Journal of Information Systems Engineering and Management
2025, 10(28s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Multilevel Optimization with Hybrid Stack Model for Lung Cancer Classification

Aanchal Vij, Kuldeep Singh Kaswan, Anand Nayyar
School of Computing Science and Engineering
Galgotias University, Greater Noida, Uttar Pradesh, India
aanchal.vij04@gmail.com
School of Computing Science and Engineering
Galgotias University, Greater Noida, Uttar Pradesh, India
kaswankuldeep@gmail.com
Graduate School
Duy Tan University, Dà Nang 550000, Vietnam
anandnayyar@duytan.edu.vn

ARTICLE INFO

Received: 20 Dec 2024
Revised: 22 Feb 2025
Accepted: 26 Feb 2025

ABSTRACT

Lung cancer is the leading cause of cancer-related mortality worldwide, and detecting the disease can still save lives. if a lung cancer diagnosis has been made. The illness known as lung cancer occurs when healthy lung cells transform into dangerous aberrant cells known as cancer cells. Tumors are collections of cancerous cells that grow over time. Today, medical imaging scans are being interpreted with the use of artificial intelligence and contemporary data science techniques. In contrast to conventional techniques, which depend on the subjective and time-consuming visual examination of radiologists, the emphasis now is on creating reliable automated diagnostic tools. This change is in line with the core objective of radiomics, a developing field of study that combines customized medicine with medical imaging. Improvements in lung cancer screening provide a glimmer of hope for life-saving treatments. Together, in stage one, the survival rate is 70%. Stage 2 drops to 50% as stage 3 drops. And stage four for the most part is not curable disease, but there are some patients who might be alive around five years.The aforementioned problems are efficiently addressed by the merging of LDCT followed by AI based Multilevel Optimization with hybrid stack model. This paper introduces a hierarchical reference architecture for Lung Cancer Classification. The proposed approach (GLCM) features are extracted from LDCT further investigated with M-GWO technique to figure-out best features. The best solution obtained from this hybrid stacked model is use to classify input image as normal or abnormal.

Keywords: LDCT , Lung Cancer, Machine Learning , Optimization , GLCM.

1. INTRODUCTION

1. Lung cancer (LC) is the world's biggest cause of cancer-related mortality, but improvements in lung cancer screening provide a glimmer of hope for life-saving treatments. The lungs are two critical organs located in the chest that are fundamental to the complicated structure of our bodies [1]. They allow for the exchange of carbon dioxide and oxygen, which is necessary for life. By breathing in oxygen and breathing out carbon dioxide through tiny air sacs, our lungs regulate this vital gas exchange. The onset of lung cancer, a deadly disease in which

healthy lung cells transform into cancerous cells that multiply into tumors and cause havoc by encroaching on and damaging healthy lung tissue, can, however, upset this harmonious function [2].

Lung cancer's journey is not limited to the lung; there is a significant possibility that it may spread via blood or lymph to other regions of the body. Two different varieties of lung cancer are small cell lung cancer (SCLC) and non-small cell lung cancer (NCLC). These are identified by the cell types that are affected and how they appear under the microscope. The most common kind of lung cancer is non-small cell, but SCLC is more aggressive and grows and spreads quickly [3]. The root cause of LC are smoking, family history of lung cancer, HIV infection, exposure to hazardous chemicals, and second hand smoking, radon, asbestos, and air pollution [4]. These all contribute to an elevated risk profile. Lung cancer is a sneaky disease; in its early stages, it seldom exhibits any signs. However, if the illness worsens, symptoms include a persistent cough, chest discomfort, breathing problems, blood in the cough, hoarseness, appetite loss, trouble swallowing, weight loss, exhaustion, and swelling in the face or neck may appear, indicating the need for immediate medical assistance [5].

The spectrum of symptoms associated with lung cancer encompasses fatigue, sudden weight loss, dyspnea, chest pain, persistent cough, and hemoptysis, indicative of coughing up blood. Various root cause of LC, including exposure to cigarette smoke, nickel, arsenic, air pollution, radon, and a personal or family history of the disease [6]. Additionally, asbestos exposure is a notable risk factor. Small cell carcinoma often presents at an advanced metastatic stage and involves neuroendocrine cells. Non-small cell carcinoma includes adenocarcinoma, which is common and affects gland cells, and squamous cell carcinoma, which occurs in proximal lung tissue [7]. These Large cell carcinoma (LCC), characterized as large tumor cells, can grow in either proximal or peripheral lung tissue. Radiographic imaging, such as chest X-rays, aids in identifying lung cancer features like pulmonary opacity, hilum enlargement, pleural effusion, and lung collapse, depending on tumor size. lung cancers can prompt the transformation of lung cells into neuroendocrine cells, leading to paraneoplastic syndrome [9]. Neuroendocrine cancer cells release hormone-like substances, such as parathyroid hormone-like substance causing hypercalcemia, ACTH-like substance stimulating cortisol production, and anti-diuretic hormone increasing water retention, mirroring normal hormone functions.

Lung cancer screening emphasizes on two importance method for identifying patients who require an immediate referral to a lung specialist. Specifically, this includes individuals who have had a chest X-ray indicating potential lung cancer and those aged 40 and above experiencing unexplained hemoptysis (coughing up blood) [10]. For instance, a 50-year-old woman presenting with a week-long history of hemoptysis should prompt an urgent two-week referral without delay. The second part of the guideline pertains to patients who necessitate an urgent chest X-ray within two weeks to evaluate for lung cancer. Notably, The helpful list of symptoms to monitor, particularly in individuals aged 40 and older. Moreover, if a patient exhibits two unexplained symptoms or has a history of smoking alongside one or more unexplained symptoms, an urgent chest X-ray is warranted. The guideline outlines specific clinical indicators that warrant consideration for a chest X-ray within two weeks. These include thrombocytosis (increased platelet count) in patients 14 years of age and older, finger clubbing, supraclavicular lymphadenopathy, the continuation of fical lymphadenopathy, and chest symptoms suggestive of lung cancer [11]. For instance, a patient with a persistent chest infection unresponsive to multiple antibiotic courses should raise suspicion for underlying LC. Finding the stage and diagnosis of lung cancer are essential elements in formulating a successful treatment plan. The progression of NCLC involves several phases, ranging from the occult stage, in which cancer cells are limited to lung fluids, to advanced stages, where the tumors invade distant organs and lymph nodes [12]. On the other hand, small cell lung cancer is categorized into localized and regional phases according to the degree of dissemination both

inside the chest and to other locations [13]. It typically occurs between the ages of 55 and 84 years, with the peak incidence seen in the 7th and 8th decades of life, around 65 to 74 years. One crucial aspect to understand is the strong association between lung cancer and cigarette smoking. When someone inhales tobacco, particularly the polycyclic hydrocarbons and nitrosamines present in it, they are exposed to potent carcinogens. Polycyclic hydrocarbons, in particular, are known to bind to nuclear DNA, causing mutations and acting as mutagens. Benzopyrene compounds, another component of tobacco, then act as tumor promoters, further contributing to the development of lung cancer.Despite advances in medical technology, lung cancer continues to be a severe challenge in the field of oncology, taking a considerable number of lives each year. The goal of early lung cancer detection provides pathway to introduce novel techniques, most notably the use of artificial intelligence (AI) into diagnostic medical imaging. This study explores the vital significance of early lung cancer diagnosis, emphasizing how AI-driven solutions have the potential to transform screening practices and enhance patient outcomes [14]. The startling data on lung cancer survival and diagnosis rates highlight the critical need for game-changing therapies. Every year, around 2 million individuals worldwide receive a lung cancer diagnosis, and a sizable portion of them lose their lives to the disease. With just a small percentage of patients living past diagnosis, the five-year survival statistics, especially in developed countries such as the UK, provide a sobering picture of the obstacles that patients confront. Lung cancer symptoms in their late stages are a major factor in this dismal situation since they frequently appear after the illness has moved to an advanced level, which reduces the effectiveness of therapy and the likelihood of survival [15].

An ray of hope has emerged in the fight against lung cancer with the use of AI technology in healthcare. By leveraging AI algorithms, medical professionals can analyze chest CT-Scan with unprecedented accuracy, identifying subtle nodules that may indicate early-stage lung cancer. This capability holds immense potential in circumventing the late diagnosis dilemma, paving the way for timely interventions and improved survival rates. However, while AI-enabled chest CT-Scan analysis is a significant advancement and it offers even greater precision in detecting and characterizing pulmonary nodules [16]. One of the key advantages of AI-enabled chest CT-Scan analysis lies in its applicability to incidental nodule diagnosis. Given the widespread use of chest CT-Scan for various medical purposes globally, integrating AI algorithms into existing imaging workflows can facilitate the early identification of suspicious nodules indicative of lung cancer [17]. This proactive approach streamlines the triage process, enabling healthcare providers to prioritize high-risk patients for further diagnostic evaluations such as chest CT scans or biopsies [18]. Incorporating CT scans into the diagnostic pathway enhances the accuracy of lung cancer detection, particularly in distinguishing benign nodules from malignant ones and assessing the extent of disease progression [19]. Therefore, while AI-driven chest CT-Scan analysis is a valuable tool in early lung cancer detection, the integration of CT scans provides a comprehensive diagnostic approach that maximizes the chances of identifying and treating lung cancer at its earliest stages. This multidimensional strategy underscores the importance of leveraging advanced imaging technologies to combat this pervasive disease effectively [20].

