PREDICTION OF LUNG CANCER AND COVID-19

Deepali Agarwal, Mukul Kumar, Mohd Umair Khan,
Nagendra, Mohammed Wasil

Computer Science and Engineering Department, MIT, Moradabad, India

formyscholars4u@gmail.com[1]
mk8271148@gmail.com[2]
uk8065@gmail.com[3]
palnishu32@gmail.com[4]
mohdwasil063@gmail.com[5]

ABSTRACT

In recent years, there has been a growing trend in the use of deep learning (DL) and Convolutional Neural Network (CNN) algorithms for COVID-19 and lung cancer prediction. These prediction models' main goal is to improve prognosis and early identification, which might lead to better patient outcomes and healthcare management techniques. By analysing complicated medical imaging data linked to COVID-19 and lung cancer, DL and CNN approaches use sophisticated computer algorithms to offer important insights on the course, severity, and response to therapy of these diseases.

KEYWORD: *Deep Learning, CNN, Machine Learning, Neural Network*

1. INTRODUCTION

One of the most lethal malignancies to human health and life is lung cancer. Based on information provided by the World Health Organization's international agency for research on cancer (IARC), lung cancer has become the malignant tumour with the second incidence rate and the first mortality among all malignant tumours [1]. The use of predictive modelling is growing in the fight against two of the deadliest illnesses of our time: COVID-19 and lung cancer. One of the main causes of cancer-related death worldwide, lung cancer is costly to treat and presents substantial obstacles to early identification for healthcare systems. On the other hand, due to its extensive spread, severe disease, and high death rates, the COVID-19 pandemic has caused hitherto unseen disruptions to public health and put a heavy strain on the global healthcare system.

In order to tackle these problems, researchers and healthcare providers have been developing predictive models that can enhance early diagnosis, prognosis, and treatment planning for COVID-19 and lung cancer using state of-the-art computational techniques, particularly deep learning (DL) and convolutional neural networks (CNNs). For the early diagnosis of lung cancer, machine-learning techniques present a viable approach [2]. Machine learning algorithms have shown promise in the analysis of complex datasets and the discovery of patterns suggestive of the emergence or progression of sickness. The patient's life span will be extended whether he or she receives an early prognosis [3][4]. These algorithms encompass a wide range of statistical and computational techniques. CNNs are a subclass of deep learning algorithms that have revolutionized medical imaging analysis by automatically identifying features in images and simplifying the process of identifying small abnormalities and illness flags. The world population standardization rate is 27.87 among a hundred thousand people [5, 6].

According to the data statistics, if the disease can be found in the early stage and the operation can be carried out in time, Patients with stage 1A lung cancer had a 5-year survival rate of more than 80%, compared to just 5% for patients with intermediate and advanced stages. [7]. According to the WHO report, lung cancer was the deadliest cancer, and almost 1.80 million people died from lung cancer in 2020, affecting 2.21 million [8].

Palani and Venkata Lakshmi [9] have given predictive modeling of lung cancer illness by continuous monitoring. This review paper's objective is to present a thorough examination of the approaches, model structures, and applications used in COVID-19 and lung cancer prediction modeling. This work aims to clarify the problems and efficacy of using CNN and ML algorithms for illness prediction by combining current research findings with new developments in the area.

Machine learning is a subfield of Artificial Intelligence [10]. Machine Learning is also used for complex data classification and decision making [11][12]. The contribution of CNN and DL algorithms to bettering clinical outcomes and prediction accuracy is at the heart of our investigation. Predictive models are able to produce insights into the course, severity, and responsiveness to treatment of diseases by utilizing a variety of data sources, such as clinical records, medical imaging, and demographic data. These insights are obtained by the use of computer algorithms. Consequently, these perceptions can guide clinical judgment, maximize resource distribution, and eventually enhance patient outcomes in the settings of COVID-19 and lung cancer.

