Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Predicting Lung Cancer Death Rates with a Novel Deep Learning Approach to Assist Medical Decision-Making

Suhas Gupta[1], Raviraj Singh Kurmi[2], Aseem Purohit[3], Sourav Rampal[4], Lakshi Goswami[5], Dr. Moovendhan V[6]

[1]Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India suhas.gupta.orp@chitkara.edu.in https://orcid.org/0009-0004-9791-2416
[2]Assistant Professor, Department of CSE(Data Science), Noida Institute of Engineering and Technology, Greater Noida, Uttar Pradesh, India, Email Id- raviraj.singh@niet.co.in, Orcid Id- 0009-0005-1122-1700
[3]Professor of Practice, Department of Management Studies, Vivekananda Global University, Jaipur, India, Email Id- aseem.purohit@vgu.ac.in, Orcid Id- 0009-0004-6312-8121
[4]Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh-174103 India sourav.rampal.orp@chitkara.edu.in https://orcid.org/0009-0000-6270-102X
[5]Assistant Professor, Department of Electrical Engineering, Sanskriti University, Mathura, Uttar Pradesh, India, Email Id- laxmigoswami.ee@sanskriti.edu.in, Orcid Id- 0000-0003-3762-4389
[6]Associate Professor, Department of Business Analytics, JAIN (Deemed-to-be Univesity), Bangalore, Karnataka, India, Email Id- dr.moovendhan_v@cms.ac.in, Orcid Id- 0000000228137717

Introduction: The aim of this proposed approach Elevating Naïve Bayes Optimization with Improved Convolutional Neural Networks (ENBO-ICNN), is to predict lung cancer (LC) death rates while considering the complexities and ethical concerns inherent in the function of deep learning (DL) algorithms in the healthcare domain.
Objective: To address the limitations of data quality, disease complexity and the evolving nature of research in forecasting LC death rates by implementing the ENBO-ICNN approach.
Method: The study gathered data from the NCRP dataset, preprocessing using Min-Max normalization and employing Principal Component Analysis (PCA) for feature extraction to develop the ENBO-ICNN approach for predicting LC mortality rates.
Result: The efficacy of the ENBO-ICNN technique is demonstrated with its enhanced performance measures, which include significant recall, accuracy, precision, and F1-score in improving LC death rate predictions compared to existing methods.
Conclusion: The proposed ENBO-ICNN approach signifies a significant advancement in predicting LC death rates, offering valuable insights into the complexities of the disease and addressing critical

ethical considerations associated with DL algorithms in healthcare.

Keywords: Lung Cancer (LC), Death Rates, Deep Learning (DL), Improved Convolutional Neural Networks (ICNN).

1. Introduction

Lung cancer (LC) remains a significant obstacle for healthcare programs all around the world despite the considerable progress that should be made. According to the World Health Organization (WHO), which claims that 1.8 million people pass away each year from LC, the disease is a significant public health hazard.[1] Despite the considerable breakthroughs in medical technology and treatment methods that have occurred over the last several years, the prognosis for many patients who have LC is relatively poor. The predominance of adverse outcomes [2] can attribute mainly to late-stage diagnosis and lack of appropriate prognostic tools. There is reason for renewed confidence in battle against LC because of recent technological advances, such as predictive analytics and machine learning(ML). [3-4] The current status of LC treatment might be altered by combining meticulous data analysis and predictive modeling. It can improve treatment plans, personalize screening processes, as well as lessen the social and financial toll that this crippling disease exacts to create a workable plan. [5]