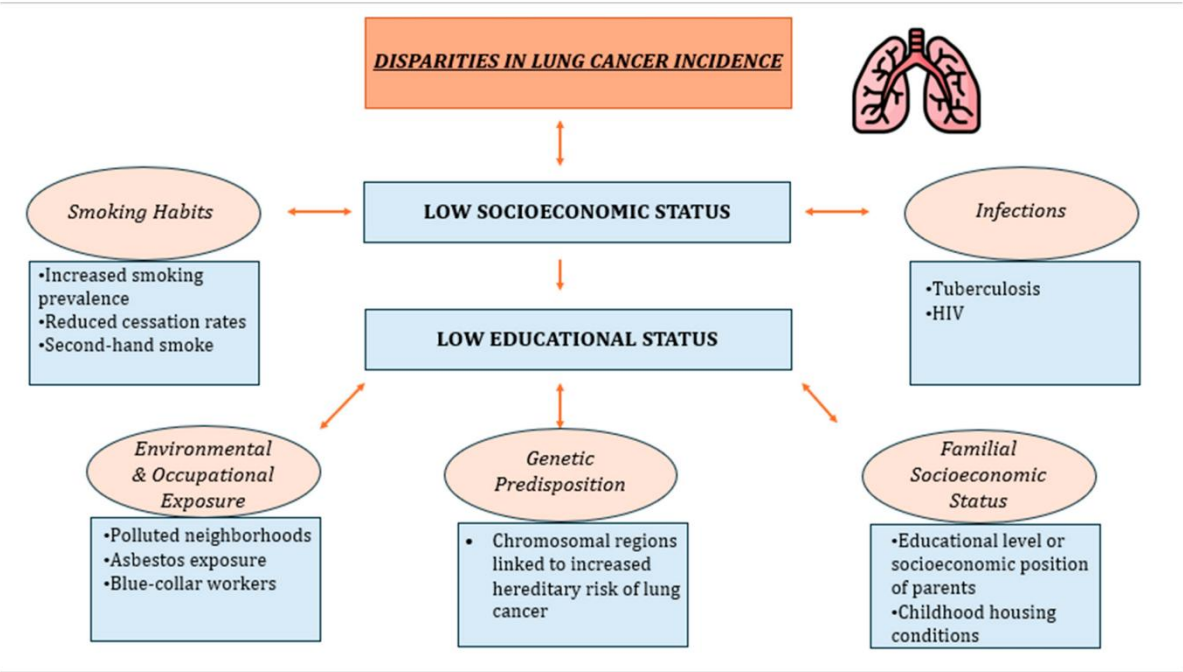


Figure 1. Root cause analysis of Lung Cancer

2. ROLE OF AI IN SCREENING OF LUNG CANCER:

The role of artificial intelligence (AI) in screening for lung cancer has become increasingly prominent due to its potential to enhance early detection and improve patient outcomes. AI techniques integrated into medical imaging systems to aid in the interpretation of lung scans. Here are some key aspects of AI's role in lung cancer screening:

Risk Stratification: AI can assist in stratifying patients based on their risk of developing lung cancer. By analyzing various factors, such as imaging features, patient history, and genetic markers, AI models can provide personalized risk assessments and guide healthcare providers in implementing targeted screening strategies [21].

Integration with Clinical Workflow: AI tools can be seamlessly integrated into existing clinical workflows, allowing for efficient screening processes and streamlined communication between healthcare professionals. This integration facilitates prompt follow-up and treatment planning for patients identified with suspicious findings [22].

Enhanced Accuracy: Studies have shown that AI-based lung cancer screening systems can achieve high levels of sensitivity and specificity, reducing false-positive and false-negative rates compared to traditional methods. This increased accuracy helps avoid unnecessary interventions while ensuring that potential cases of lung cancer are not missed [23].

Continuous Learning: AI models can continuously learn from new data and updates, improving their performance over time. This adaptability enables ongoing refinement of screening algorithms and ensures that they remain up-to-date with the latest advancements in LC detection and management [24].

This paper is structured as follows: second Section provides the motivation behind this research work, third section provide the related works, fourth section presents the proposed technique, fifth section analyzes the results, and last section concludes the paper with discussions.

3. MOTIVATION:

The motivation behind the advancements in lung cancer diagnosis stems from a deep commitment to enhancing patient outcomes, streamlining healthcare delivery, and improving medical decision-making processes. By embracing innovative technologies, we aim to tackle the multifaceted challenges associated with lung cancer, spanning from accurate diagnosis to personalized treatment strategies. Our focus is on leveraging these advancements to revolutionize healthcare delivery and ultimately improve the quality of life for patients battling lung cancer.

4. OUR CONTRIBUTIONS:

Hierarchical Reference Architecture Leveraging Lung Cancer Classification:

We have introduced a novel hierarchical reference architecture that integrates lung cancer classification methodologies. This structured framework not only facilitates accurate diagnosis but also enhances the management of complex healthcare systems efficiently. By leveraging insights from lung cancer classification leads to improved patient treatment .

M-GWO Feature Selection Technique:

Our proposed Hybrid_PSO feature selection technique plays a pivotal role in optimizing overall system performance. By employing advanced algorithms, we can identify and prioritize key features relevant to lung cancer diagnosis and treatment. This optimized feature selection process not only improves the accuracy of diagnostic tools but also contributes to more efficient healthcare delivery, reducing unnecessary procedures and enhancing resource utilization.

Machine Learning-Based Stacked Model for Scalability and Flexibility:

We have developed a machine learning-based stacked model designed to offer scalability and flexibility to healthcare institutions. This model empowers hospitals to adapt swiftly to evolving technological advancements in the medical domain. By leveraging machine learning capabilities, hospitals can streamline workflows, enhance predictive analytics for early detection, and customize treatment protocols based on real-time data insights. This scalability and flexibility enable healthcare facilities to stay at the forefront of lung cancer diagnosis and treatment, ensuring optimal patient care in a rapidly changing healthcare landscape.

5. Related work:

The medical landscape, particularly in cancer treatment, has witnessed remarkable strides through AI-driven analyses of vast datasets. These technological advancements have not only enhanced diagnostic accuracy but also paved the way for tailored therapies based on individual patient profiles. recent studies have demonstrated AI's prowess in early lung cancer detection, transcending the traditional association with smoking. Non-smokers, too, can fall victim to lung cancer, emphasizing the criticality of vigilant symptom monitoring and timely medical intervention.

In the realm of LC classification using deep learning and image analysis techniques, various research endeavors have been undertaken to address the challenges of accuracy and efficiency in diagnosis. Raza et al. (2023) introduced Lung-EffNet, a transfer learning-based predictor leveraging the EfficientNet architecture. This approach represents a significant leap in accuracy, as Lung-EffNet achieved remarkable results in accurately classifying LC from CT scans. By integrating top layers in the classification head of the model, Lung-EffNet demonstrated the potential of transfer learning in enhancing predictive capabilities, showcasing its effectiveness in the critical domain of early cancer detection. Pandit et al. (2023) contributed to the field by enhancing lung cancer classification through multispace image pooling and an autoencoder system integrated into their CNN model. This innovative

approach not only improved overall accuracy but also addressed the issue of processing time, a crucial factor in real-time diagnosis and treatment planning. By leveraging advanced techniques like multispace image pooling and autoencoders, Pandit et al. showcased the importance of combining traditional CNN frameworks with cutting-edge methodologies to achieve optimal results in medical image analysis. Naseer et al. (2023) focused on lobe segmentation and nodule detection, essential components in accurate lung cancer diagnosis. Their modified U-Net and SVM-based model demonstrated promising results in segmenting lobes and identifying nodules, contributing significantly to the automated detection and classification of LC from CT scans. This work highlights the importance of robust segmentation techniques in preprocessing medical images and lays the foundation for improved diagnostic tools. Mohamed et al. (2023) proposed a hybrid EOSA-CNN model, showcasing the potential of metaheuristic optimization in improving CNN performance for lung cancer classification. By addressing challenges related to parameter selection and model architecture, the EOSA-CNN hybrid model achieved high accuracy and demonstrated superior specificity and sensitivity compared to other methods. This research signifies the importance of optimizing deep learning models using metaheuristic algorithms for enhanced performance in medical image analysis.

AR et al. (2023) introduced LCD-CapsNet, an innovative combination of CNN and Capsule Network, aimed at improving LC detection accuracy. By leveraging the unique capabilities of Capsule Networks in recognizing fine-grained spatial correlations, LCD-CapsNet achieved high precision and recall rates, showcasing its potential as a robust tool for automated LC diagnosis from CT scans. This work highlights the importance of exploring diverse deep learning architectures to achieve optimal results in medical image classification tasks. Prakash et al. (2023) contributed to the field by developing an EESNN-FSOA-LCC model using genetic algorithm optimization. This approach demonstrated superior accuracy and processing efficiency compared to existing methods, showcasing the potential of hybrid metaheuristic and CNN algorithms in improving lung cancer classification. By integrating genetic algorithm optimization into the model training process, Prakash et al. highlighted the importance of fine-tuning model parameters for optimal performance in medical image analysis. Bushara et al. (2023) introduced VGG-CapsNet, an ensemble method for lung cancer detection that leveraged the capabilities of both VGG and Capsule Networks. This novel approach demonstrated high precision, recall, and specificity rates, underscoring its effectiveness in accurately identifying and classifying lung cancer from CT scans. By combining different deep learning architectures, Bushara et al. showcased the potential of ensemble methods in improving diagnostic accuracy in medical imaging. Dunn et al. (2023) utilized deep learning and radiomic analysis for automated lung cancer subtype classification, achieving high accuracy in distinguishing different subtypes. Their study underscored the importance of leveraging advanced AI techniques like deep learning and radiomics to extract meaningful features from medical images, enabling accurate classification of lung cancer subtypes. This research contributes to the ongoing efforts in personalized medicine by providing tools for precise diagnosis and treatment planning based on subtype-specific characteristics.

The collective efforts of researchers in the field of LC classification using deep learning and image analysis techniques signify the rapid advancements and ongoing innovations in medical image analysis. These endeavors not only focus on improving diagnostic accuracy but also emphasize the importance of efficiency, processing speed, and model optimization. By exploring diverse methodologies, combining different deep learning architectures, and integrating metaheuristic optimization techniques, researchers are paving the way for more accurate, efficient, and personalized approaches to lung cancer diagnosis and treatment.