2. LITERATURE REVIEW

In recent years, predictive modeling has become an effective tool in healthcare, offering the ability to predict disease outcomes and guide clinical decisions. With the emergence of the COVID-19 pandemic and the ongoing challenges of lung cancer, there is an urgent need to develop accurate prognostic models for these diseases. The purpose of this literature review is to examine the existing research on predictive modeling for lung cancer and COVID-19 and to highlight key methods, challenges and potential future research opportunities.

A. Prognostic modelling of lung cancer

Lung cancer remains one of the leading causes of cancer death worldwide, emphasizing the importance of early detection and intervention. Several studies have investigated predictive modeling techniques to identify individuals at high risk of developing lung cancer. For example, Aberle et al. (2011) demonstrated the effectiveness of a predictive model based on demographic factors, smoking history, and lung nodule characteristics in identifying lung cancer risk. Similarly, Bach et al. (2012) developed a risk prediction model that incorporates smoking intensity, duration, and age to estimate the likelihood of a lung cancer diagnosis.

B. Predictive modelling of COVID-19

The emergence of the COVID-19 pandemic has rapidly increased research aimed at predicting disease progression, severity, and outcomes. Predictive modelling has played a critical role in predicting epidemiological trends, guiding public health interventions, and prioritizing health care resources. Early research focused on epidemiological models to predict the spread of the virus and evaluate health services (Ferguson et al., 2020; Wu et al., 2020).

As the pandemic progressed, efforts shifted to developing models that predict clinical. results in patients with COVID-19. Machine learning methods have been widely used to predict disease severity, mortality risk, and the need for intensive care. For example, Liang et al. (2020) developed a machine learning model that uses clinical and laboratory data to predict the progression of COVID-19 patients to severe disease.

3. PROPOSED METHODOLOGY

A. Module for Preprocessing Data

Function 1: {preprocess_lung_cancer_data()}- Summary: Preprocesses, normalizes, and extracts features from imaging data related to lung cancer.

Function 2: {preprocess_covid_data()} - Description: Handles artifacts and inconsistencies in COVID-19 imaging data before preparing it for analysis.

B. Conventional Machine Learning Module

Function 1: {train_ml_model()} - Description: Uses pre-processed data to train conventional ML models (such as logistic regression and decision trees).

Function 2: {evaluate_ml_model()} - Summaries: Assesses machine learning models' performance by utilizing relevant measures, such as accuracy, precision, and recall.

C. Module on Convolutional Neural Nets (CNNs)

Function 1: {build_lung_cancer_cnn_model()} - Description: Builds CNN architectures with 2D or 3D CNN layers for the prediction of lung cancer.

Function 2: {Build_covid_cnn_model()} - Description: Creates CNN models with adjustable layers and parameters specifically for COVID-19 detection.

Function 3: {train_cnn_model()} - Description: Utilizing pre-processed imaging data for COVID-19 and lung cancer, CNN models are trained.

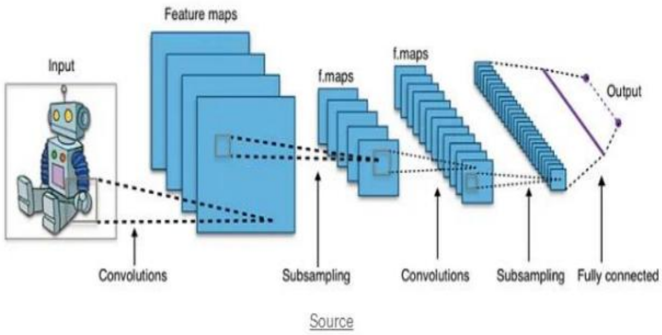

Figure 1: Convolutional Neural Network (CNN) Architecture

D. Model Integration Module

Function 1: {integrate_predictions()} - Description: Uses ensemble approaches or fusion techniques to combine predictions from CNN and ML models.