The study [6] provided a combination of LDA and Optimal Deep Neural Networks (ODNNs) for Evaluating CT Lung Images. The research [7] determined the use of various ML classifiers to investigate whether or not data from the UCI ML repository could be used to categorize LC cases as benign or malignant reliably. The study [8] adopted a novel technique based on multi-omics data to estimate the prognosis of LC patients accurately. The article [9-10] examined over 65 studies that use ML techniques to forecast various illnesses. Future LC diagnosis methods must be developed utilizing medical IoT and the investigation primarily focuses on different ML algorithms used to detect multiple diseases. The work [11-12] utilized a pioneering interdisciplinary technique deployed initially to LC, using a combination of metabolomics and ML methods to find diagnostic biomarkers for early-stage LC. The study [13] presented a DL strategy based on the Adaptive Hierarchical Heuristic Mathematical Model (AHHMM). Automated radiation adaptation techniques have been developed for NSCLC. The research [14] described DL-powered survival analysis system. There are three primary parts to the proposed method. The research [15] verified a model for predicting OS in LUAD. They are recruited from various groups of LUAD patients after applying ML algorithms to 492 patients' data from Cancer Genome Atlas (CGA).

The study [16] provided a deep neural network (DNN) created by combining disparate collections of laboratory and genetic information for accurate prognosis of non-small cell LC (NSCLC) patients' overall survival. The article [17] utilized CNN architecture that they've dubbed Deep LRHE to evaluate histo-pathological images of patients to predict the probability of LC recurrence. The study [18] assessed the role of HRQOL in model for predicting LC patient's odds of survival five years after diagnosis using ML techniques (MLTs). The study [19-20] provided a novel multi-pronged deep learning-based strategy for

determining whether or not nodules are cancerous. They have constructed two deep 3D CMixNet architectures for nodule recognition and classification in the lungs, inspired by recent achievements in image processing using deep CNN.

2. Methodology

The study proposes an ENBO-ICNN to forecast LC mortality rates, focusing on data quality and disease complexity. NCRP dataset data is collected and normalized using Min-Max and PCA is used to extract data features. Figure 1 shows the outline of our proposed work.

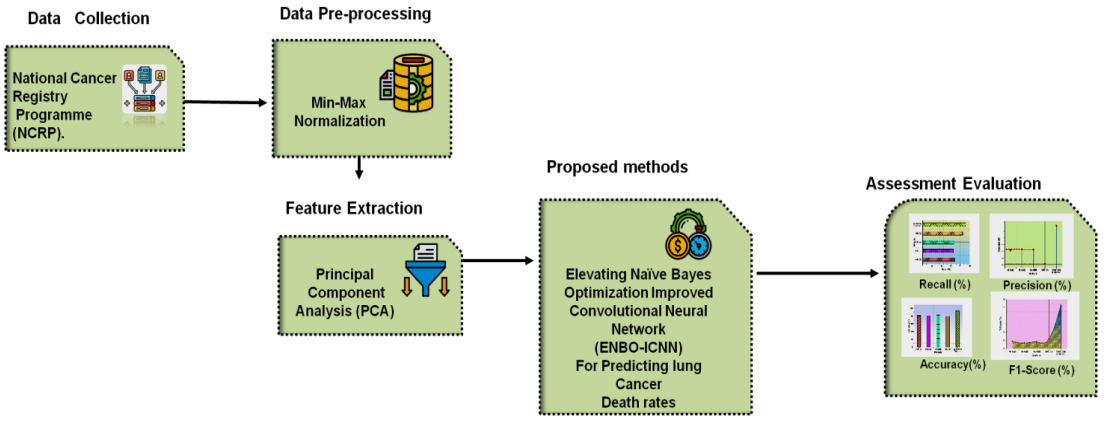

Figure 1. Outline of the proposed methods (Source: Author)

Dataset

The majorities of the recent NCRP data spans a five-year (2012-2016) reporting period and were compiled from 58 hospital-based cancer registries (HBCRs) and 28 population-based cancer registries (PBCRs). The 28 PBCRs have been divided into the following regions: north, east, central, west, south, and northeast regions (NER). An HBCR, established at a specific medical facility, compiles statistics regarding cancer patients who obtain care through various healthcare specialties. They include essential details on the conditions of diagnostic manifestation, available treatments, and its prognoses, including mortality and survival.[21] Considering the requirements established by the International Classification of Illnesses for Cancer, the NCRP records tumors with a behavior code of 3.

