparameters, approaching optimal solutions[12]. This algorithm is based on a variety of bio-inspired methodologies, including natural selection and Darwinism. To compare two deep learning models that were built manually, Deep Local-Global Network and FractalNet were trained. The best GA-CNN achieved a remarkable classification accuracy, surpassing both of the models that were developed manually, as seen below. Based on the findings, GAs are a viable option for diagnosing issues; in the future, this process might be fully automated by using GAs to construct and optimise CNN structures for various therapeutic applications.

Adaptive Firefly Optimization with Enhanced Convolutional Neural Network (AFOECNN)

This paper suggests to increase the classification accuracy by means of Adaptive Firefly Optimization with Enhanced Convolutional Neural Network (AFOECNN). Three primary phases preprocessing, feature subset selection and classification form part of the proposed system. K-Means Clustering (KMC) method is used preprocessing to lower the noise data from the provided gene expression dataset. Using k-means centroid values, it manages the missing features[13]. It helps to more successfully raise the categorization accuracy. The technique of feature subset selection for the preprocessed data helps to identify more instructive elements from the lung cancer dataset. It is carried out using AFO technique and computes the required and conspicuous feature depending on the best fitness values utilizing the objective function. It turns the data into pictures then enhances those visuals.

Recurrent Neural Network (RNN)

To choose a subset of features, proposed original framework of feature selection based on RNN. In particular, the framework has been used to choose micro-array data characteristics for cell categorization[14]. Under the framework, there are four feature selection models with varying topologies of recurrent neural networks under which GRU, LSTM, RNN and bi-directional LSTM find resonance. Using actual micro-array datasets, the benefits of the system are shown.

Stacked Denoising Autoencoder (SDAE)

The paper proposed a deep learning method for identifying genes crucial for the diagnosis of breast lung cancer and for lung cancer detection overall. Using Stacked Denoising Autoencoder (SDAE), we first painstakingly retrieved functional traits

from high-dimensional gene expression patterns. The next step was to evaluate the acquired representation's performance using supervised classification models[15] to validate the new features' usefulness in lung cancer detection. An SDAE connection matrix analysis led to the discovery of a cluster of highly interdependent genes. Based on the results and analysis, it seems that these interacting genes might be useful lung cancer biomarkers for detecting breast lung cancer that warrants more investigation..

Binary Firefly Algorithm(BFA)

To find the best subset of features to use in a Random Forest classifier for breast, cervical, and hepatocellular carcinoma (liver) lung cancer diagnosis, it looked at adding a penalty function to the existing fitness function that promotes the Binary Firefly Algorithm. Compared to other modern methods based on Deep Learning and Information Gain, it showed an improvement in classification accuracy and feature reduction [16].

The paper shows how effectively merging deep learning models with optimization methods could help to detect lung cancer. BPSO-DT and AFOECNN are able to raise the accuracy of diagnostic jobs by means of evolutionary techniques. Unlike the SDAE, which identifies important genes for breast lung cancer screening, the RNN framework has great promise in the micro-array feature selection. BFA emphasizes the possible consequences of these methods in the detection of lung cancer by proving a higher degree of feature reduction and classification accuracy than the approaches already used.

3. PROPOSED METHOD

The paper includes a collection of deep learning algorithm and data processing approach-based diagrams illustrating numerous complex ways for lung cancer detection and analysis. Every figure shows a different framework or method emphasizing the effectiveness of many approaches in improving classification accuracy and performance across several lung cancer kinds. Mostly grounded on deep learning architectures, feature selection, and preprocessing, these novel ideas

Contribution 1: Develop and Propose CC-ODLA Framework

The paper propose the CC-ODLA framework for classifying several lung cancer types using advanced deep learning techniques with an eye toward kidney renal clear cell carcinoma, breast invasive carcinoma, lung squamous cell carcinoma, lung adenocarcinoma and uterine corpus endometrial carcinoma.

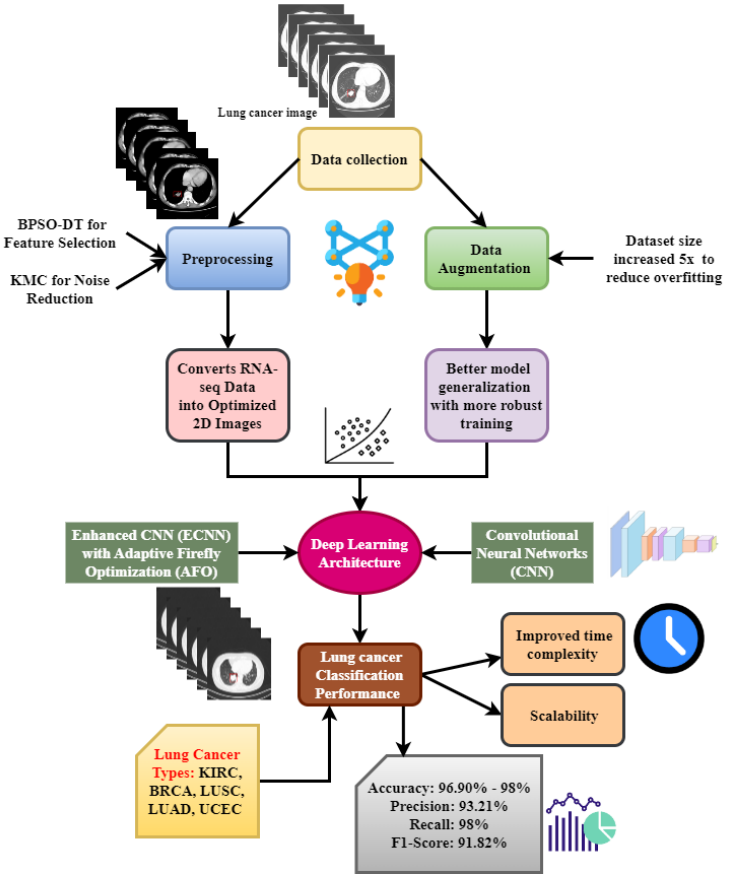

Figure 1:Framework for Enhanced Lung cancer Classification Using Optimized Deep Learning Approaches (CC-ODLA)

Figure 1 offers a framework for best deep learning-based lung cancer categorization. The preprocessing step in the figure consists mostly of two phases: noise reduction using the KMC approach and feature selection using BPSO-DT. K-means clustering, often known as KMC, helps remove noise from the data that only relevant trends show up for further analysis. BPSO-DT reduces the dimensionality of the dataset by selecting the most relevant characteristics, therefore conserving significant gene expression data. Following that, ideal 2D images derived from the processed RNA-seq data are fit for deep learning models. This translation allows the framework better identify tumors by using CNNs and other image-based deep learning architectures. Preprocessing assures ordered, appropriate, and clean input, hence enhancing the learning capacity and performance of the model.