Table 1. Comparative analysis among literature survey the done

Reference	Objective	Methodology	Advantage	Limitations
[25] Raza et al. (2023)	LC classification using EfficientNet from CT-scan images	Proposed Lung-EffNet based on EfficientNet architecture with top layer modification, evaluated on IQ-OTH/NCCD dataset, handled class imbalance	Achieved 99.10% accuracy, outperformed other CNNs in terms of accuracy and efficiency	Lack of detailed comparison with other state-of-the-art models
[26] Pandit et al. (2023)	Enhanced lung cancer classification using CNN with multispace image pooling and autoencoder system	Used CNN with multispace image pooling and autoencoder for improved accuracy and reduced processing time	Increased accuracy to 99.5%, reduced processing time	Limited discussion on scalability and generalizability
[27] Naseer et al. (2023)	LC classification using modified U-Net and SVM	Developed a modified U-Net for lobe segmentation and nodule detection, used SVM for classification	Achieved promising results in lobe segmentation and nodule detection	Limited dataset description and external validation
[28] Mohamed et al. (2023)	LC classification using hybrid metaheuristic and CNN	Proposed EOSA-CNN hybrid model for lung cancer classification, optimized with EOSA	Achieved 0.9321 accuracy, demonstrated improved specificity and sensitivity	Limited discussion on algorithm complexity and training time
[29] AR et al. (2023)	Detection and classification of LC using LCD-CapsNet	Introduced LCD-CapsNet combining CNN and CapsNet for LC detection and classification	Achieved 94% accuracy, demonstrated high precision and recall	Lack of detailed comparison with other CapsNet variants
[30] Prakash et al. (2023)	Enhanced Elman Spike Neural Network for lung cancer classification	Proposed EESNN-FSOA-LCC model using genetic algorithm and flamingo search optimization	Achieved high accuracy compared to existing methods	Limited discussion on model interpretability
[31] Cao et al. (2023)	Weakly supervised deep CNN for LC classification	Developed E2EFP-MIL model for efficient and accurate LC subtype classification	Achieved high AUCs and accuracy, demonstrated generalizability	Limited discussion on model explainability and clinical validation
[32] Jagadeesh et al. (2023)	Genetic algorithm-based LC segmentation and classification	Developed improved model using genetic algorithm for LC segmentation and PNN for classification	Outperformed existing methods in accuracy and processing time	Limited discussion on algorithm scalability
[33] Bushara et al. (2023)	Ensemble method with CapsNet for LC detection	Proposed VGG-CapsNet ensemble for LC detection, achieved high accuracy and precision	Demonstrated superiority over basic CapsNet and CNN combinations	Lack of detailed analysis on model robustness
[34] Dunn et al. (2023)	Automated LC subtype classification using deep learning	Applied iMRRN for image segmentation and various classification algorithms for subtype classification	Achieved high accuracy in classifying lung cancer subtypes	Limited discussion on data preprocessing steps and model explainability

6. **Proposed work:** As meticulous approach has been adopted, employing a feature selection method known as M-GWO in algorithm 1. This acronym may signify an optimization algorithm designed for feature selection. Through the implementation of M-GWO, a comprehensive analysis has been undertaken, resulting in the identification and selection of eleven crucial features deemed essential for optimizing the system's performance. Subsequently, the data corresponding to these selected features has been compiled into a new dataset named M-GWO. This dataset serves as the foundation for the next phase of the stack model utilizing the insights gained from the M-GWO dataset. The hybrid_stack model implies a strategic arrangement or configuration of features. This systematic and data-driven approach reflects a commitment to enhancing the performance and responsiveness of the framework in question, showcasing a methodical integration of advanced optimization techniques in the pursuit of efficiency.

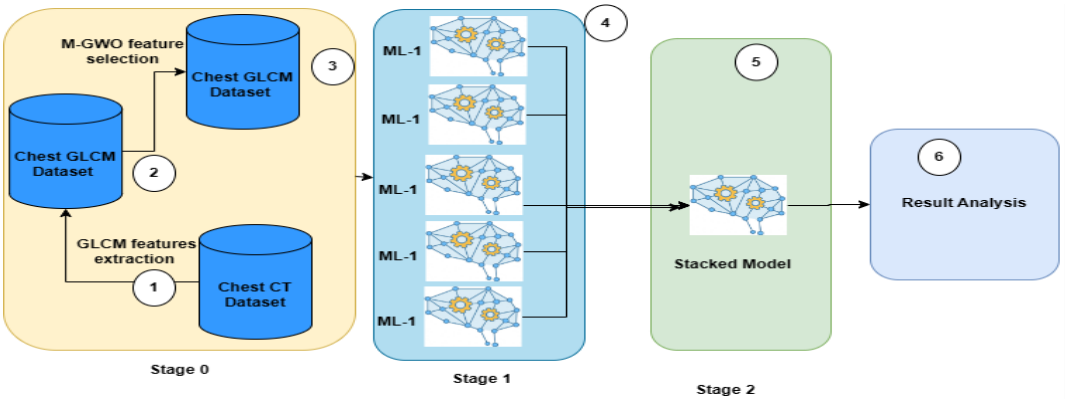


Figure 2. Work flow of proposed model.

Data Description: The LC dataset from the Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) was gathered during a three-month period in autumn 2019 at the aforementioned specialty institutions. It comprises both healthy individuals' and patients' CT scans that have been diagnosed with LC at various stages. In these two centers, radiologists and oncologists marked IQ-OTH/NCCD slides. A total of 1190 pictures, or CT scan slices from 110 instances, are included in the collection (see Figure 1). The three groups of these instances are malignant, benign, and normal. Out of them, 45 instances have been classed as normal cases, 40 as malignant cases, and 15 as benign ones. Originally, the CT scans were gathered in DICOM format. [35].

GLCM Feature Extraction:

i. *Autocorrelation Equation* (1):

- Autocorrelation measures the similarity of a pixel to its neighboring pixel throughout the image.
- It quantifies how the pixel values are correlated with each other at different spatial distances.
- A higher autocorrelation value indicates smoother and more uniform texture

$$P(x,y) = \frac{\sum_{u=0}^{N}\sum_{v=0}^{N} I(u,v)I(u+x,v+y)}{\sum_{u=0}^{N_{g-1}}\sum_{v=0}^{N_{g-1}} I^2(u,v)} \quad (1)$$

ii. *Contrast* (Equation 2):

- Contrast measures the intensity difference between a pixel and its neighbors.
- It quantifies the local variations or changes in pixel intensity within the image.
- Higher contrast values indicate a more distinct boundary between different texture regions.

$$Cont. = \sum_{i}\sum_{j} |i-j|^2 p(i,j) \quad (2)$$

iii. *Correlation* (Equation 2):

- It quantifies how much the pixel values are correlated with each other across the image.

- A high correlation value indicates a strong linear relationship between neighboring pixels.

$$\text{Correl} = \sum_{i}\sum_{j} \frac{(i - \mu_i)(j - \mu_j)p(i,j)}{\sigma_i \, \sigma_j}$$

(3)

iv. *Cluster Prominence* (Equation 4):

- It measures the skewness of the GLCM matrix, indicating the presence of clustered or dispersed texture patterns.
- Higher values suggest more asymmetry and clustering in the texture.

$$\text{Pro} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \left(i + j - u_x - u_y\right)^4 p(i,j)$$

(4)

v. *Cluster Shade* (Equation 5):

- Cluster Shade measures the skewness of the GLCM matrix and the uniformity of pixel distribution.
- It quantifies the asymmetry of the GLCM around its mean value, indicating the uniformity or non-uniformity of texture.
- Higher values indicate greater asymmetry and non-uniformity in texture distribution.

$$\text{Sha} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \left(i + j - u_x - u_y\right)^3 p(i,j)$$

(5)

vi. *Dissimilarity* (Equation 6):

- Dissimilarity measures the average absolute difference in grayscale values between a pixel and its neighbors.
- It quantifies the heterogeneity or dissimilarity of texture within the image.
- Higher values indicate greater dissimilarity or variation in texture patterns.

$$\text{Dissimilarity} = \sum_{i}\sum_{j} |i - j| \, p(i,j)$$

(6)

vii. *Energy* (Equation 7):

- Energy measures the uniformity and orderliness of texture by summing the squared elements of the GLCM.
- It quantifies the presence of repetitive patterns or regularity in texture.

- Higher energy values indicate more orderliness and repetitive patterns.

$$\text{Energy} = \sum_{i,j} p(i,j)^2 \tag{7}$$

viii. *Entropy* (Equation 8):

- Entropy measures the randomness or uncertainty in the distribution of pixel values within the image.
- It quantifies the degree of disorder or unpredictability in texture patterns.
- Higher entropy values indicate greater randomness and less predictability in texture.

$$\text{Ent} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j) \log(p(i,j)) \tag{8}$$

ix. *Homogeneity* (Equation 9):

- Homogeneity measures the closeness of the distribution of GLCM elements to the diagonal.
- It quantifies the similarity or homogeneity of texture within the image.
- Higher homogeneity values indicate more uniform and homogeneous texture.

$$Homog = \sum_i \sum_j \frac{1}{1 + |i-j|^2} P(i,j) \tag{9}$$

x. *Maximum Probability* (Equation 10):

- The maximum probability of a certain pair of pixels occurring in the picture is measured.
- It quantifies the most frequently occurring texture pattern or pixel pair.
- Higher values indicate a higher probability of occurrence for a specific texture pattern.

$$Max.Probability = max.p(i,j) for\ all\ (i,j) \tag{10}$$

xi. *Sum of Squares Variance* (Equation 11):

- It quantifies the heterogeneity or variability in texture patterns based on gray level sums.
- Higher values indicate greater variability and dispersion in texture patterns.

$$Variance = \sum_{i=1}^{N_g} (i-\mu)^2 p(i,j)$$

(11)

xii. *Sum Average* (Equation 12):

- It quantifies the average gray level sum in the texture patterns.
- Higher values indicate a higher average gray level sum in the texture.

$$sum\ average = \sum_{i=2}^{2N_g} ip_{x+y^{(i)}} \quad (12)$$

7. Data Cleaning:

Managing and Resolving Missing Data: One of the most important phases in data analysis is handling missing values. Choosing whether to remove rows or columns with missing data or use sophisticated machine learning-based approaches to fill in the blanks will depend on the situation. The missing value is replaced by mean value.