Data preprocessing

Preprocessing data is crucial for standardizing it in classification, accelerating learning and addressing numerical issues like precision loss due to human computation mistakes. Normalization is essential for gradient calculations, where a broader quality range is preferred over a smaller one. This process is necessary for ensuring accurate classification as shown in (Equation 1).

$$w = \min_{new} + (\max_{new} - \min_{new}) * \left(\frac{w - \min_x}{\max_x - \min_x}\right) \quad (1)$$

This method's power comes from its dependably maintaining the relationships between the different data pieces. There will be no misleading effect on the final result.

Feature Extraction

PCA is a linear technique that reduces dimensionality by projecting data in the direction of the most variability, starting with raw data and sorted eigenvectors.

Step 1: Samples from a signal M × Min M dimensions can have their covariance matrices calculated as (Equation 2):

$$\Sigma = \frac{1}{2}\{(w - \overline{w})\ (w - \overline{w})^{S}\} \quad (2)$$

Where, w is the provided signal matrix with N data points of size M and w is the mean vector.

Step 2: Determine the covariance matrix Q using the eigenvectors and diagonal components of matrix E as provided by (Equation 3)

$$U^{-1}\Sigma U = C \quad (3)$$

Step 3: The eigenvectors of C should be sorted from highest to lowest Eigenvalue.

Step 4: It's possible to convert the data into a vector space by taking the dot product of the input data and the ordered Eigenvectors.

Step 5: Based on proportion of variance they explain, choose the first few significant components.

Improved Convolutional neural network (ICNN)

The ICNN accurately predicts LC mortality by extracting complex spatial information from 3D lung scans and analyzing imaging data to improve prognosis by addressing geographical links and subtle patterns as shown in Figure 2.

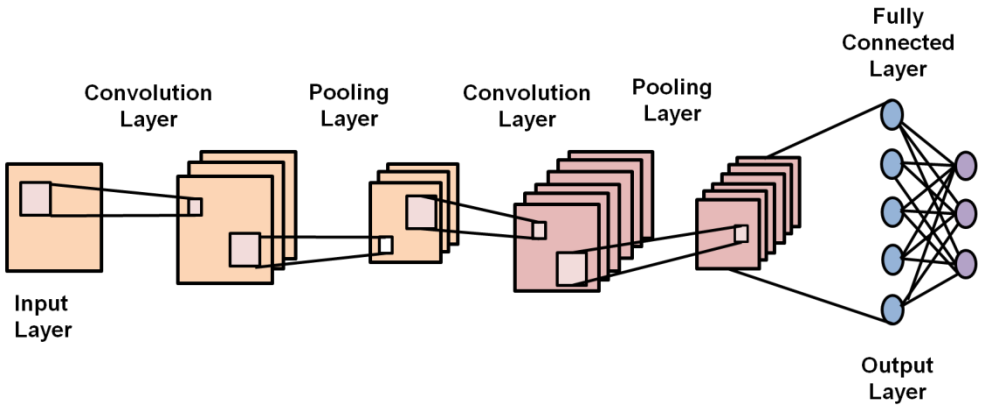

Figure 2.Architecture for ICNN (Source: https://www.researchgate.net/figure/Schematic-diagram-of-a-basic-convolutional-neural-network-CNN-architecture-26_fig1_336805909)

Mini-batch training is employed to handle the enormous quantities of data needed to train a CNN. Let $W = [w_1, \ldots., W_N]$ To represent the K-dimensional activation features of an M-image training set and w_j is $Land z = [z_1, \ldots., z_M], z_j \in \{-1,1\}$ to represent a vector holding the truth labels. The boosting approach uses a robust classifier, $G(\cdot)$, the weighted sum of weak classifiers, $g(\cdot)$ to produce the prediction as shown in (Equation 4).

$$G(W_j) = \sum_{i=1}^{L} \alpha_i \, g(w_{ji,}\lambda_i); g(w_{ji,}\lambda_i) = \frac{e(w_{ji,}\lambda_i)}{\sqrt{e(w_{ji,}\lambda_i)^2+\eta^2}} \quad (4)$$