Data augmentation thus five-fold increases the dataset size, hence enhancing model generalization and reducing overfitting. Main deep learning architecture combines Enhanced CNN (ECNN) with Adaptive Firefly Optimization (AFO) and standard CNNs to improve classification accuracy and efficiency. Showing its flexibility, the method identifies many lung cancer types including Kirc, BRCA, LusC, LUAD, and UCEC. Great accuracy (96.90% – 98%), precision (93.21%), recall (98%), and F1-score (91.82%) are shown by performance statistics. Moreover perfect for real-world, extensive lung cancer classification projects, the framework offers improved scalability and time complexity[17].

$$|\partial qw - pk''|: Jk < P - rty'' + Nk > \quad (1)$$

This is the CC-ODLA framework's equation 1 representation of the input feature-output categorization connection$\partial qw - pk''$. To increase the accuracy of lung cancer classificationJk, it shows how the optimization parameters (such as biases and weights)$P - rty''$ interact inside the modelNk. Improved prediction accuracy in deep learning applications is made possible by equation 1, which successfully correlates RNA sequence data with lung cancer kinds.

$$\partial_2 Ep < Ty - pkj'' > : Wa < Vx - zui''' > \quad (2)$$

By highlighting the changes made to optimize feature selection$Ty - pkj''$ and classification performanceWa, the equation 2showsthe energy potential ($\partial_2 Ep$) is connected to the model parameters and how sensitive it is$Vx - zui'''$. The need for dynamic optimization in developing effective AI-driven healthcare solutions is shown by this equation.

$$-Ok_2 < Pl - rw'' > : Hg < Wq + uyt'' > \quad (3)$$

In the context of lung cancer classification$-Ok_2$, equation 3 describes the interplay$Pl - rw''$ between optimization criteriaHg and loss functions$Wq + uyt''$. This equation highlights the significance of achieving a balance between precision and computing efficiency within the CC-ODLA framework, which enables effective decision-making while diagnosing lung cancer[18].

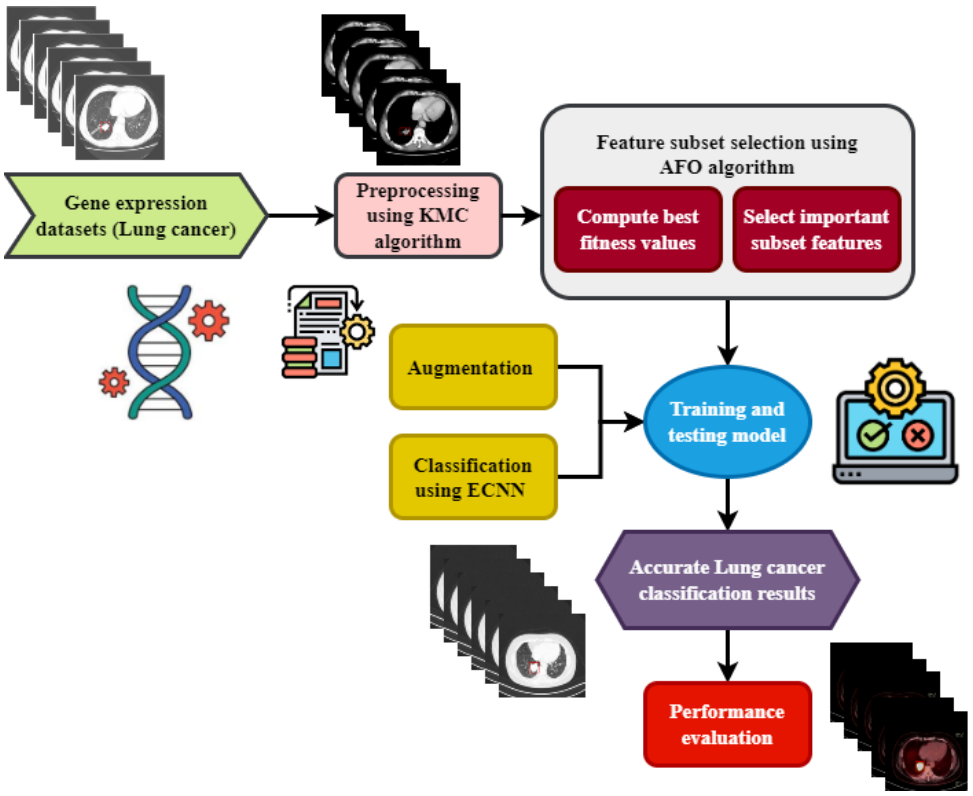

Figure 2: AFO-ECNN Framework for Lung cancer Classification

The suggested lung cancer classification based on gene expression data system consists of three main stages. Beginning is data collecting for gene expression databases from tumor tissues shown in figure 2. In the **preprocessing stage**, normalizing numbers for consistency, imputation replacing missing values and outlier detection and removal assist to prepare the data thereby ensuring reliable analysis. The preparation stage ensures uniform, neat, appropriate for data on gene expression in studies. It covers outlier detection to remove aberrant data points that might skew results, imputation to handle missing data by projecting appropriate values, and consistent normalization of scale values. These techniques ensure that only relevant and accurate features are used for further investigation, thereby improving the data quality and dependability and therefore the performance of the lung cancer classification model.