Normalization: Scaling Numerical characteristics: Preventing the dominance of particular characteristics during model training by normalizing numeric features to make sure they are on a comparable scale. To handle this problem, the suggested model uses the Standard Scaler normalization technique in the context of sensor data traffic characteristics with different magnitudes. By using this technique, feature observations are converted to have a standard deviation of 1 and a mean incoming traffic value of 0. By doing this, biases from incoming traffic are removed while maintaining the statistical properties of the data. This scaling is implemented by using Equation 13 (the transformation function).

$$S_S(z) = \frac{O_V(z) - M(z)}{S_D(z)} \quad (13)$$

In such case, z is the feature Z is a member of the set {z_1,z_2,...,z_N}, and the standard score for z features is S_S (z). The observed value is the O_V. The feature mean is denoted by M. The standard deviation of the characteristics is denoted by S_D. Here is how M(z) and S_D (z) are calculated:

$$M(z) = \frac{\sum_{z=1}^{N} O_V(z)}{N} \quad (14)$$

$$S_D(z) = \sqrt{\frac{1}{N}\sum_{z=1}^{N} (O_V(z) - M(z))^2} \quad (15)$$

8. **Grey Wolf Optimization:** GWO is a metaheuristic optimization algorithm inspired by the social hierarchy and hunting behavior of grey wolves. It aims to optimize complex problems by mimicking the leadership hierarchy and collaborative hunting strategies observed in wolf packs. this algorithm is use to select the important features. Typically, grey wolves exhibit a preference for living in packs, with an average group size ranging from 5 to 12 individuals. Within the social structure of a grey wolf pack, there are stringent rules governing dominance and hierarchy. The pack consists of the following roles:

I. Alphas, serving as leaders, are responsible for decision-making, and their directives are followed by the entire pack.
II. Betas, subordinate to alphas, assist in decision and other group happenings. Betas can be either gender, and they often emerge as strong contenders for assuming the alpha role.
III. Omegas play the role of scapegoats and are required to submit to all remaining dominant wolves. They are the preceding in line to partake in meals.
IV. Alphas and betas may have authority over deltas, but deltas exert control over omegas. This group comprises scouts, sentinels, elders, hunters, and caretakers. Scouts vigilantly watch over territorial borders, warning the pack of potential threats. Sentinels are dedicated to safeguarding and ensuring the safety of the pack. Elders, having previously held alpha or beta roles, bring experience and wisdom. Hunters collaborate with alphas and betas in capturing

prey to sustain the pack. Caretakers, on the other hand, attend to the needs of weak, ill, and injured pack members.

The framework of Grey Wolf Optimization model, the most optimal solution is denoted as alpha (α). The subsequent two best solutions are referred to as beta (β) and delta (δ). The remaining candidate results are collectively designated as omega (ω). The strategy used for hunting is guided by these four entities—α, β, δ, and ω—each following one of the three aforementioned candidates. To simulate the encircling behavior necessary for the bunch to hunt prey, the mathematical modeling of this behavior is represented by the following equations, namely Eqs. (16)– (19).

$$\vec{P}(t+1) = \vec{P}\rho(t) + \vec{W}.\vec{Z} \quad (16)$$

In Eq. (17), $\vec{D}$ is defined, where t represents the iteration number, and $\vec{W}$ and $\vec{Y}$ are vectors of coefficients. $\vec{P}\rho$ denotes the prey position, while $\vec{P}$ signifies the grey wolf position.

$$\vec{W} = \left|\vec{Y}.\ \vec{P}\rho(t) - \vec{P}(t)\right| \quad (17)$$

The $\vec{W}.\vec{Y}$ vector are calculated as in equation 18 and 19

$$\vec{x} = 2x.\vec{r1} - x \quad (18)$$

$$\vec{Y} = 2\vec{r2} \quad (19)$$

The process involves linearly decreasing the parameter 'a' from 2 to 0 throughout the iterations, with 'r1' and 'r2' representing random vectors within the range [0, 1]. Generally, the alpha guides the hunting, with occasional participation from beta and delta. The updating of the wolves' positions is expressed as in Eq. (20).

$$\vec{P}(t+1) = \frac{\vec{P_1}+\vec{P_2}+\vec{P_3}}{3} \quad (20)$$

$\vec{P_1}, \vec{P_2}, \vec{P_3}$ are define as in equation (21)-(23), respectively.

$$\vec{P_1} = \left|\vec{P_\alpha} - \vec{W_1}.\vec{Z_1}\right|, \quad (21)$$

$$\vec{P_2} = \left|\vec{P_\beta} - \vec{W_2}.\vec{Z_1}\right|, \quad (22)$$

$$\vec{P_3} = \left|\vec{P_\gamma} - \vec{W_3}.\vec{Z_1}\right|, \quad (23)$$

Where $\vec{P_\alpha}, \vec{P_\beta}, \vec{P_\gamma}$ are the first three best solution in the swarm at a given iteration t, $\vec{W_1}, \vec{W_2}, \vec{W_3}$ are defined as in eq. (3), and $\vec{Z_\alpha}, \vec{Z_\beta}, \vec{Z_\delta}$ are defined using equation (24) – (26), respectively.

$$\vec{Z}_\alpha = \left|\vec{Y_1}.\ \vec{P_\alpha} - \vec{P}\right| \quad (24)$$

$$\vec{Z}_\beta = \left|\overrightarrow{Y_2}.\ \overrightarrow{P_\beta} - \vec{P}\right| \quad (25)$$

$$\vec{Z}_\gamma = \left|\overrightarrow{Y_3}.\ \overrightarrow{P_\gamma} - \vec{P}\right| \quad (26)$$

Where $\vec{Y}, \overrightarrow{Y_2}, \overrightarrow{Y_3}$ are define as in equation (17)

A concluding note regarding the GWO pertains to the adjustment of the parameter 'x,'. The parameter 'a' undergoes a linear update during each iteration, transitioning from 2 to 0 in accordance with Eq. (27).

$$\alpha = 2 - t\frac{2}{MaxIter} \quad (27)$$

Algorithm 1: M-Grey wolves optimization Model

Input:

n: Total amount grey wolves in the group,

N-Iter: count of repetitions for optimization.

Output:

Pα: Location of the best grey wolf,

f(Pα): Optimal fitness cost.

Initialization:

Randomly initialize positions for n grey wolves in the population.

Identify Alpha, Beta, and Delta:

a. Determine the solutions for Alpha, Beta, and Delta based on the fitness number.

Optimization Loop:

While the termination criteria are not satisfied, perform the following:

a. Update Positions:

i. Update the position of each wolf i in the pack using Eq. (17).

b. Update Parameters:

i. Adjust the parameters a, A, and C.

c. Evaluate Positions:

i. Compute the location of individual wolves.

d. Update:

i. Update the values of α, β, and δ.

End of Optimization Loop

Note: The algorithm iteratively updates the positions of the grey wolves, modifies parameters, evaluates fitness for each individual, and updates the alpha, beta, and delta solutions until the specified stopping criteria are met.

9. **Improved Grey Wolf Optimization (M-GWO):** Wolf locations are continually adjusted to any point in space using the GWO technique. Certain unique issues, such feature selection, have solutions that can only be found in the binary {0, 1} range, which is why a unique version of the GWO is needed. This paper proposes a novel M-GWO for the job of feature selection. Each wolf is drawn to the top three best solutions by the wolves updating equation, which is a function of three location vectors, xα, xβ, and xδ. The pool of solutions in the M-GWO is always in binary form, with every solution located on a hypercube's corner.

$$P_i^{t+1} = Crossover(P1, P2, P3) \quad (28)$$

where Crossover is required cross over among solutions x, y, and z and P1; P2; P3 Binary vectors, indicating the impact of the wolf movement towards alpha, beta, and delta wolves in sequence, are represented by P1, P2, and P3. The calculations for these vectors are determined through Eqs. (29), (32), and (35), respectively.

$$P_i^d = \begin{cases} 1 \; if \; (P_\alpha^d + bstep_\alpha^d) \geq 1 \\ 0 \; Otherwise \end{cases} \quad (29)$$

Here, P_α^d denotes the position vector of the alpha wolf in dimension d, and $bstep_\alpha^d$ is a binary step in dimension d, which can be computed according to the formula presented in Eq. (30).

$$bstep_\alpha^d = \begin{cases} 1 \; if \; cstep_\alpha^d \geq rand \\ 0 \; Otherwise \end{cases} \quad (30)$$

In this context, "rand" represents a randomly generated number from a uniform distribution. Meanwhile, $cstep_\alpha^d$ stands for the continuous-valued size of step associated with dimension (d). This step size can be determined by employing a sigmoidal function, as outlined in Eq. (31).

$$cstep_\alpha^d = \frac{1}{1+e^{-10(A_1^d(D_\alpha^d-0.5)}} \quad (31)$$

where W_1^d are calculated using Eqs. (15), and (21) in the dimension d.

$$P_2^d = \begin{cases} 1 \; if \; x_\beta^d \geq 1 \\ 0 \; Otherwise \end{cases} \quad (32)$$

In this scenario, P_2^d signifies the position vector of the beta wolf in dimension d, and $bstep_\alpha^d$ is a binary step in dimension d, the computation of which is specified in Eq. (33).

$$bstep_\beta^d = \begin{cases} 1 \; if \; cstep_\beta^d \geq rand \\ 0 \; Otherwise \end{cases} \quad (33)$$

In this context, "rand" denotes a randomly generated number from a uniform distribution. Additionally, $cstep_\alpha^d$ represents the continuous-valued size of step is associated with dimension d. The calculation of this step size involves the utilization of a sigmoid function, as described in Equation (34).

$$cstep_\beta^d = \frac{1}{1+e^{-10(A_1^d(D_\beta^d-0.5)}} \quad (34)$$

where W1 d, and Zd β are computed with Eqs. (15) and (21) in dimension d.

$$P_3^d = \begin{cases} 1 \; if \; (P_\delta^d + bstep_\delta^d) \geq 1 \\ 0 \; Otherwise \end{cases} \quad (35)$$

In this context, P_δ^d designates the positional vector used for the delta wolf in dimension d, while $bstep_\beta^d$ represents a major binary step in the dimension d. The computation of this binary step is outlined in Eq. (36). $bstep_\delta^d = \begin{cases} 1 \; if \; cstep_\delta^d \geq rand \\ 0 \; Otherwise \end{cases}$ (36)

Two different approaches are used to execute the changed version. The first technique uses binarization of steps toward the top three solutions, followed by a stochastic crossover between these three basic motions to calculate the updated position of the grey wolf. On the other hand, the second approach involves using a sigmoidal function to transform the continuously updated position. The resulting values are stochastically thresholded to ascertain the updated position of the grey wolf. Both methods for the modified Grey Wolf Optimization (M-GWO) are applied within the field of feature selection.