The $w_{ji} \in w_j$ feature of image j is lower than the i^t activation feature. The weak classifier $g(w_{ji,}\lambda_i)$ stands for each component and its result can be anywhere from $(-1,1)$ $\frac{e(\cdot)}{\sqrt{e(\cdot)^2+\eta^2}}$ is used to compute the derivative of sign approximatively $(\cdot)$ function to optimize using gradient descent. In this work, we provide $(w_{ji,}\lambda_i)^2 \in r$, a typical AdaBoost function described as a decision tree with a single node (a decision stump) and a threshold of λ_i. The parameter is stated $e(\cdot)$ as $\eta = \frac{\sigma}{d}$, where is the standard deviation of $e(\cdot)$ and d is a constant supplied by the distribution of $e(\cdot)$ and defines the slope η of the function $\frac{e(\cdot)}{\sqrt{e(\cdot)^2+\eta^2}}$ in (Equation 5-7).

$$\epsilon^A = \beta\varepsilon^A_{strong} + (1-\beta)\epsilon_{weak} \quad (5)$$

$$\varepsilon^A_{strong} = \frac{1}{N}\sum_{j=1}^{M}(G(W_j) - z_j)^2 \quad (6)$$

We get the weak-classifier loss when the weak classifiers' losses are added together.

$$\varepsilon_{weak} = \frac{1}{MN}\sum_{j=1}^{N}\sum_{\substack{1\le i\le l \\ \alpha_i>0}} [g(w_{ji,}\lambda_i) - z_i]^2 \quad (7)$$

When the loss does not include dormant neurons due to the limitation $\alpha_i > 0$, the active neurons' weights as well as thresholds, among other things, are reset and refreshed with new Information.

Elevating Naïve Bayes Optimization (ENBO)

Thomas Bayes invented ENBO. The Bayesian optimization technique in (Equation 8-9) estimates class membership probability using the Bayes theorem.

$$O(Z\backslash W) = \frac{o(W\backslash Z)o(z)}{o(w)} \quad (8)$$

W = Information without a class

Z = Data premise W pertains to a particular class

$O(Z\backslash W)$= the probability of hypothesis Z given condition W

$O(Z)$= Probability of belief Z

It should be noted that the classification process requires numerous suggestions to identify which class is suitable for the data this ENBO algorithm is examining in Equation (9-11).

$$O(Z_I|W_1, \ldots., W_m) = \frac{o(z)o(W_1,\ldots,W_m|Z_i)}{o(W_1,\ldots,W_m)} \quad (9)$$

$$\text{Posterior} = \frac{\text{prior likehood}}{\text{evidence}} \quad (10)$$

The evidence value remains constant for each sampled class. The sample class to be categorized will be determined by comparing the posterior value to other posterior class values. A multiplication rule can be used to build on the Bayesian formula$(Z_I|W_1, \ldots., W_m)$ as shown in Equation (11-13).

$$O(Z_I|W_1, \ldots., W_m) \quad (11)$$

$$O(W_j|Z_i) = \frac{o(w_j \cap w_i)}{o(w_j)} = \frac{o(w_j)o(w_i)}{o(w_j)} = o(w_j) \quad (12)$$

Since j ≠ i, we have

$$O(W_j|Z, W_i) = O(W_j|Z_i) \quad (13)$$

Since (Equation 12-16) above assumes naive independence, it's possible to argue that this simplifies the probability requirements, making the computation feasible. Furthermore, $O(W_j|Z, W_i) = O(W_j|Z_i)$can be expressed in a shorter form as:

$$O(Z_i|W_1, \ldots. W_n) = O(Z_i)O(W_1|Z_i)O(W_2|Z_i)O(W_3|Z_i) \ldots$$

$$O(W_m|Z_i) = (O(Z_i) \prod_{j=1}^{m} O(W_j|Z_i)) \quad (14)$$

Information:

$O(Z_i|Z_1, \ldots. W_m) =$ Backward Probability

$O(W_j|Z_i)$= possibility

$O(Z_i) =$ Advance Probability

$Z_{MAP} =$ The class with the highest probability of A posteriori

$$Z_{MAP} = \text{argmax}_{zi \in z}(O(Z_i) \prod_{j=1}^{m} O(W_j|Z_i)) \quad (15)$$

It's from the preceding equation, will be used in the classification procedure.

$$O(Z_i|Z_1, \ldots. W_m) = \text{PosteriorProbability} \quad (16)$$

3. Results

We provide an ENBO-ICNN to predict LC Death Rates. It performs better than other traditional methods such as random forest classification (RFC) [(22)], convolutional neural network (CNN) [(22)], recurrent neural network (RNN) [(22)] and artificial neural network (ANN) [(22)].

The accuracy of LC death rate estimate is how well-projected values match observed values. The graphical accuracy values are shown in Figure 3 and Table 1 ENBO-ICNN is higher than other existing methods.

Table 1. Numerical Outcomes of Accuracy (Source: Author)

Methods	Accuracy (%)
ANN [22]	82
RNN [22]	81
CNN [22]	83
RFC [22]	80
ENBO-ICNN (Proposed)	94

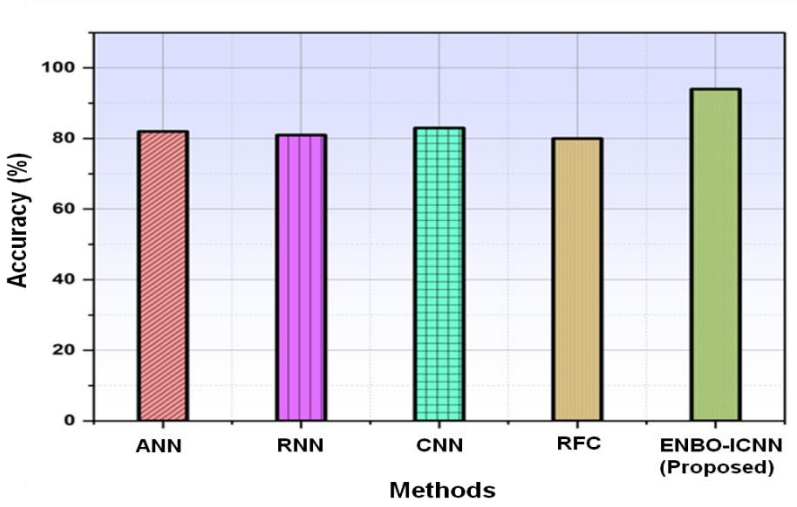

Figure 3. Graphical Outcomes of Accuracy (Source: Author)

Precision assesses a model's optimistic LC death rate predictions. Calculate the true positives or false positives ratio. Precision is the percentage of optimistic model predictions that succeed. Figure 4 and table 2 displays the exact precision results and, suggested method's ENBO-ICNN is higher than other existing methods.

Table 2. Numerical Outcomes of Precision (Source: Author)

Methods	Precision (%)
ANN [22]	70,52
RNN [22]	71,5
CNN [22]	71,06
RFC [22]	60,29
ENBO-ICNN (Proposed)	89

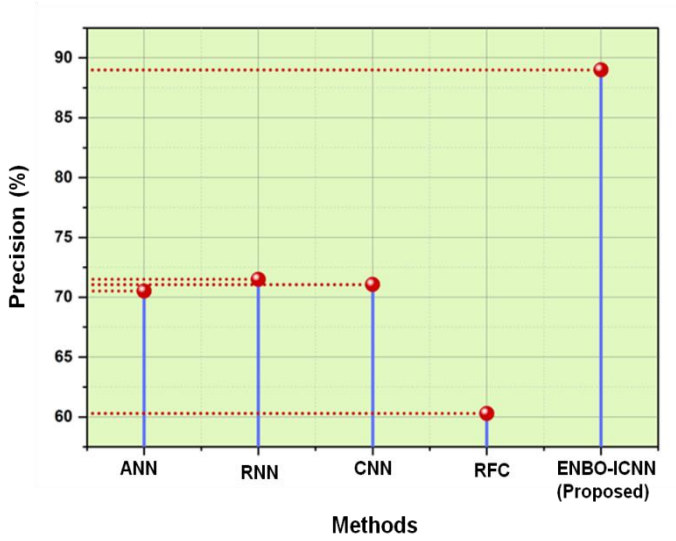

Figure 4. Graphical Outcomes of Precision (Source: Author)