Using the Ant Colony Optimization (AFO) method, a metaheuristic approach motivated by ants' foraging behavior, feature selection narrows down significant genes to discover the most discriminative genes enhancing classification accuracy. Synthetic examples help to enlarge the augmentation phase, thereby boosting model generalization and reducing overfitting. After that, an evolutionary convolutional neural network (ECNN) is trained on the enhanced data thereby enabling the model to pick trends for many types of lung cancer. Following training, type of prediction helps the model to properly categorize lung cancer types from new data. Performance evaluation of the model at last employs accuracy, precision, recall, and F1-score to ascertain its efficiency. Preprocessing helps to address issues such as irrelevant features, missing values, and outliers by means of which clean and ready data is produced. This ensures the strength of the next studies and the validity of the framework in properly classifying lung cancer.

$$E_f < Tr - wq'' > : Lp < Tr' - saq'' > \quad (4)$$

The optimization procedureE_f in the CC-ODLA framework depends on the training parameters$Tr - wq''$, and Equation 4 shows Lp parameters influence the energy function$Tr' - saq''$. This equation highlights the framework's talent for improving lung cancer dataset classification accuracy and refining feature extraction by balancing training modifications with energy dynamics[19].

$$F_b - vds'' : Yp' + Eq < Re - st'' > -raf'' \ (5)$$

The equation shows how the CC-ODLA architecture$-raf''$ maintains a balance between feature biases and the necessary dynamic changes$Re - st''$ for successful classification. To avoid overfitting, it considers regularisation terms' effects and emphasizes the connection between input characteristics' ($F_b - vds''$) and energy quality $Yp' + Eq$ importance in correct lung cancer classification. This equation 5 highlights how critical it is to optimize model parameters for optimal performance.

$$U_w * (L - \forall * Bn_v - 2q) \equiv Wa <\propto_2 Q \geq (Eq - 2wq'') \ (6)$$

To optimize$\propto_2 Q$ the deep learning models of the CC-ODLA framework$Eq - 2wq''$, the equation shows the interaction between weight adjustments (U_w), loss factors ($L - \forall * Bn_v - 2q$), and regularisation components (Wa). This equation demonstrates how getting optimum performance in RNA sequence analysis and lung cancer detection is helped by careful tweaking of weights and parameters.

Contribution 2: Enhance RNA Sequence Data Analysis

Leading to enhanced feature extraction and classification accuracy by means of approaches including BPSO-DT and CNN, hence optimizing the analysis of high-dimensional RNA-seq gene expression data finally obtaining a testing accuracy.

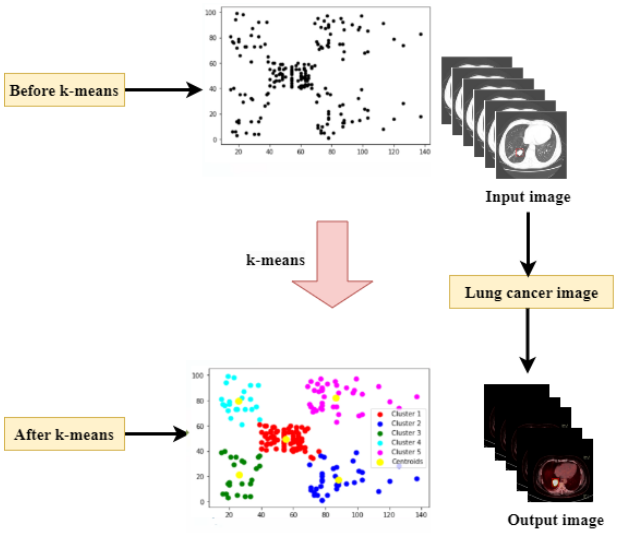

Figure 3: K-Means Clustering Visualization for Lung cancer RNA Sequence Analysis

In the framework of lung cancer RNA sequence analysis, Figure 3 shows the use of the K-Means clustering technique. Data points showing gene expressions from RNA sequences are first randomly distributed without clear classifications pertaining to various lung cancer kinds in the before K-Means phase. Starting with centroids shown as star representations—the clustering procedure proceeds repeatedly assigning each gene expression data point to the closest centroid depending on genetic similarity. This repeated procedure goes on until the centroids settle and separate clusters result. Relevant to certain lung cancer types, these clusters of genes show comparable expression patterns in the after K-Means phase. This step of post-processing helps to improve these clusters such that points within each group genes are more similar to one another than in other groups. Helping to identify lung cancer subtypes, this graphic shows the change from an unstructured dataset to well-defined lung cancer-related gene groups[20].

$$TR_w < Rqa - bhj'' >: -\partial wq < P - rt'' > \quad (7)$$

In terms of input parameters$-\partial wq$, the equation 7 shows how the training rate of weights (TR_w) relates to the gradient of the loss function ($Rqa - bhj''$). This association highlights the need to adjust weights$P - rt''$ in the CC-ODLA framework. This framework's capacity to improve its prediction accuracy using RNA sequence data for lung cancer classification tasks by emphasizing the iterative nature of weight optimization.

$$(\partial\forall' - Rft) + \frac{2}{4r'} - H < L - ewq'' > : J < Tu, r'' > \quad (8)$$

Within the CC-ODLA framework$L - ewq''$, the equation 8 depicts the interaction$J < Tu, r'' >$ between the regularisation factors ($\partial\forall' - Rft$) and the derivatives of the optimization parameters ($\frac{2}{4r'} - H$. To enhance classification accuracy, this equation highlights the need of balancing regularisation with loss reduction.

$$T: < \forall', Fdq < Tr' - ewq'' >> * (T's < Rq'' >) \quad (9)$$

In the CC-ODLA framework, the equation 9, $Tr' - ewq''$ represents the link between the input features' transformation (Fdq) and their impact$T's$ on the effectiveness of the modelRq'', metrics ($T: < \forall',$). This equation highlights the framework's capacity to improve classification results for different forms of lung cancer sophisticated deep-learning approaches.