Function 2: {Evaluate_integrated_model()} - Description: Uses holdout validation or cross-validation to assess how well the integrated model performs.

E. Module for Deployment

Function 1: {deploy_model()} - Description: This function uses CNN and machine learning models that have been developed to predict COVID-19 and lung cancer in real time.

Function 2: {monitor_model_performance()} - Description: Tracks deployed models' performance and offers suggestions for improving them.

4. CASE STUDIES AND THEIR USES

In order to predict COVID-19 and lung cancer, in comparison with related work, we obtain a good result in this work with dataset and methods that we used. However, researcher in [13] obtained 94% CT scan images dataset and SURF (Speeded Up Robust Features) for feature selection.in addition the researchers in [14],[15] they used the same dataset, the researchers in [14] obtained 90.9% with used Delta Radiomics method for feature extraction. But researchers in [15] could obtained better accuracy by using GLCM function for feature extraction and (MLP 98%, SVM 70.45%, & KNN 99.2%) classifier convolutional neural networks (CNNs) and machine learning (ML) algorithms have been used realistically. This has resulted in a number of case studies and practical applications in various clinical settings. These case studies highlight the promise of predictive modelling while also illuminating the difficulties and key lessons learned from using these models to illness diagnosis and treatment.

A noteworthy example study uses clinical, radiological, and demographic data along with machine learning algorithms to evaluate the risk of lung cancer. By means of a retrospective examination of patient data, scientists created a prediction model that included factors including age, smoking history, tumour size, and histological characteristics to determine the likelihood of lung cancer progression. Health care professionals may now conduct focused screening and monitoring programs to identify and treat lung cancer early on due to the model's high accuracy in identifying lung cancer risk variables.

Similar to this, using DL and CNN algorithms, prognosis and disease severity have been predicted for COVID 19 based on imaging and clinical data. For example, scientists created a CNN model that can analyse chest X-rays or CT scans to find abnormalities associated with COVID-19 and gauge the illness's course. The method helped physicians prioritize patients and use resources more wisely by providing important insights into the severity and prognosis of the ailment by establishing a connection between imaging data and clinical outcomes.

Applying predictive models in clinical contexts has presented several difficulties, notwithstanding the efficacy of these case studies. Data availability and quality continue to be major obstacles, especially in the COVID-19 setting where imaging datasets may be erratic or limited. Moreover, healthcare professionals need to understand the underlying causes influencing model predictions in order to make educated clinical decisions; for this reason, the model's interpretability and transparency are crucial elements to take into account. In addition, it is imperative to

prioritize ethical and legal aspects such as algorithmic bias and patient privacy to enable the proper implementation and utilization of predictive models in clinical settings.

5. FLOWCHART FOR PROPOSED CNN ALGORITHM

The proposed CNN algorithm involves several key steps, including data loading, model initialization, forward and backward propagation, weight updates, and convergence criteria. The flowchart illustrates the iterative training process employed by CNN algorithms to learn features and predict disease outcomes.