Recall assesses a model's ability to predict the relevant deaths in a dataset for LC mortality rate. Figure 5 and table 3 demonstrates the exact recall results and, suggested method's ENBO-ICNN is higher than other existing methods.

Table 3. Numerical Outcomes of Recall (Source: Author)

Methods	Recall (%)
ANN [22]	68,39
RNN [22]	65,56
CNN [22]	67,09
RFC [22]	85,48
ENBO-ICNN (Proposed)	92

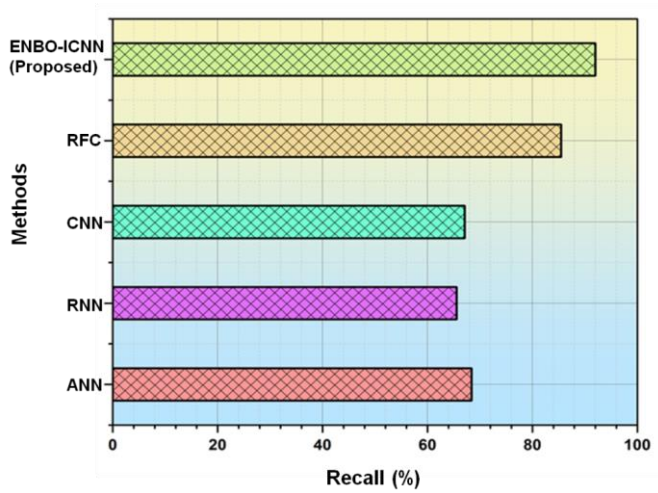

Figure 5. Graphical Outcomes of Recall (Source: Author)

The F1-score is used to evaluate the categorization of LC cases into two groups for mortality prediction. It evaluates test accuracy's precision and recall. Figure 6 and table 4 shows the f1-score's results and, suggested method's ENBO-ICNN is higher than other existing methods.

Table 4. Numerical Outcomes of F1-score (Source: Author)

Methods	F1-Score (%)
ANN [22]	69,44
RNN [22]	68,4
CNN [22]	69,02
RFC [22]	70,71
ENBO-ICNN (Proposed)	92

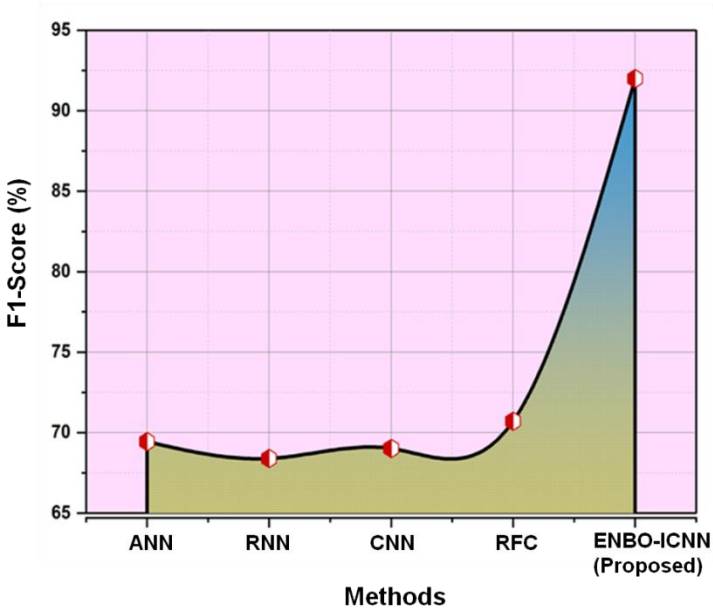

Figure 6. Graphical Outcomes of F1-score (Source: Author)