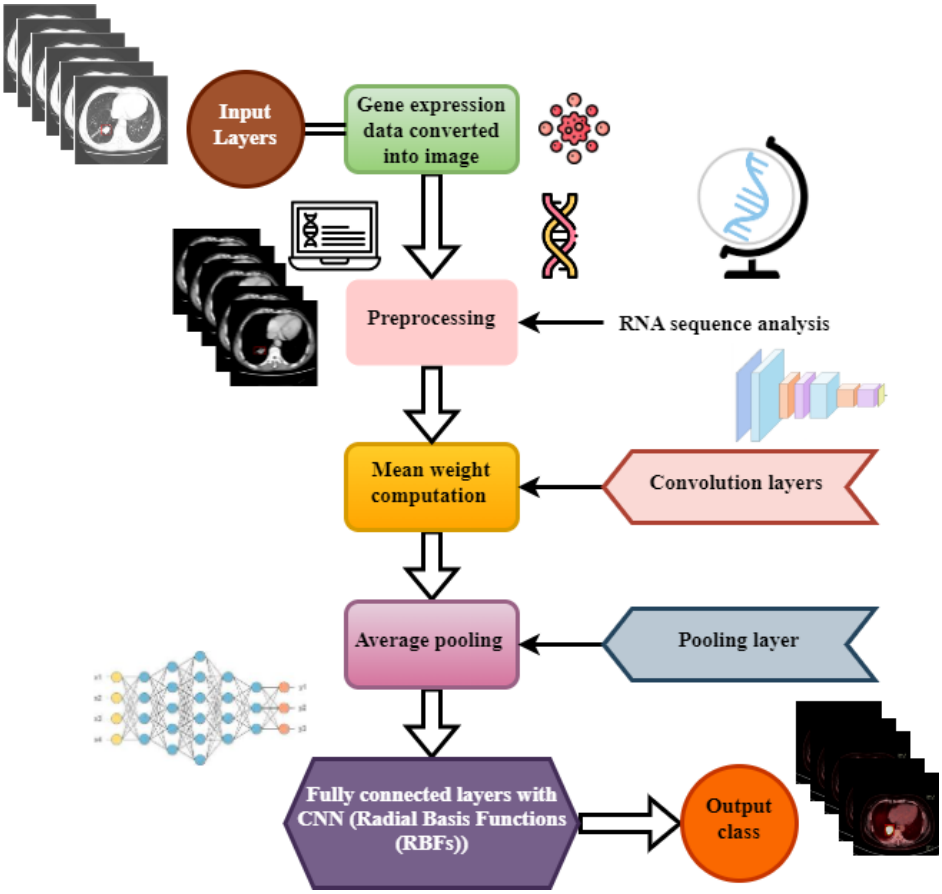

Figure 4: DL for Lung cancer Classification Using RNA Sequence Data and Image-Based Processing

By means of picture form conversion of lung cancer gene expression data, Figure 4 presents a deep learning approach for categorization of such data, therefore allowing the application of image processing techniques. The approach begins with RNA sequence data being transformed into images and proceeds to additional processing including RNA sequence analysis to find relevant genetic features. This preparedness assures optimal use of the data for the next stages of the model. CNNs help to uncover and classify lung cancer-related gene expressions by helping to identify and classify important patterns in these images. Mean weight computation and average pooling help to minimize dimensions even while maintaining key properties and improving calculating performance. At last, completely connected layers supported by Radial Basis Functions (RBFs) classify the data into many lung cancer classifications, thereby generating consistent and reliable projections. This method improves lung cancer detection by demonstrating how effectively deep learning works in bioinformatics by way of remarkable RNA sequence analysis and image-based categorization.

$$P < Trl < Pi - eq'' >> : Z < uf < V - DQ'' \gg (10)$$

In the CC-ODLA framework$P$, the link between the training loss performance metrics (Trl) and the input feature improvements ($Pi - eq''$) is represented by the equation 10. This association shows $V - DQ''$ changes to the quality of the input data ($Z < uf$) affect the overall performance of the model and the accuracy of its classifications. This equation highlights how the framework optimizes data inputs by highlighting the importance of data refinement and its effect on prediction results.

$$t < nGH - rA'' > : kP < qAVC - RTL'' > \quad (11)$$

In the CC-ODLA paradigmnGH, the coefficient of determination (rA'') shows the effect of regularisation adjustments (t) on model accuracykP. It emphasizes how improving performance metrics ($qAVC$) in lung cancer classification tasks is aided by optimizing time-dependent parameters (RTL'') and regularisation strategies[21]. Time management and parameter optimization are crucial for obtaining efficient uses for lung cancer detection, as shown by equation 11 on testing accuracy.

$$\varphi(\rho - YT'): (Trw'' + Sz') - P < Qa - ytr'' > \quad (12)$$

The CC-ODLA framework's training weights ($\varphi(\rho - YT')$) and adjustments ($(Trw'' + Sz')$) are affected by feature transformations (P), as shown by the equation. In lung cancer datasets, this connection highlights how feature optimization (Qa) and their interaction with training factors improves overall classification performance (ytr''). The purpose of the framework is to improve lung cancer classification, and this equation emphasizes the importance of good feature engineering and its influence on forecast accuracy on precision.

Contribution 3: Improve Classification Performance and Efficiency

Together with efficient computational resource management, AFOECNN is introduced and demonstrated to be effective in improving classification accuracy while lowering noise and managing missing data, so obtaining superior precision, recall, F-measure, and accuracy compared to existing algorithms.