10. Result analysis:

Analyzing the results of a model using metrics like accuracy (ACC), Matthews correlation coefficient (MCC), and F1 score provides valuable insights into its performance. Accuracy gives an overall view of correct predictions, but it may be biased by imbalanced datasets. MCC offers a balanced measure, especially useful for imbalanced data, with values closer to 1 indicating better predictions. F1 score balances precision and recall, making it suitable when false positives and false negatives hold equal importance. Each metric contributes unique perspectives, guiding adjustments or optimizations based on the specific needs and characteristics of the dataset and problem domain.

GLCM Feature extraction

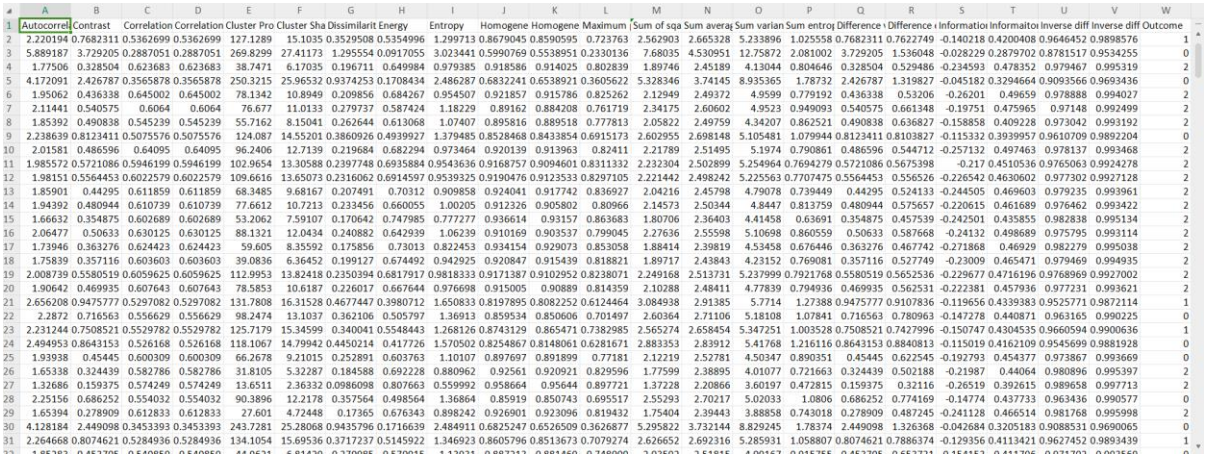
GLCM feature extraction: In the first step we have extracted the GLCM features using Matlab code. Indifferent GLCM features are discussed in the above section that we have extracted. GLCM features help us to figure out the lung cancer. The sample records are shown below in figure 3.

S.N	Autocorrel	Contrast	Correlation	Correlation	Cluster Pro	Cluster Sha	Dissimilarity	Energy	Entropy	Homogeneit	Homogeneit	Maximum p	Sum of squ	Sum averag	Sum varian	Sum entrop	Difference v	Difference e	Information m	Informaiton	Inverse diffe	Inverse diffe	Outcome
1	2.44218	0.9046722	0.5258718	0.5258718	141.6015	16.78738	0.4108035	0.4746637	1.468893	0.84718	0.8366847	0.6779527	2.851264	2.786022	5.599684	1.146449	0.9046722	0.8399917	-0.1306213	0.4295983	0.9589376	0.9880976	Bengin_cases
2	2.434147	0.881744	0.5334237	0.5334237	132.091	16.26532	0.4052658	0.4825256	1.453119	0.8489996	0.8385536	0.6841242	2.831757	2.778573	5.600168	1.134705	0.881744	0.833795	-0.1318477	0.4301686	0.9594314	0.9883024	Bengin_cases
3	2.42883	0.9125211	0.4981537	0.4981537	124.0559	14.82333	0.4338216	0.4412696	1.535675	0.8355353	0.8243565	0.6504845	2.841364	2.811353	5.350508	1.190744	0.9125211	0.8717501	-0.1153254	0.4134015	0.9562863	0.9878429	Bengin_cases
4	6.689918	4.487211	0.2303385	0.2303385	271.6991	26.32836	1.48553	0.06826681	3.227386	0.5594061	0.505501	0.1839154	8.855705	4.906512	14.64256	2.173487	4.487211	1.624973	-0.01948888	0.2480853	0.8621335	0.944384	Bengin_cases
5	7.077254	4.863946	0.217123	0.217123	272.5774	25.96149	1.559155	0.0634641	3.303189	0.5488795	0.4922875	0.1781997	9.427091	5.060736	15.65685	2.215591	4.863946	1.661846	-0.01759789	0.2383284	0.8565596	0.9401633	Bengin_cases
6	6.054184	4.036836	0.2205962	0.2205962	237.2	23.49078	1.401557	0.07545082	3.117321	0.5712275	0.5201363	0.1944125	7.996971	4.683125	12.97201	2.104375	4.036836	1.579556	-0.01911186	0.2409759	0.8683887	0.9492581	Bengin_cases
7	6.310974	4.170372	0.2333727	0.2333727	254.8382	25.06086	1.424772	0.07327908	3.156728	0.5684214	0.5166515	0.1920764	8.322267	4.764964	13.64087	2.132214	4.170372	1.59318	-0.02033994	0.2500381	0.8667258	0.9478369	Bengin_cases
8	6.310974	4.170372	0.2333727	0.2333727	254.8382	25.06086	1.424772	0.07327908	3.156728	0.5684214	0.5166515	0.1920764	8.322267	4.764964	13.64087	2.132214	4.170372	1.59318	-0.02033994	0.2500381	0.8667258	0.9478369	Bengin_cases
9	6.543733	4.537297	0.2006785	0.2006785	249.2456	24.04225	1.504756	0.06746807	3.220636	0.5550038	0.4996245	0.1823836	8.736038	4.888419	14.26165	2.159702	4.537297	1.632251	-0.01728202	0.233263	0.8603616	0.9435966	Bengin_cases
10	6.543733	4.537297	0.2006785	0.2006785	249.2456	24.04225	1.504756	0.06746807	3.220636	0.5550038	0.4996245	0.1823836	8.736038	4.888419	14.26165	2.159702	4.537297	1.632251	-0.01728202	0.233263	0.8603616	0.9435966	Bengin_cases
11	6.430398	4.253799	0.2345454	0.2345454	256.6226	25.17184	1.438848	0.07217425	3.178117	0.5670626	0.5147636	0.1919884	8.479082	4.807782	13.94314	2.1465	4.253799	1.601828	-0.02030276	0.2510932	0.8657413	0.9469488	Bengin_cases
12	6.642247	4.424577	0.2312555	0.2312555	266.5517	25.88718	1.475753	0.06871936	3.218314	0.5603571	0.5067009	0.1845722	8.7779	4.889478	14.49349	2.167929	4.424577	1.619075	-0.01985135	0.2495471	0.8627942	0.945017	Bengin_cases
13	2.443239	0.8893306	0.5311091	0.5311091	134.3469	16.36884	0.4070389	0.4775697	1.464226	0.84844	0.8379317	0.6803807	2.844492	2.785375	5.600544	1.143253	0.8893306	0.8358613	-0.132428	0.432008	0.9592744	0.9882371	Bengin_cases
14	6.749646	4.600214	0.2096713	0.2096713	257.4946	24.64919	1.517413	0.06538509	3.25064	0.5529712	0.4971454	0.1788603	8.972719	4.955576	14.74958	2.179177	4.600214	1.638093	-0.0178287	0.2381823	0.8593829	0.9429009	Bengin_cases
15	2.438957	0.7937423	0.5510406	0.5510406	111.0351	14.40919	0.4036267	0.456714	1.490571	0.8429477	0.8330175	0.6626283	2.787831	2.793961	5.398046	1.16464	0.7937423	0.8359871	-0.134943	0.439842	0.9588959	0.9891619	Bengin_cases
16	2.419604	0.7804458	0.553145	0.553145	111.6455	14.38701	0.3964805	0.4609444	1.474902	0.8454166	0.835876	0.6658854	2.762217	2.782943	5.364241	1.155023	0.7804458	0.825756	-0.138245	0.4416352	0.9596168	0.9893684	Bengin_cases
17	2.419604	0.7804458	0.553145	0.553145	111.6455	14.38701	0.3964805	0.4609444	1.474902	0.8454166	0.835876	0.6658854	2.762217	2.782943	5.364241	1.155023	0.7804458	0.825756	-0.138245	0.4416352	0.9596168	0.9893684	Bengin_cases
18	2.502512	0.8258023	0.543903	0.543903	112.5546	14.5148	0.4273418	0.428126	1.554479	0.8320999	0.8218619	0.6380936	2.867272	2.835403	5.442789	1.208617	0.8258023	0.8641178	-0.1282698	0.4371191	0.9563405	0.9887016	Bengin_cases
19	2.502512	0.8258023	0.543903	0.543903	112.5546	14.5148	0.4273418	0.428126	1.554479	0.8320999	0.8218619	0.6380936	2.867272	2.835403	5.442789	1.208617	0.8258023	0.8641178	-0.1282698	0.4371191	0.9563405	0.9887016	Bengin_cases
20	2.541425	0.8663028	0.5340625	0.5340625	115.0242	14.75164	0.4415939	0.4167293	1.58892	0.8275567	0.8168325	0.6285309	2.925947	2.859881	5.504383	1.231604	0.8663028	0.8810993	-0.1237245	0.4342507	0.9549914	0.9882026	Bengin_cases
21	2.460212	0.8333984	0.5084865	0.5084865	100.329	12.82001	0.452742	0.3933825	1.607137	0.818396	0.8078294	0.6056794	2.829192	2.848828	5.143705	1.240959	0.8333984	0.8912772	-0.1084421	0.4071245	0.9533782	0.9884776	Bengin_cases
22	2.50181	0.8497281	0.5133732	0.5133732	102.8621	13.14044	0.453596	0.395248	1.621361	0.8199708	0.8088438	0.6090227	2.878524	2.865935	5.244955	1.253012	0.8497281	0.894101	-0.1136837	0.4152076	0.9534382	0.9882723	Bengin_cases
23	2.531733	0.8569068	0.5179459	0.5179459	104.4903	13.41513	0.4564319	0.3916833	1.633646	0.8189243	0.807829	0.6057521	2.912744	2.878315	5.312643	1.261693	0.8569068	0.8974532	-0.1144229	0.4201261	0.953163	0.9881769	Bengin_cases

Figure 3: Extracted GLCM features

To shuffle dataset and perform label encoding, we can leverage scikit-learn's utilities. Utilize sklearn.utils.shuffle to randomize the order of your dataset, ensuring a balanced representation across training and testing sets. Concurrently, employ sklearn.preprocessing. LabelEncoder to transform categorical labels into numerical values, facilitating compatibility with machine learning algorithms. This process streamlines data preparation, enhancing model performance and interpretability while maintaining data integrity and consistency

throughout the analysis pipeline.shuffling of GLCM dataset & output lable encoding. The sample records are shown below in figure 4.