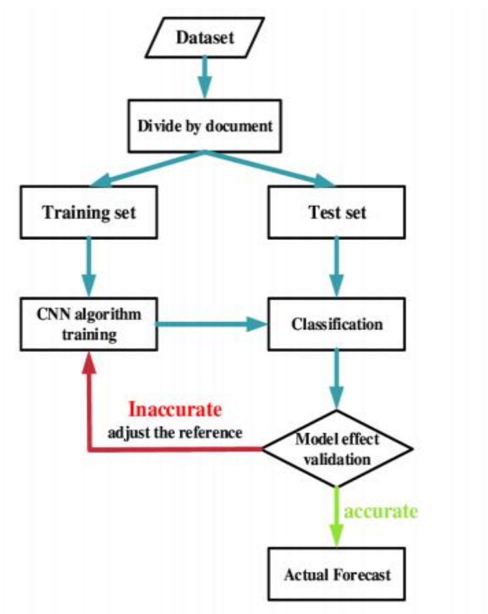

Figure 2: Flow Chart of CNN Algorithm

6. RESULT

SNAPSHOTS

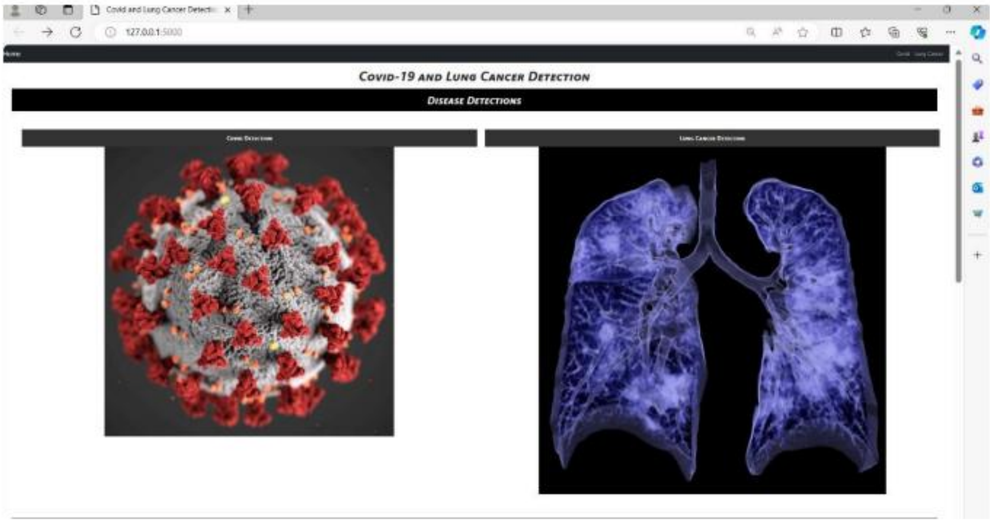

Figure 3: HOME PAGE

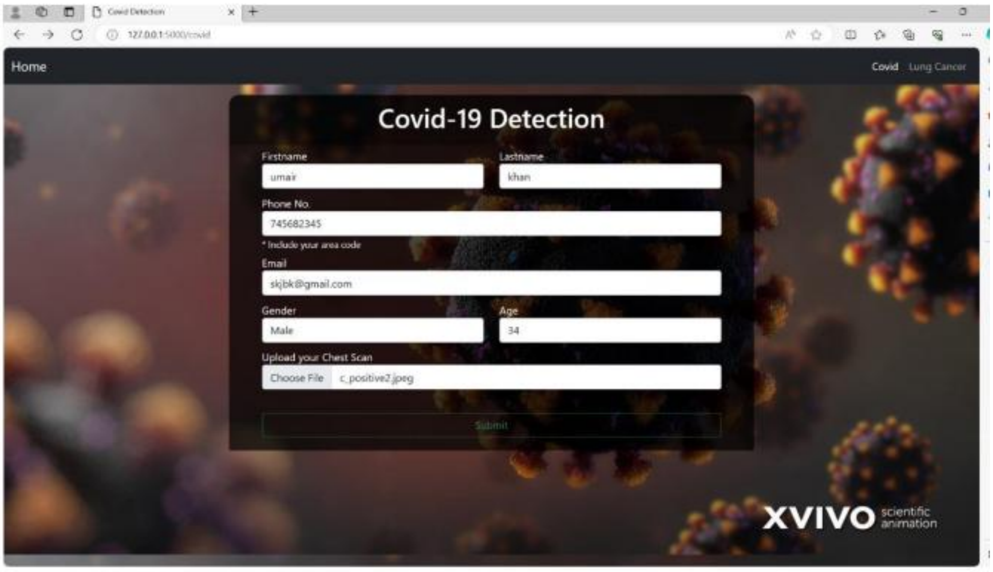

Figure 4: COVID-19 PATIENT DETAILS

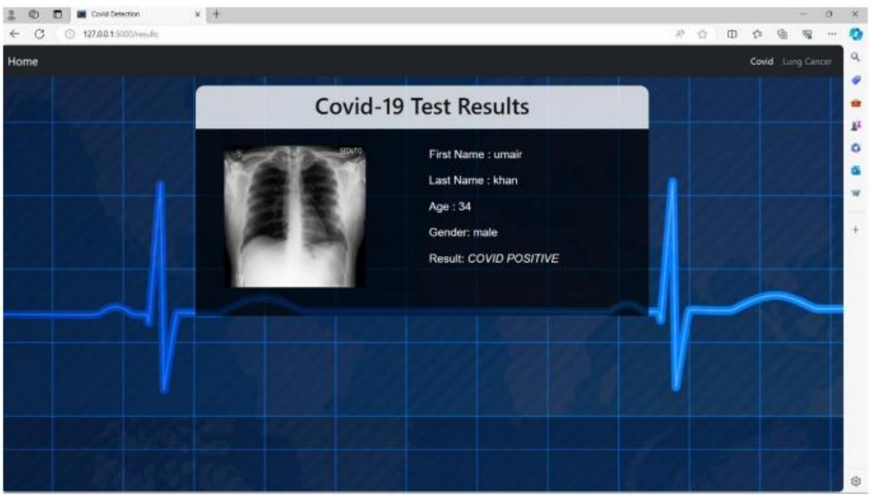

Figure 5: COVID-19 TEST RESULTS

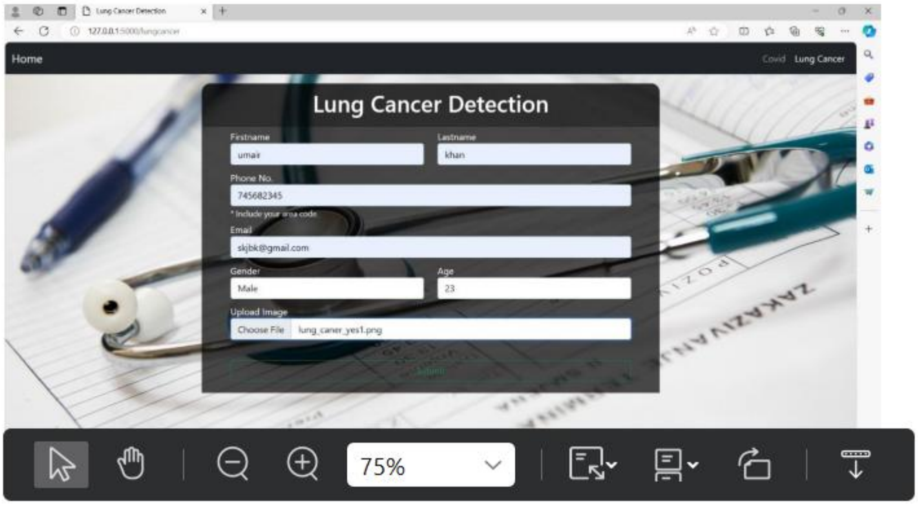

Figure 6: LUNG CANCER PATIENT DETAILS

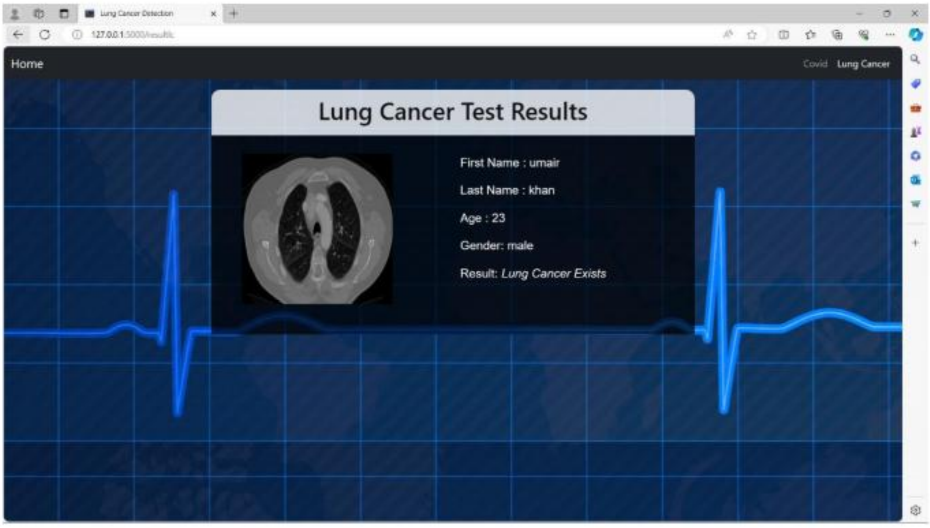

Figure 7: LUNG CANCER TEST RESULTS

7. FUTURE SCOPE

There are constantly new advances and ongoing challenges with COVID-19 and lung cancer prediction models that politicians, medical practitioners, and academics must address. This section examines the obstacles and potential next steps in predictive modelling for certain illnesses, with the goal of guiding the creation of more effective and therapeutically relevant models.

The integration of multimodal data sources is a novel concept in COVID-19 and lung cancer predictive modelling that aims to increase therapeutic usefulness and accuracy. Although medical imaging data has typically been the primary focus of predictive modelling, other data modalities such as genetic, proteomic, and clinical indicators are increasingly acknowledged for their potential benefits. Integrating these diverse data sources can lead to a better understanding of illness causes and make it simpler to build personalized prediction models based on specific patient characteristics.