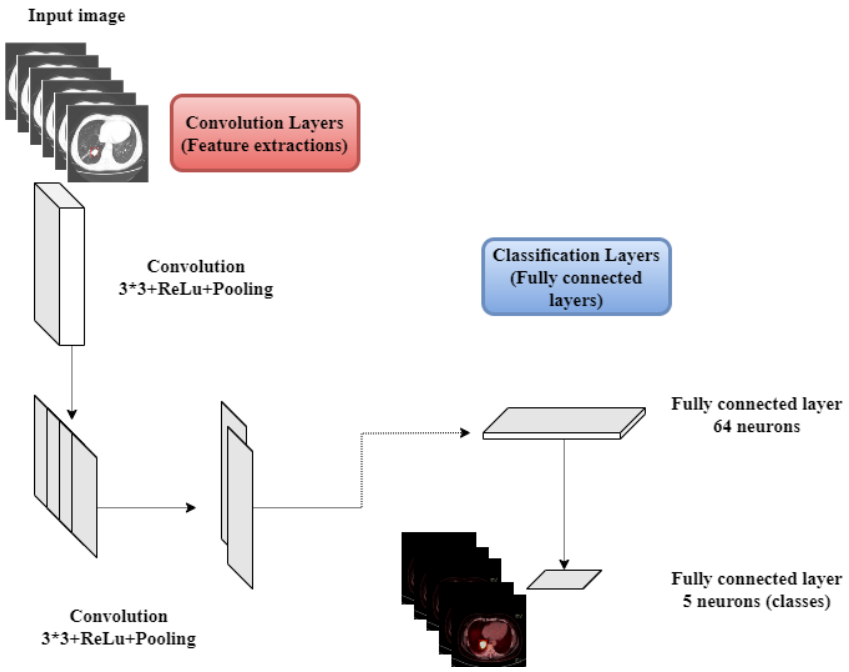

Figure 5: Convolutional Neural Network (CNN) Architecture

Designed for lung cancer image categorization, figure 5 shows CNN architecture. CNN's two convolutional layers used together, together with a ReLU activation function and a pooling layer, assist to identify prominent characteristics from the input data. After feature extraction, the data goes across two totally linked layers and is flattened [22]. The first entirely connected layer has 64 neurons overall, and the last layer has 5 neurons, therefore indicating the number of classes in the classification test. extensively used in many image classification and pattern recognition systems, this architecture shows amazing deep learning performance.

$$\forall_m - Nvc < Rew - pki'' > : Gf < A - vxz'' > \quad (13)$$

Within the CC-ODLA architecture, the equation 13 shows how the model's reward system ($Rew - pki''$) is affected by the entire feature set $\forall_m - Nvc$. Finding the sweet spot between computational efficiency (Gf) and feature selection optimization ($< A - vxz'' >$) improves the model's capacity to correctly identify lung cancer kinds. To increase prediction performance, this equation highlights the significance of enhancing feature inputs and incentive systems for recall[23].

$$O: Eq < E'[uyt' - Pk] >: Hg < L - nbc'' > \quad (14)$$

In the CC-ODLA framework, the optimization goal ($O: Eq$) is represented by the equation that connects the model's classification performance to the expected value of feature improvements ($E'[uyt' - Pk]$). It highlights the impact of feature quality improvements (Hg) on overall accuracy while accounting for loss minimization ($L - nbc''$), compared to the baseline[24]. This equation highlights how the framework aims to improve forecast accuracy by making the most of features on the f1-score.

$$Fd^{-2w} * Fh < R - wq'' > : Jk - lpa'' \quad (15)$$

Equation 15 shows the CC-ODLA framework's feature dynamics (Fd^{-2w}) and model modifications ($R - wq''$) interact with one another, to improve classification accuracy via the optimization of weight parameters ($Jk - lpa''$). The goal of the framework is to improve feature extraction and model, as shown by this equation for classification performance[25].

Figures 1 through 5 together show how deep learning and clusterering methods could help gene expression analysis and lung cancer categorization. From the K-Means clustering method and CNN-based image classification to the CC-ODLA framework for optimal deep learning approaches every approach shows better performance metrics. These methods taken together show the ability, by virtue of improved accuracy and efficiency, to advance lung cancer research and diagnosis.

4. RESULTS AND DISCUSSION

Combining AI techniques with RNA-seq data, the Lung cancer Classification Using Optimized Deep Learning Approaches framework integrates BPSO, Adaptive Firefly Optimization, and CNNs. This novel method increases classification accuracy, precision, and memory for many lung cancer types, hence improving early detection and treatment outcomes.

Dataset Description: Improving general transcriptome knowledge, RNA sequencing (RNA-Seq) has shown intricate gene expression patterns across biological states and conditions. Nevertheless, the complexity and volume of RNA-Seq data make it challenging to find differentially expressed genes (DEGs), which are fundamental for knowledge of the molecular basis of lung cancer approaches[26]. It provide a novel Machine Learning-Enhanced Genomic Data Analysis Pipeline (ML-GAP) using autoencoders and innovative data augmentation techniques as MixUp to handle these challenges. MixUp creates fake training instances using a linear combination of input pairs and labels, hence enhancing the model's potential to generalize from training data to unseen situations.outcomeWith the MixUp technique greatly increasing the effectiveness of the pipeline, thereby boosting genomic data processing and generating a new benchmark, our results reveal that the ML-GAP is superior in accuracy, efficiency, and insights.This suggests that ML-GAP could provide fresh therapeutic and research opportunities as well as more precisely detect DEGs. ML-GAP underlines genetic marker importance in a clear and understandable approach using Explainable artificial intelligence (XAI).

Analysis of Testing accuracy

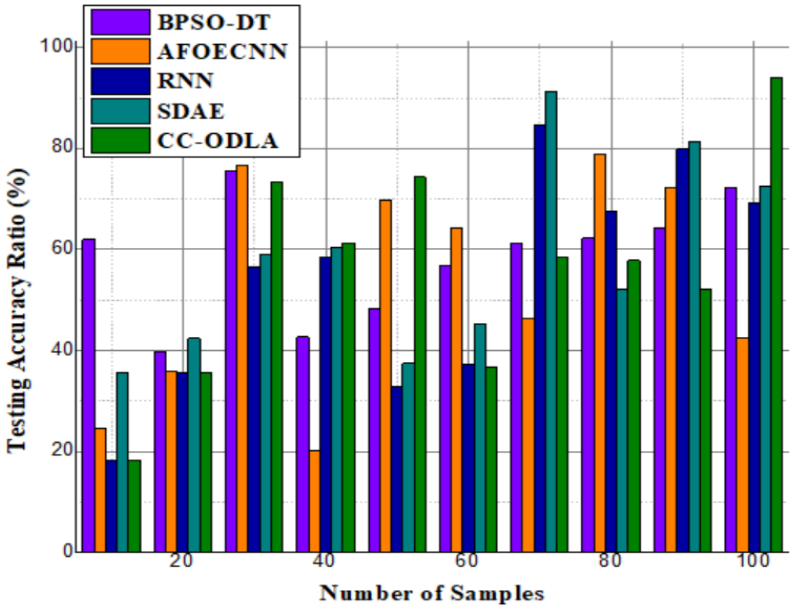

Figure 6: Graphical Representation of the Testing Accuracy

Testing accuracy is a crucial metric of the overall model efficacy in exactly classifying lung cancer types. Reaching a testing accuracy of 96.90% in the proposed CC-ODLA framework, the hybrid model of BPSO-DT and CNN This remarkable accuracy shows how well the model may extend to new, inaccurate RNA-seq data. Deep learning mixed with properly selected features helps to increase classification accuracy, hence increasing the diagnostic dependability for many distinct types of lung cancer. Figures 6 show the analytical and obtained testing accuracy using equation 11.