4. Discussion

When working with large sequences or time-series data, RNNs can have trouble collecting long-range dependencies, hindering their ability to understand complicated correlations in LC data.[22] Over fitting with high-dimensional data and computing expenses for large datasets limit the usefulness of ANN in processing immense amounts of LC data.[22] High dimensionality might hinder the capacity of Random Forest models to capture complicated relationships among characteristics in LC datasets.[22] They need help in capturing temporal dependencies or connections between subsequent data sets, which is essential for comprehending illness progression and treatment effects.[22] The benefits of employing ENBO-ICNN in LC analysis include enhanced accuracy in predicting mortality rates, improved early detection of potential identification and a more comprehensive understanding of complex imaging data, facilitating more effective treatment strategies and

better patient outcomes.

5. Conclusion

Our findings highlight the potential use of a DL approach in predicting LC death rates, which has substantial implications for clinical judgments. Taking into account data quality, ethical concerns and the challenges of treating patients from diverse backgrounds, we propose using ENBO-ICNN, there is strong evidence that this strategy can enhance patient care and outcome because of its high levels of recall, precision, accuracy and F1-score. In addition, our study highlights the need to combine cutting-edge healthcare technologies and a multidisciplinary approach to manage the obstacles of treating severe medical conditions. We introduce an ENBO-ICNN for predicting the death rate of LC. It outperforms traditional methods like RFO, CNN, RNN and ANN.Integrating genomes, transcriptomics, proteomics and metabolomics can illuminate LC molecular pathways. This technique could assist in constructing more accurate prediction models by considering the complex interplay of biological components. PET, MRI and radiomics studies can reveal tumor features and behavior. Imaging data can improve prognosis and prediction.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2021 May;71(3):209-49.https://doi.org/10.3322/caacc.216025.
2. Chinnappan R, Mir TA, Alsalameh S, Makhzoum T, Alzhrani A, Alnajjar K, et al. Emerging Biosensing Methods to Monitor LC Biomarkers in Biological Samples: A Comprehensive Review. Cancers. 2023 Jun 29;15(13):3414.https://doi.org/10.3390/cancers15133414.
3. Srivastava S, Jayaswal N, Kumar S, Sharma PK, Behl T, Khalid A, et al. Unveiling the potential of proteomic and genetic signatures for precision therapeutics in lung cancer management. Cellular Signalling. 2023 Oct 21:110932.https://doi.org/10.1016/j.cellsig.2023.110932.
4. Liefaard MC, Lips EH, Wesseling J, Hylton NM, Lou B, Mansi T, et al. The way of the future: personalizing treatment plans through technology. American Society of Clinical Oncology Educational Book. 2021 Apr 1;41:12-23. https://doi.org/10.1200/EDBK_320593.
5. Iqbal MJ, Javed Z, Sadia H, Qureshi IA, Irshad A, Ahmed R, et al. Clinical applications of artificial intelligence and ML in cancer diagnosis: looking into the future. Cancer Cell International. 2021 Dec;21(1):1-1.https://doi.org/10.1186/s12935-021-019811.
6. Lakshmanaprabu SK, Mohanty SN, Shankar K, Arunkumar N, Ramirez G. Optimal deep learning model for classification of lung cancer on CT images. Future Generation Computer Systems. 2019 Mar 1;92:374-82. https://doi.org/10.1016/j.future.2018.10.009.
7. Patra R. Prediction of lung cancer using machine learning classifier. InComputing Science, Communication and Security: First International Conference, COMS2 2020, Gujarat, India, March 26–27, 2020, Revised Selected Papers 1 2020 (pp. 132-142). Springer Singapore. https://doi.org/10.1007/978-981-15-6648-6_11.
8. Takahashi S, Asada K, Takasawa K, Shimoyama R, Sakai A, Bolatkan A, et al. Predicting

deep learning based multi-omics parallel integration survival subtypes in lung cancer using reverse phase protein array data. Biomolecules. 2020 Oct 19;10(10):1460. https://doi.org/10.3390/biom10101460.