Autocorrel	Contrast	Correlation	Correlation	Cluster Pro	Cluster Sha	Dissimilarit	Energy	Entropy	Homogene	Homogene	Maximum	Sum of sqa	Sum avera	Sum varian	Sum entro	Difference	Difference	Informatio	Informatio	Inverse diff	Inverse diff	Outcome
2.220194	0.7682311	0.5362699	0.5362699	127.1289	15.1035	0.3529508	0.5354996	1.299713	0.8679045	0.8590595	0.723763	2.562903	2.665328	5.233896	1.025558	0.7682311	0.7622749	-0.140218	0.4200408	0.9646452	0.9898576	1
5.889187	3.729205	0.2887051	0.2887051	269.8299	27.41173	1.295554	0.0917055	3.023441	0.5990769	0.5538951	0.2330136	7.68035	4.530951	12.75872	2.081002	3.729205	1.536048	-0.028229	0.2879702	0.8781517	0.9534255	0
1.77506	0.328504	0.623683	0.623683	38.7471	6.17035	0.196711	0.649984	0.979385	0.918586	0.914025	0.802839	1.89746	2.45189	4.13044	0.804646	0.328504	0.529486	-0.234593	0.478352	0.979467	0.995319	2
4.172091	2.426787	0.3565878	0.3565878	250.3215	25.96532	0.9374253	0.1708434	2.486287	0.6832241	0.6538921	0.3605622	5.328346	3.74145	8.935365	1.78732	2.426787	1.319827	-0.045182	0.3294664	0.9093566	0.9693436	0
1.95062	0.436338	0.645002	0.645002	78.1342	10.8949	0.209856	0.684267	0.954507	0.921857	0.915786	0.825262	2.12949	2.49372	4.9599	0.779192	0.436338	0.53206	-0.26201	0.49659	0.978888	0.994027	2
2.11441	0.540575	0.6064	0.6064	76.677	11.0133	0.279737	0.587424	1.18229	0.89162	0.884208	0.761719	2.34175	2.60602	4.9523	0.949093	0.540575	0.661348	-0.19751	0.475965	0.97148	0.992499	2
1.85392	0.490838	0.545239	0.545239	55.7162	8.15041	0.262644	0.613068	1.07407	0.895816	0.889518	0.777813	2.05822	2.49759	4.34207	0.862521	0.490838	0.636827	-0.158858	0.409228	0.973042	0.993192	2
2.238639	0.8123411	0.5075576	0.5075576	124.087	14.55201	0.3860926	0.4939927	1.379485	0.8528468	0.8433854	0.6915173	2.602955	2.698148	5.105481	1.079944	0.8123411	0.8103827	-0.115332	0.3939957	0.9610709	0.9892204	0
2.01581	0.486596	0.64095	0.64095	96.2406	12.7139	0.219684	0.682294	0.973464	0.920139	0.913963	0.82411	2.21789	2.51495	5.1974	0.790861	0.486596	0.544712	-0.257132	0.497463	0.978137	0.993468	2
1.985572	0.5721086	0.5946199	0.5946199	102.9654	13.30588	0.2397748	0.6935884	0.9543636	0.9168757	0.9094601	0.8311332	2.232304	2.502899	5.254964	0.7694279	0.5721086	0.5675398	-0.217	0.4510536	0.9765063	0.9924278	2
1.98151	0.5564453	0.6022579	0.6022579	109.6616	13.65073	0.2316062	0.6914597	0.9539325	0.9190476	0.9123533	0.8297105	2.221442	2.498242	5.225563	0.7707475	0.5564453	0.556526	-0.226542	0.4630602	0.977302	0.9927128	2
1.85901	0.44295	0.611859	0.611859	68.3485	9.68167	0.207491	0.70312	0.909858	0.924041	0.917742	0.836927	2.04216	2.45798	4.79078	0.739449	0.44295	0.524133	-0.244505	0.469603	0.979235	0.993961	2
1.94392	0.480944	0.610739	0.610739	77.6612	10.7213	0.233456	0.660055	1.00205	0.912326	0.905802	0.80966	2.14573	2.50344	4.8447	0.813759	0.480944	0.575657	-0.220615	0.461689	0.976462	0.993422	2
1.66632	0.354875	0.602689	0.602689	53.2062	7.59107	0.170642	0.747985	0.777277	0.936614	0.93157	0.863683	1.80706	2.36403	4.41458	0.63691	0.354875	0.457539	-0.242501	0.435855	0.982838	0.995134	2
2.06477	0.50633	0.630125	0.630125	88.1321	12.0434	0.240882	0.642939	1.06239	0.910169	0.903537	0.799045	2.27636	2.55598	5.10698	0.860559	0.50633	0.587668	-0.24132	0.498689	0.975795	0.993114	2
1.73946	0.363276	0.624423	0.624423	59.605	8.35592	0.175856	0.73013	0.822453	0.934154	0.929073	0.853058	1.88414	2.39819	4.53458	0.676446	0.363276	0.467742	-0.271868	0.46929	0.982279	0.995038	2
1.75839	0.357116	0.603603	0.603603	39.0836	6.36452	0.199127	0.674492	0.942925	0.920847	0.915439	0.818821	1.89717	2.43843	4.23152	0.769081	0.357116	0.527749	-0.23009	0.465471	0.979469	0.994935	2
2.008739	0.5580519	0.6059625	0.6059625	112.9953	13.82418	0.2350394	0.6817917	0.9818333	0.9171387	0.9102952	0.8238071	2.249168	2.513731	5.237999	0.7921768	0.5580519	0.5652536	-0.229677	0.4716196	0.9768969	0.9927002	2
1.90642	0.469935	0.607643	0.607643	78.5853	10.6187	0.226017	0.667644	0.976698	0.915005	0.90889	0.814359	2.10288	2.48411	4.77839	0.794936	0.469935	0.562531	-0.222381	0.457936	0.977231	0.993621	2
2.656208	0.9475777	0.5297082	0.5297082	131.7808	16.31528	0.4677447	0.3980712	1.650833	0.8197895	0.8082252	0.6124464	3.084938	2.91385	5.7714	1.27388	0.9475777	0.9107836	-0.119656	0.4339383	0.9525771	0.9872114	1
2.2872	0.716563	0.556629	0.556629	98.2474	13.1037	0.362106	0.505797	1.36913	0.859534	0.850606	0.701497	2.60364	2.71106	5.18108	1.07841	0.716563	0.780963	-0.147278	0.440871	0.963165	0.990225	0
2.231244	0.7508521	0.5529782	0.5529782	125.7179	15.34599	0.340041	0.5548443	1.268126	0.8743129	0.865471	0.7382985	2.565274	2.658454	5.347251	1.003528	0.7508521	0.7427996	-0.150747	0.4304535	0.9660594	0.9900636	1
2.494953	0.8643153	0.526168	0.526168	118.1067	14.79942	0.4450214	0.417726	1.570502	0.8254867	0.8148061	0.6281671	2.883353	2.83912	5.41768	1.216116	0.8643153	0.8840813	-0.115019	0.4162109	0.9545699	0.9881928	0
1.93938	0.45445	0.600309	0.600309	66.2678	9.21015	0.252891	0.603763	1.10107	0.897697	0.891899	0.77181	2.12219	2.52781	4.50347	0.890351	0.45445	0.622545	-0.192793	0.454377	0.973867	0.993669	0
1.65338	0.324439	0.582786	0.582786	31.8105	5.32287	0.184588	0.692228	0.880962	0.92561	0.920921	0.829596	1.77599	2.38895	4.01077	0.721663	0.324439	0.502188	-0.21987	0.44064	0.980896	0.995397	2
1.32686	0.159375	0.574249	0.574249	13.6511	2.36332	0.0986098	0.807663	0.559992	0.958664	0.95644	0.897721	1.37228	2.20866	3.60197	0.472815	0.159375	0.32116	-0.26519	0.392615	0.989658	0.997713	2
2.25156	0.686252	0.554032	0.554032	90.3896	12.2178	0.357564	0.498564	1.36864	0.85919	0.850743	0.695517	2.55293	2.70217	5.02033	1.0806	0.686252	0.774169	-0.14774	0.437733	0.963436	0.990577	0
1.65394	0.278909	0.612833	0.612833	27.601	4.72448	0.17365	0.676343	0.898242	0.926901	0.923096	0.819432	1.75404	2.39443	3.88858	0.743018	0.278909	0.487245	-0.241128	0.466514	0.981768	0.995998	2
4.128184	2.449098	0.3453393	0.3453393	243.7281	25.28068	0.9435796	0.1716639	2.484911	0.6825247	0.6526509	0.3626877	5.295822	3.732144	8.829245	1.78374	2.449098	1.326368	-0.042684	0.3205183	0.9088531	0.9690065	0
2.264668	0.8074621	0.5284936	0.5284936	134.1054	15.69536	0.3717237	0.5145922	1.346923	0.8605796	0.8513673	0.7079274	2.626652	2.692316	5.285931	1.058807	0.8074621	0.7886374	-0.129356	0.4113421	0.9627452	0.9893439	1

Figure 4: shuffle dataset

Feature selection using M-GWO: Feature selection using the M-GWO (Modified Grey Wolf Optimizer) algorithm involves a metaheuristic approach to identify the most relevant features from a given dataset. The M-GWO algorithm is an enhancement of the traditional Grey Wolf Optimizer (GWO), integrating modified strategies to improve feature selection accuracy. figure 5 provide the code to perform feature selection using M-GWO.

```
def GWO(SearchAgents_no,Max_iter,ub,lb,dim,Cost_fun,X,y):

    Alpha_pos=np.zeros(dim)
    Alpha_score=np.inf

    Beta_pos=np.zeros(dim)
    Beta_score=np.inf

    Delta_pos=np.zeros(dim)
    Delta_score=np.inf

    Positions=initialization(SearchAgents_no,dim)
    # print(Positions)
    Y1=np.zeros(dim)
    Y2=np.zeros(dim)
    Y3=np.zeros(dim)


    Convergence_curve=np.zeros(Max_iter)
    l=0
    while l<Max_iter:
        for i in range(0,SearchAgents_no):
            Flag4ub=Positions[i]>ub
            Flag4lb=Positions[i]<lb
            Positions[i]=(Positions[i]*(~(Flag4ub+Flag4lb)))+ub*Flag4ub+lb*Flag4lb
#             print(Positions[i])
```

Figure 5:code to perform feature selection using M-GWO.