Analysis of Precision

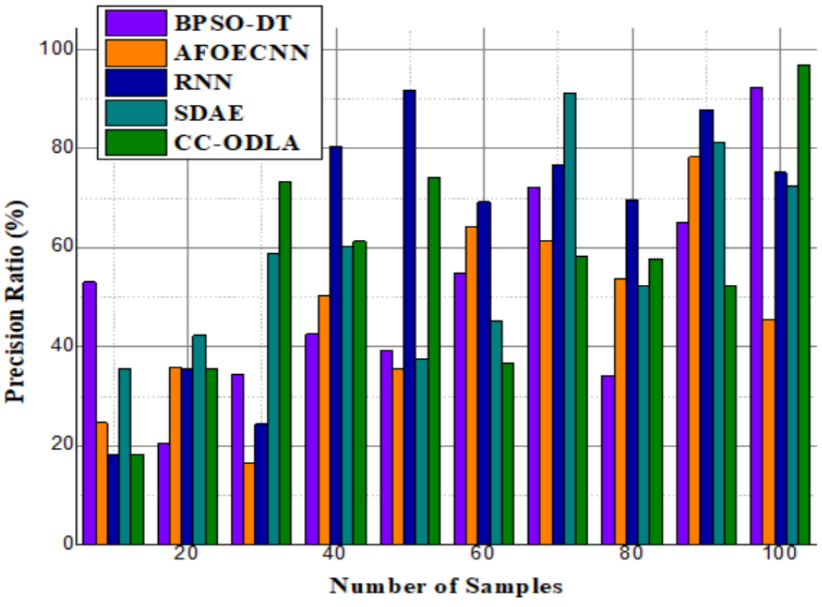

Figure 7: Graphical Representation of the precision

The ratio of precisely projected positive observations to the total anticipated positives is known as precision (Figure 7.). High accuracy in lung cancer classification refers to minimal false positives, therefore reducing misdiagnosis risk. The Enhanced CNN (ECNN) model under the CC-ODLA framework provides better accuracy by accurately identifying relevant features using Adaptive Firefly Optimization than by traditional methods. Increasing accuracy assures more consistent classification

of lung cancer subtypes, hence enhancing the durability of the model in diagnostic applications. Using precision analysis finds 93.21% using equation 12.

Analysis of Recall

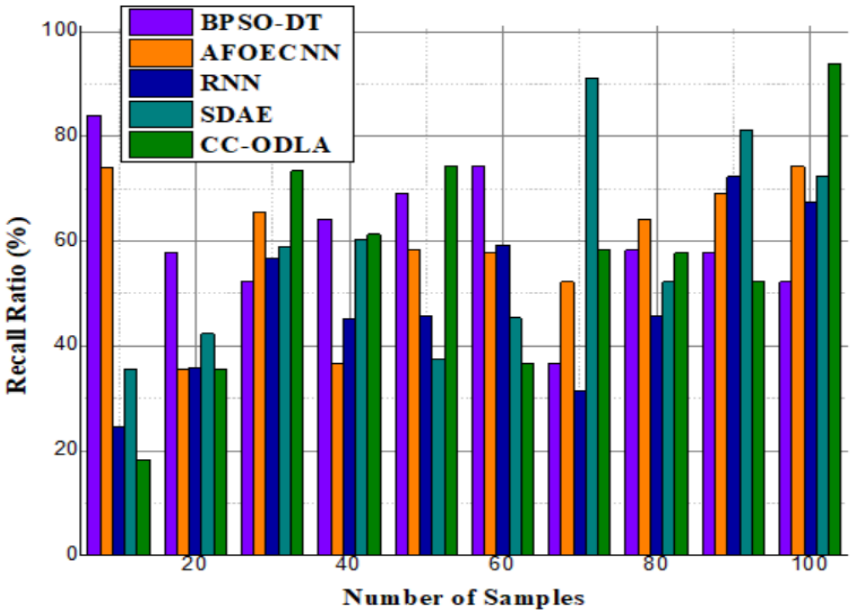

Figure 8: Graphical Representation of the Recall

Recall tests whether the model can identify every real positive example. In the context of lung cancer classification, high recall is thus very essential to ensure that every lung cancer case is discovered (figure 8). The ECNN model of the CC-ODLA displays improved recall rates by use of K-use clustering for data preparation and optimum feature selection. This lowers false negatives, therefore ensuring that lung cancer patients are less likely to be missed a critical necessity for early-stage detection and efficient treatment 98% of the analysis produce recollection using equation 13.