Furthermore, developments in computational techniques, particularly deep learning and reinforcement learning, hold promise for increasing the predictive capacity of DL and CNN systems. Attention mechanisms, graph neural networks, and generative adversarial networks are new ways to collecting complex relationships in medical imaging data and improving model interpretability. Reinforcement learning approaches also enable adaptive decision-making based on clinical outcome data, paving the way for dynamic and responsive prediction models.

Despite these achievements, major challenges remain in the field of predictive modelling for lung cancer and COVID-19. Data quality and availability remain significant challenges, particularly in the case of COVID-19, where imaging datasets may be few or heterogeneous. To solve these issues, efforts must be made to standardize data collection procedures, increase data sharing activities, and develop rigorous data preparation techniques capable of reducing biases and inconsistencies.

Model interpretability and openness are also critical considerations in the development and application of prediction models for lung cancer and COVID-19. To make informed clinical decisions and create trust in prediction models, physicians must grasp the underlying rationale behind model forecasts. Explaining model predictions via methodologies such as attention processes, saliency maps, and feature attribution methods can increase model interpretability while also promoting clinical acceptability and adoption.

Furthermore, scalability and generalizability are important considerations when implementing predictive models across several healthcare settings and patient populations. Models developed in one therapeutic setting may not be transferable to another, necessitating validation and adaptation in many situations. Furthermore, predictive models' ability to scale to massive amounts of data and make real-time choices is crucial for widespread adoption and therapeutic effectiveness.

8. CONCLUSION

To summary, this review research conducted a detailed investigation of the application of deep learning (DL) and Convolutional Neural Network (CNN) algorithms for predicting lung cancer and COVID-19. An assessment of methodology, model architectures, case studies, future directions, and challenges revealed valuable insights into predictive modelling efficacy and potential in combating these two lethal disorders. The importance of deep
learning (DL) and CNN algorithms for predicting lung cancer and COVID-19 cannot be overstated. These new computational techniques provide unprecedented opportunities to improve early diagnosis, prognosis, and treatment planning, ultimately improving patient outcomes and lowering costs for healthcare systems worldwide. Predictive models can provide helpful insights into sickness onset, severity, and treatment response by combining various data sources such as medical imaging, clinical records, and genetic information, allowing clinicians to make more informed decisions and enhance patient care.