Fine-tune the M-GWO algorithm by adjusting parameters like the number of iterations (num_iter), population size, and convergence criteria. Shown in figure 6.

```
Agents=25
MaxIter=10
ub=1
lb=0
dim=X_train.shape[1]

Best_score,Best_pos,Cg=GWO(Agents,MaxIter,1,0,dim,CostFunction,X,y)
```

```
Iteration 1 -- 0.12700000643730164 Num feature: 11.0
Iteration 2 -- 0.12700000643730164 Num feature: 11.0
Iteration 3 -- 0.12700000643730164 Num feature: 11.0
Iteration 4 -- 0.12250001072883605 Num feature: 15.0
Iteration 5 -- 0.12250001072883605 Num feature: 15.0
Iteration 6 -- 0.12250001072883605 Num feature: 15.0
Iteration 7 -- 0.11800001502037048 Num feature: 15.0
Iteration 8 -- 0.11800001502037048 Num feature: 15.0
Iteration 9 -- 0.11800001502037048 Num feature: 15.0
Iteration 10 -- 0.11800001502037048 Num feature: 15.0
```

```
plt.plot(Cg)
plt.show()
Selected=Best_pos
print(Selected)
```

Figure 6:M-GWO algorithm adjusting parameters

Iterate this process to achieve optimal feature subsets that maximize model performance. Adjust parameters and experiment with different evaluation metrics to optimize feature selection outcome shown in figure 7.

	A	B	C	D	E	F	G	H	I	J	K
1	Autocorrela	Contrast	Correlation	Cluster Pror	Cluster Shac	Entropy	Homogenei	Sum of squ	Sum varianc	Sum entrop	Outcome
2	2.220194	0.7682311	0.5362699	127.1289	15.1035	1.299713	0.8590595	2.562903	5.233896	1.025558	1
3	5.889187	3.729205	0.2887051	269.8299	27.41173	3.023441	0.5538951	7.68035	12.75872	2.081002	0
4	1.77506	0.328504	0.623683	38.7471	6.17035	0.979385	0.914025	1.89746	4.13044	0.804646	2
5	4.172091	2.426787	0.3565878	250.3215	25.96532	2.486287	0.6538921	5.328346	8.935365	1.78732	0
6	1.95062	0.436338	0.645002	78.1342	10.8949	0.954507	0.915786	2.12949	4.9599	0.779192	2
7	2.11441	0.540575	0.6064	76.677	11.0133	1.18229	0.884208	2.34175	4.9523	0.949093	2
8	1.85392	0.490838	0.545239	55.7162	8.15041	1.07407	0.889518	2.05822	4.34207	0.862521	2
9	2.238639	0.8123411	0.5075576	124.087	14.55201	1.379485	0.8433854	2.602955	5.105481	1.079944	0
10	2.01581	0.486596	0.64095	96.2406	12.7139	0.973464	0.913963	2.21789	5.1974	0.790861	2
11	1.985572	0.5721086	0.5946199	102.9654	13.30588	0.9543636	0.9094601	2.232304	5.254964	0.7694279	2
12	1.98151	0.5564453	0.6022579	109.6616	13.65073	0.9539325	0.9123533	2.221442	5.225563	0.7707475	2
13	1.85901	0.44295	0.611859	68.3485	9.68167	0.909858	0.917742	2.04216	4.79078	0.739449	2
14	1.94392	0.480944	0.610739	77.6612	10.7213	1.00205	0.905802	2.14573	4.8447	0.813759	2
15	1.66632	0.354875	0.602689	53.2062	7.59107	0.777277	0.93157	1.80706	4.41458	0.63691	2
16	2.06477	0.50633	0.630125	88.1321	12.0434	1.06239	0.903537	2.27636	5.10698	0.860559	2
17	1.73946	0.363276	0.624423	59.605	8.35592	0.822453	0.929073	1.88414	4.53458	0.676446	2
18	1.75839	0.357116	0.603603	39.0836	6.36452	0.942925	0.915439	1.89717	4.23152	0.769081	2
19	2.008739	0.5580519	0.6059625	112.9953	13.82418	0.9818333	0.9102952	2.249168	5.237999	0.7921768	2
20	1.90642	0.469935	0.607643	78.5853	10.6187	0.976698	0.90889	2.10288	4.77839	0.794936	2
21	2.656208	0.9475777	0.5297082	131.7808	16.31528	1.650833	0.8082252	3.084938	5.7714	1.27388	1
22	2.2872	0.716563	0.556629	98.2474	13.1037	1.36913	0.850606	2.60364	5.18108	1.07841	0
23	2.231244	0.7508521	0.5529782	125.7179	15.34599	1.268126	0.865471	2.565274	5.347251	1.003528	1
24	2.494953	0.8643153	0.526168	118.1067	14.79942	1.570502	0.8148061	2.883353	5.41768	1.216116	0

Figure 7: Optimize feature selection outcome

Figure 8, helps to determine how quickly the algorithm reaches a stable solution, whether there are any oscillations or fluctuations in the optimization process, and whether the algorithm exhibits any convergence and divergence patterns.

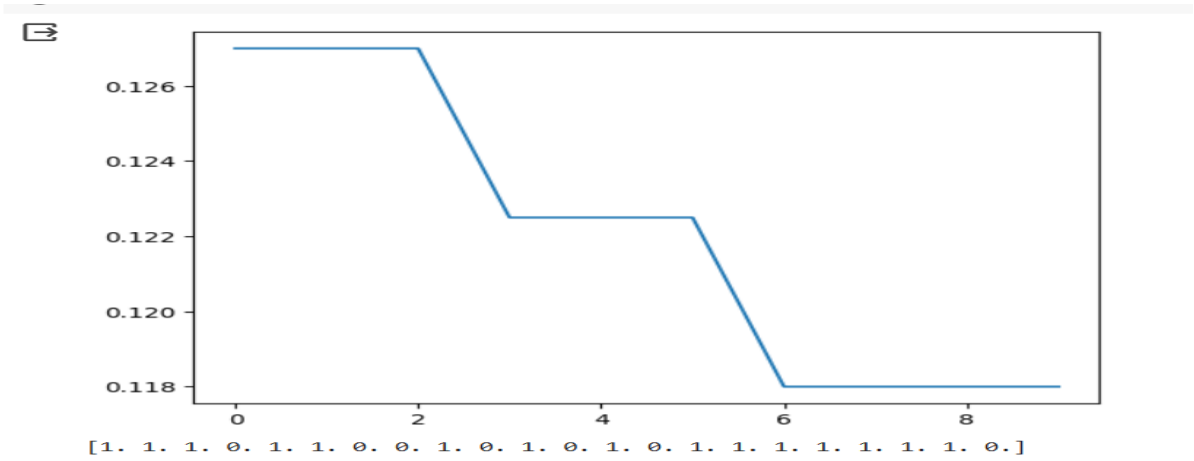


Figure 8: convergence and divergence patterns

11. Data Preprocessing:

This section incorporates data preprocessing steps all the listed methodology is implemented as in methodology section. We have implemented M-GWO for the feature selection. Figure 9 determine class count at output label. The dataset contains three different class, Bengin_cases 120 instant, Malignant_cases 561 instance and Normal_cases 561 instants.

```
df['Outcome'].value_counts()
```

```
Outcome
2    561
0    416
1    120
Name: count, dtype: int64
```

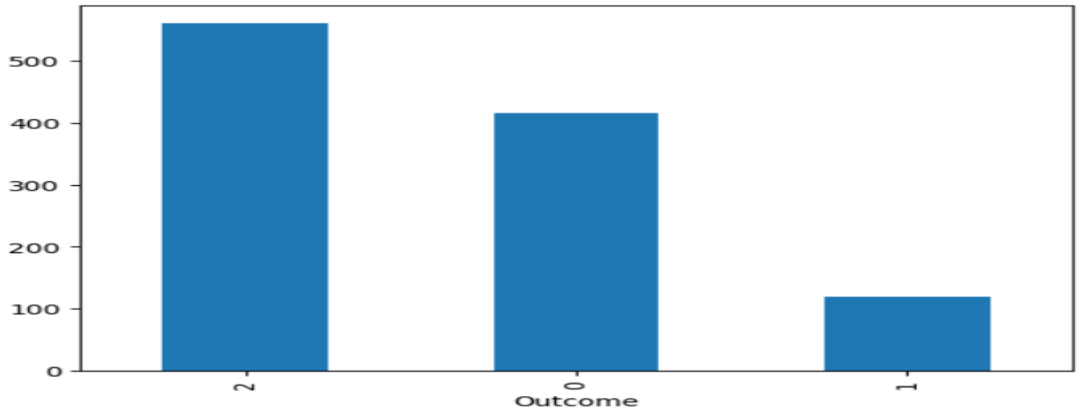


Figure 9:Count of output label

A correlation matrix is shown in figure 10, displaying the correlation values among the variables. The correlation between two variables is displayed in each cell of the table. The number falls between -1 and 1. A complete positive linear link exists between two variables if their correlation coefficient is 1.