9. Pradhan K, Chawla P. Medical Internet of things using machine learning algorithms for lung cancer detection. Journal of Management Analytics. 2020 Oct 1;7(4):591-623. https://doi.org/10.1080/ 23270012. 2020. 1811789.
10. Asuntha A, Srinivasan A. Deep learning for lung Cancer detection and classification. Multimedia Tools and Applications. 2020 Mar;79:7731-62.. https://doi.org/10.1007/s11042-019-083943
11. Pradhan K, Chawla P, Rawat S. A deep learning-based approach for detection of lung cancer using self adaptive sea lion optimization algorithm (SA-SLnO). Journal of Ambient Intelligence and Humanized Computing. 2023 Sep;14(9):12933-47.https://doi.org/10.1007/s12652-022-04118y
12. Xie Y, Meng WY, Li RZ, Wang YW, Qian X, Chan C, et al. Early lung cancer diagnostic biomarker discovery by machine learning methods. Translational oncology. 2021 Jan 1;14(1):100907.https://doi.org/10.1016/j.trabnon.2020.100907.
13. Yu H, Zhou Z, Wang Q. Deep learning assisted predict of lung cancer on computed tomography images using the adaptive hierarchical heuristic mathematical model. IEEE Access. 2020 May 5;8:86400-10. https://doi.org/10.1109/ACCESS.2020.25992645.
14. Cui L, Li H, Hui W, Chen S, Yang L, Kang Y, et al. A deep learning-based framework for lung cancer survival analysis with biomarker interpretation. BMC bioinformatics. 2020 Dec;21:1-4. https://doi.org/10.1186/s12859-020-3431z
15. Li Y, Ge D, Gu J, Xu F, Zhu Q, Lu C. A large cohort study identifying a novel prognosis prediction model for lung adenocarcinoma through machine learning strategies. BMC cancer. 2019 Dec;19:1-4.https://doi.org/10.1186/s12885-019-61017.
16. Lai YH, Chen WN, Hsu TC, Lin C, Tsao Y, Wu S. Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning. Scientific reports. 2020 Mar 13;10(1):4679.https://doi.org/10.1038/s41598-020-61588w
17. Wu Z, Wang L, Li C, Cai Y, Liang Y, Mo X, et al. DeepLRHE: a deep convolutional neural network framework to evaluate the risk of lung cancer recurrence and metastasis from histopathology images. Frontiers in Genetics. 2020 Aug 25;11:768.https://doi.org/10.3389/flgene.2020.00768.
18. Sim JA, Kim YA, Kim JH, Lee JM, Kim MS, Shim YM, et al. The major effects of health-related quality of life on 5-year survival prediction among lung cancer survivors: applications of machine learning. Scientific reports. 2020 Jul 1;10(1):10693.https://doi.org/10.1038/s41598-020-676043.
19. Nasrullah N, Sang J, Alam MS, Mateen M, Cai B, Hu H. Automated lung nodule detection and classification using deep learning combined with multiple strategies. Sensors. 2019 Aug 28;19(17):3722. https://doi.org/10.3390/s19173722.
20. Obulesu O, Kallam S, Dhiman G, Patan R, Kadiyala R, Raparthi Y, et al. Adaptive diagnosis of lung cancer by deep learning classification using wilcoxon gain and generator. Journal of Healthcare Engineering. 2021 Oct 13;2021.https://doi.org/10.11558/2021/s59120581.
21. Nath A, Sathishkumar K, Das P, Sudarshan KL, Mathur P. A clinicoepidemiological profile of lung cancers in India–Results from the National Cancer Registry Programme. The Indian journal of medical research. 2022 Feb;155(2):264. https://doi.org/10.4103%2Fijmr.ijmr_1364_21.
22. Doppalapudi S, Qiu RG, Badr Y. Lung cancer survival period prediction and understanding: Deep learning approaches. International Journal of Medical Informatics. 2021 Apr 1;148:104371.https://doi.org/10.1016/ j. ijmedinf.2020.104371.