Future research and clinical practice in the domains of lung cancer and COVID-19 prediction modelling may yield a variety of suggestions. First, efforts should be made to improve data quality and availability, particularly in the case of COVID-19, where imaging datasets may be limited or varied. Standardizing data collection techniques, strengthening data exchange programs, and developing robust data preparation technologies are all necessary steps toward overcoming these challenges. Improving model interpretability and transparency should be focused to increase clinical acceptance and uptake. Clinicians want insights into the underlying factors that drive model predictions in order to make informed decisions and establish trust in predictive models. Model predictions may be better explained utilizing methodologies like attention processes, saliency maps, and feature attribution methods.

REFRENCES

[1] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 2021, 71(3): 209-249. [2] M. Guo et al., “Autologous tumor cell–derived microparticle-based targeted chemotherapy in lung cancer patients with malignant pleural effusion,” Sci. Transl. Med., vol. 11, no. 474, Jan. 2019, doi: 10.1126/scitranslmed.aat5690.

[3] S. F. Khorshid and A. M. Abdulazeez, “BREAST CANCER DIAGNOSIS BASED ON K-NEAREST NEIGHBORS: A REVIEW,” PalArch’s J. Archaeol. Egypt/Egyptology, vol. 18, no. 4, pp. 1927–1951, 2021.

[4] P. Chaudhari, H. Agarwal, and V. Bhateja, “Data augmentation for cancer classification in oncogenomics: an improved KNN based approach,” Evol. Intell., pp. 1–10, 2019.

[5] Zheng R, Zhang S, Zeng H, et al. Cancer incidence and mortality in China. J Natl Cancer Cent, 2022, 2 (1): 1-9.

[6] He J, Wei WQ. 2019 China cancer registry annual report. Beijing: People’s Health Publishing House, 2021: 59-98.

[7] Siegel RL, Miller KD, Jemal A. Cancer statistics. CA Cancer J Clin, 2017, 67(1): 7-30. [8] "Cancer," World Health Organization. [Online]. Available: https://www.who.int/newsroom/fact sheets/detail/cancer. [Accessed: 15- Apr-2022].

[9] D. Palani and K. Venkatalakshmi, "An IoT based predictive modelling for predicting lung cancer using fuzzy cluster based segmentation and classification," Journal of Medical Systems, vol. 43, no. 2, p. 21, 2019.

[10]M. R. Mahmood, A. M. Abdulazeez, and Z. ORMAN, "A NEW HAND GESTURE RECOGNITION SYSTEM USING ARTIFICIAL NEURAL NETWORK."

[11]M. Somvanshi, P. Chavan, S. Tambade, and S. V. Shinde, "A review of machine learning techniques using decision tree and support vector machine," Proc. - 2nd Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2016, 2017, doi: 10.1109/ICCUBEA.2016.7860040.

[12]D. M. Abdulqader, A. M. Abdulazeez, and D. Q. Zeebaree, "Machine Learning Supervised Algorithms of Gene Selection: A Review," Mach. Learn., vol. 62, no. 03, 2020.

[13]D. Q. Zeebaree, H. Haron, and A. M. Abdulazeez, "Gene selection and classification of microarray data using convolutional neural network," in 2018 International Conference on Advanced Science and Engineering (ICOASE), 2018, pp. 145– 150.

[14]D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and D. A. Zebari, "Trainable model based on new uniform LBP feature to identify the risk of the breast cancer," in 2019 International Conference on Advanced Science and Engineering (ICOASE), 2019, pp. 106–111.

[15]H. Tang, J. Zhao, and X. Yang, "Explore machine learning for analysis and prediction of lung cancer related risk factors," in Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence, 2018, pp. 41–45.