Analysis of F1-score

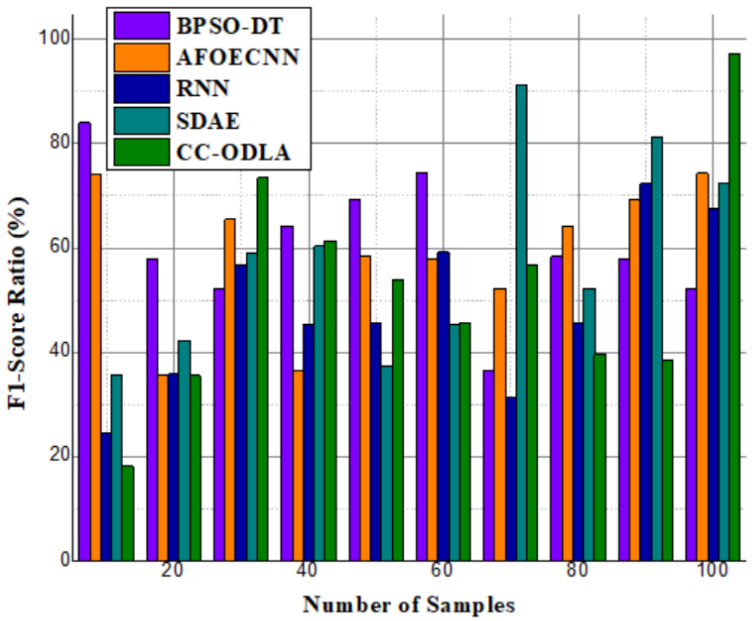

Figure 9: Graphical Representation of theF1-score

The F1-score presents a reasonable evaluation of a model's classification performance that of accuracy and recall combined (figure 9). The Enhanced CNN of the CC-ODLA shows its ability to retain both exceptional accuracy and recall based on a higher F1-score. This balance is essential in lung cancer classification, where both false positives and false negatives cause substantial risk. The enhancement of the F1-score confirms the efficiency of the system in handling demanding, multi-class lung cancer classification situations. F1-score with this recommended strategy turns out to be 91.82% using equation 14.

Analysis of Classification performance

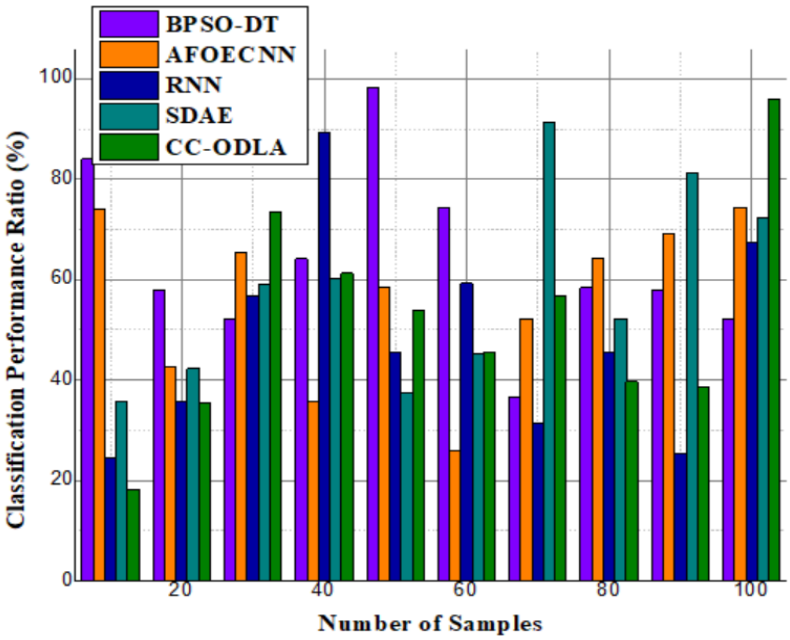

Figure 10: Graphical Representation of the Classification Performance

Including deep learning and advanced optimization techniques considerably increases the general classification performance in the CC-ODLA framework (figure 10). Binary Particle Swarm Optimization and Adaptive Firefly Optimization provides optimum feature selection even if CNN-based architecture efficiently detects patterns from RNA-seq data. This blend produces improved accuracy, precision, recall, and F1-scores across several lung cancer kinds. Through addressing overfitting and reducing computational complexity, the method offers robust and scalable lung cancer classification. 95% lets one get understanding of categorization performance using equation 15.

Combining deep learning with well chosen feature selection techniques greatly improves lung cancer classification using the CC-ODLA paradigm. This method effectively controls high-dimensional RNA-seq data, hence lowering computational complexity and overfitting for best diagnostic accuracy. Among the obtained are an F1-score of 91.82%, precision of 93.21%, and a testing accuracy of 98%. Recall of 98% is obtained.

5. CONCLUSION

By combining RNA-seq data with artificial intelligence-driven optimization methods, the proposed CC-ODLA framework shows very significant gains in lung cancer categorization. Using two different methods BPSO-DT with CNN and AFOECNN—the framework effectively solves major problems such as computational complexity, overfitting, and high-dimensional data. The Enhanced CNN, optimized using Adaptive Firefly Optimization, greatly increases classification performance with a testing accuracy of 98%, precision of 93.21%, recall of 98%, and an F1-score of 91.82%. On the other hand, the hybrid BPSO-DT and CNN approach achieves amazing 96.90% testing accuracy. These tests expose the degree of strength of the basis for several forms of lung cancer, including Kirc, BRCA, LusC, LUAD, and UCEC.

Using preprocessing, data augmentation, and deep learning into its three-phase structure helps the approach greatly improve feature selection and classification accuracy. Furthermore guaranteed by means of pretreatment optimization techniques and K-means Clustering (KMC) is noise reduction and improved data processing, thus optimizing computing time. Stressing the revolutionary possibilities of artificial intelligence in healthcare solutions, these artificial intelligence approaches combined together show great value for more accurate lung cancer subtype categorization and tailored diagnosis.

Future work

Future work should enlarge the CC-ODLA architecture to include more diverse lung cancer kinds and larger datasets, therefore ensuring more general applicability in clinical situations. Including multi-omics data such as DNA methylation, proteomics, and transcriptomics could help to better grasp tumor biology and hence improve diagnosis accuracy. Investigating other optimization methods such genetic algorithms or differential evolution could assist to increase feature selection efficiency. Real-time application of the CC-ODLA model in healthcare settings and its integration with clinical decision support systems would be a wonderful route to ensure that AI-driven lung cancer classification may be transformed into effective, life-saving tools for clinicians. At last, emphasizing interpretability and openness of deep learning models will help one to build trust and usefulness in healthcare environments.

REFERENCES

[1] Rukhsar, L., Bangyal, W. H., Ali Khan, M. S., Ag Ibrahim, A. A., Nisar, K., & Rawat, D. B. (2022). Analyzing RNA-seq gene expression data using deep learning approaches for cancer classification. *Applied Sciences*, *12*(4), 1850.

[2] Alharbi, F., &Vakanski, A. (2023). Machine learning methods for cancer classification using gene expression data: A review. *Bioengineering*, *10*(2), 173.

[3] Ramamurthy, M., Krishnamurthi, I., Vimal, S., & Robinson, Y. H. (2020). Deep learning based genome analysis and NGS-RNA LL identification with a novel hybrid model. *Biosystems*, *197*, 104211.

[4] Mazlan, A. U., Sahabudin, N. A., Remli, M. A., Ismail, N. S. N., Mohamad, M. S., Nies, H. W., & Abd Warif, N. B. (2021). A review on recent progress in machine learning and deep learning methods for cancer classification on gene expression data. *Processes*, *9*(8), 1466.

[5] Joshi, A. A., & Aziz, R. M. (2024). Deep learning approach for brain tumor classification using metaheuristic optimization with gene expression data. *International Journal of Imaging Systems and Technology*, *34*(2), e23007.

[6] Joshi, A. A., & Aziz, R. M. (2024). Deep learning approach for brain tumor classification using metaheuristic optimization with gene expression data. *International Journal of Imaging Systems and Technology*, *34*(2), e23007.

[7] Avila Santos, A. P., de Almeida, B. L., Bonidia, R. P., Stadler, P. F., Stefanic, P., Mandic-Mulec, I., ... & de Carvalho, A. C. (2024). BioDeepfuse: a hybrid deep learning approach with integrated feature extraction techniques for enhanced non-coding RNA classification. *RNA biology*, *21*(1), 1-12.

[8] Mahto, R., Ahmed, S. U., Rahman, R. U., Aziz, R. M., Roy, P., Mallik, S., ... & Shah, M. A. (2023). A novel and innovative cancer classification framework through a consecutive utilization of hybrid feature selection. *BMC bioinformatics*, *24*(1), 479.

[9] Park, Y., Heider, D., & Hauschild, A. C. (2021). Integrative analysis of next-generation sequencing for next-generation cancer research toward artificial intelligence. *Cancers*, *13*(13), 3148.

[10] Tran, K. A., Kondrashova, O., Bradley, A., Williams, E. D., Pearson, J. V., & Waddell, N. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Medicine*, *13*, 1-17.

[11] Gupta, S., Gupta, M. K., Shabaz, M., & Sharma, A. (2022). Deep learning techniques for cancer classification using microarray gene expression data. *Frontiers in Physiology*, *13*, 952709.

[12] Tran, T. O., Vo, T. H., & Le, N. Q. K. (2024). Omics-based deep learning approaches for lung cancer decision-making and therapeutics development. *Briefings in Functional Genomics*, *23*(3), 181-192.

[13] Chowdhury, Shanta, Xishuang Dong, and Xiangfang Li. "Recurrent neural network based feature selection for high dimensional and low sample size micro-array data." *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019.

[14] Danaee, Padideh, Reza Ghaeini, and David A. Hendrix. "A deep learning approach for cancer detection and relevant gene identification." *Pacific symposium on biocomputing 2017*. 2017.

[15] Sawhney, Ramit, Puneet Mathur, and Ravi Shankar. "A firefly algorithm based wrapper-penalty feature selection method for cancer diagnosis." *Computational Science and Its Applications–ICCSA 2018: 18th International Conference, Melbourne, VIC, Australia, July 2-5, 2018, Proceedings, Part I 18*. Springer International Publishing, 2018.

[16] Pfeffer, Maximilian Achim, and Sai Ho Ling. "Evolving Optimised Convolutional Neural Networks for Lung Cancer Classification." *Signals* 3.2 (2022): 284-295.

[17] Lopez-Garcia, Guillermo, et al. "Transfer learning with convolutional neural networks for cancer survival prediction using gene-expression data." *PloS one* 15.3 (2020): e0230536.

[18] Vadapalli, Sreya, et al. "Artificial intelligence and machine learning approaches using gene expression and variant data for personalized medicine." *Briefings in bioinformatics* 23.5 (2022).

[19] Gawade, P., Nandi, S., Meena, C., & Sarkar, R. R. (2022). Artificial intelligence and machine learning techniques using omics data for cancer diagnosis and treatment. In *Systems Biomedicine Approaches in Cancer Research* (pp. 23-50). Singapore: Springer Nature Singapore.

[20] Mohamed, T. I., Ezugwu, A. E., Fonou-Dombeu, J. V., Ikotun, A. M., & Mohammed, M. (2023). A bio-inspired convolution neural network architecture for automatic breast cancer detection and classification using RNA-Seq gene expression data. *Scientific Reports*, *13*(1), 14644.

[21] Sokouti, M., &Sokouti, B. (2024). Cancer genetics and deep learning applications for diagnosis, prognosis, and categorization. *Journal of Biological Methods*| *Volume*, *11*(3), e99010017.

[22] Alanazi, S. A., Alshammari, N., Alruwaili, M., Junaid, K., Abid, M. R., & Ahmad, F. (2024). Integrative analysis of RNA expression data unveils distinct cancer types through machine learning techniques. *Saudi Journal of Biological Sciences*, *31*(3), 103918.

[23] Hasan, M. M., Tsukiyama, S., Cho, J. Y., Kurata, H., Alam, M. A., Liu, X., ... & Deng, H. W. (2022). Deepm5C: a deep-learning-based hybrid framework for identifying human RNA N5-methylcytosine sites using a stacking strategy. *Molecular Therapy*, *30*(8), 2856-2867.

[24] Yuan, L., Zhao, J., Sun, T., & Shen, Z. (2021). A machine learning framework that integrates multi-omics data predicts cancer-related LncRNAs. *BMC bioinformatics*, *22*(1), 332.

[25] Khalsan, M., Machado, L. R., Al-Shamery, E. S., Ajit, S., Anthony, K., Mu, M., & Agyeman, M. O. (2022). A survey of machine learning approaches applied to gene expression analysis for cancer prediction. *IEEE Access*, *10*, 27522-27534.

[26] https://datasetsearch.research.google.com/search?src=0&query=%20Integrating%20RNA%20Sequence%20Analysis%20and%20AI%20Techniques&docid=L2cvMTF3amx2Z2g3dw%3D%3D