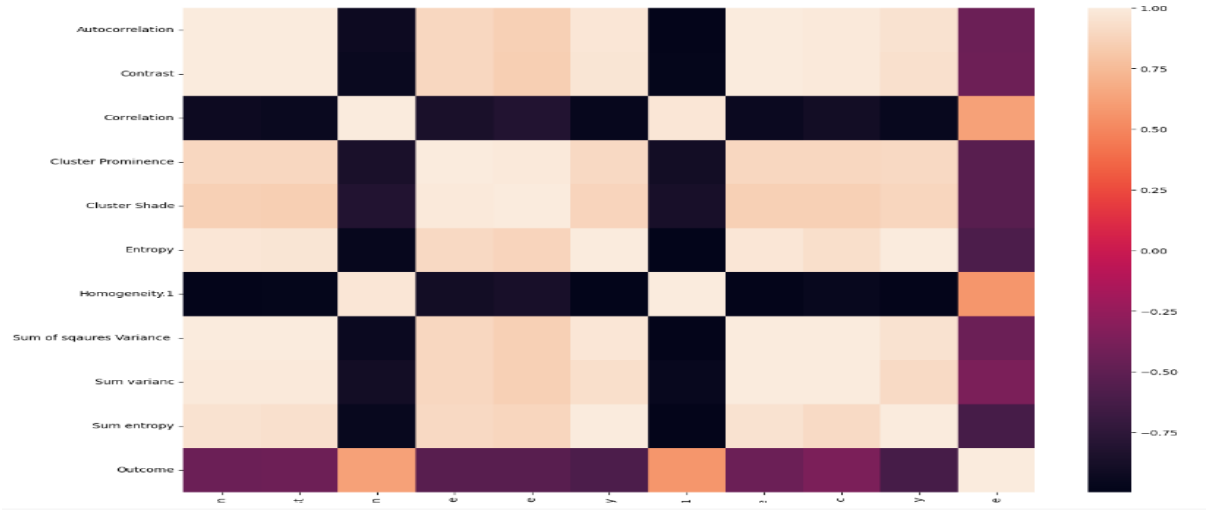


Figure 10: correlation matrix

Artificial Neural Network(ANN): This computational model is inspired by biological neural networks, such as those found in the human brain, in terms of both form and functionality. ANNs are used in artificial intelligence and machine learning to handle complicated data inputs and produce outputs that are dependent on correlations and patterns found in the data. The input layer, hidden layers, and output layer are the layers of linked nodes, or neurons, that make up these networks and collaborate to learn and make predictions or judgments. Figure 11 (a) and (b) demonstrate ANN had approx 84 % training and validation accuracy and high validation loss. This indicate that ANN is not fit for this dataset.

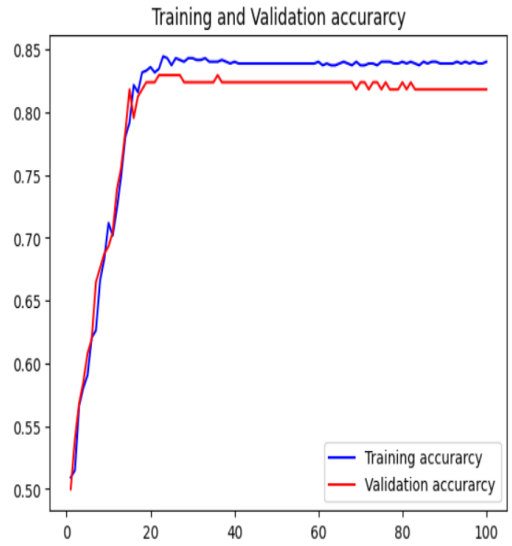


Figure 11 a) training and validation accuracy

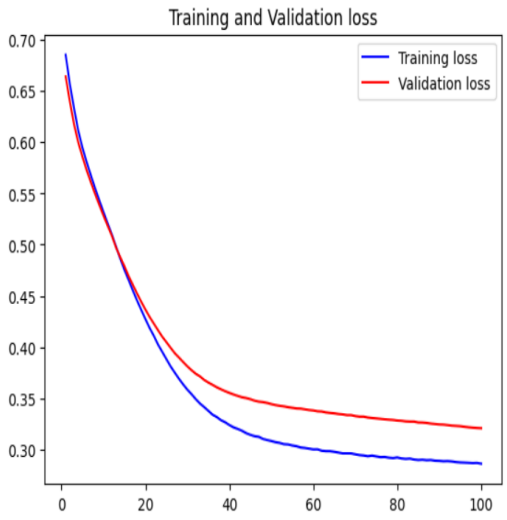
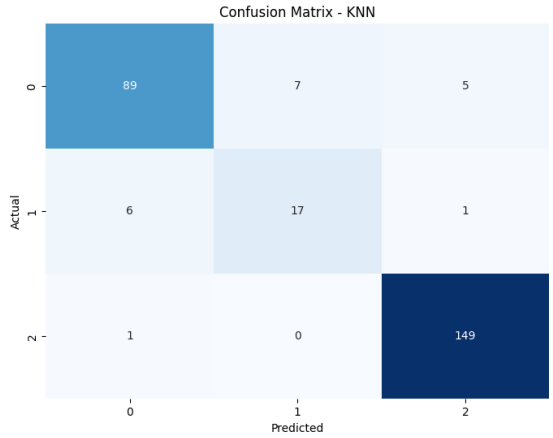


Figure 11 b) Training and validation loss

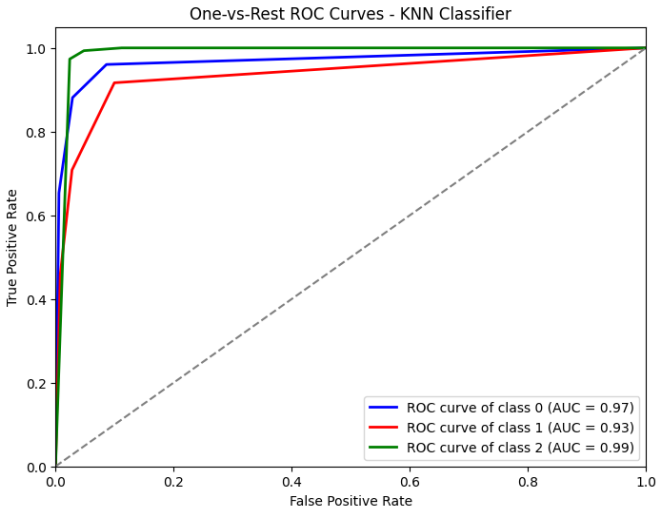
KNN: A straightforward yet effective machine learning approach for classification and regression problems is K-Nearest Neighbors (KNN). When predicting a new data point, the "K" in KNN stands for the number of nearest neighbors to take into account. Figure 12 a) demonstrate the confusion matrix score, b)train data set result achieved a high accuracy of 0.960, indicating that it correctly classified 96.0% of the samples. The MCC of 0.932 suggests a strong correlation between predicted and actual values, and the F1 score of 0.959 reflects a good balance between precision and recall and test data set result, KNN achieved an accuracy of 0.927, an MCC of 0.870, and an F1 score of 0.927. This ROC curve indicates performance across all metrics, especially in terms of accuracy, F1-score and MCC.



a) KNN Confusion matrix

```
Model performance for Training set
- Accuracy: 0.9598540145985401
- MCC: 0.9316296169752784
- F1 score: 0.9591022920826962
---------------------------------
Model performance for Test set
- Accuracy: 0.9272727272727272
- MCC: 0.8699024088436459
- F1 score: 0.9266045829014956
```

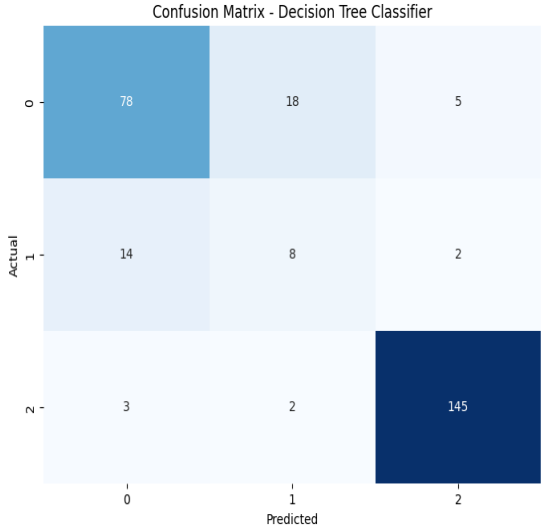
b) Training and Testing results of KNN model



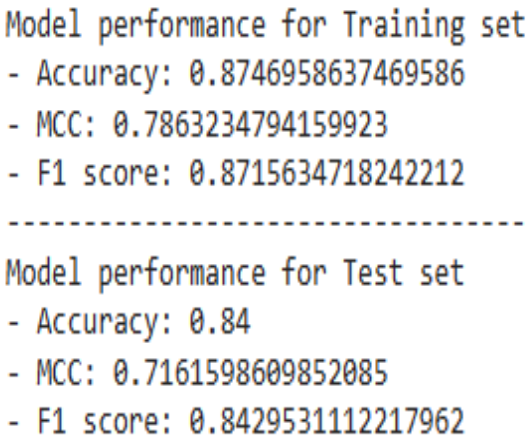
c) ROC curve for KNN

Figure 12: Overall KNN results

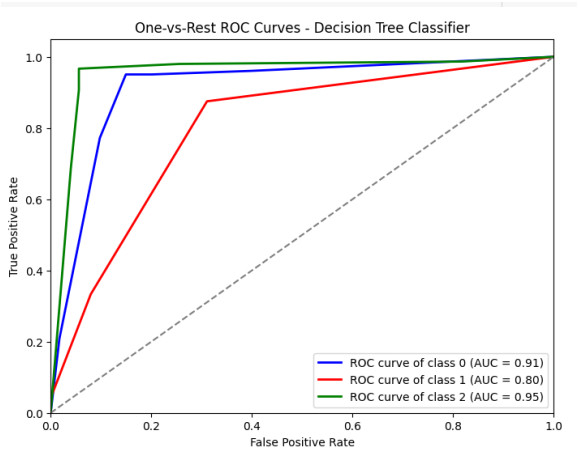
Decision Tree (DT): A Decision Tree is a hierarchical structure where each internal node represents a decision based on a feature, and each leaf node represents the outcome or class label.Figure 13 a) demonstrate the confusion matrix score, b)train data set result, DT performed well with an accuracy of 0.875, an MCC of 0.786, and an F1 score of 0.872. This suggests that the decision tree model was effective in capturing the underlying patterns in the data and making accurate predictions and test result achieved an accuracy of 0.840, an MCC of 0.716, and an F1 score of 0.843. It shows reasonable performance but falls short compared to KNN and the stacked model in terms of MCC and F1 score. This ROC curve indicates performance across all metrics, especially in terms of accuracy, F1-score and MCC.



a) DT Confusion matrix



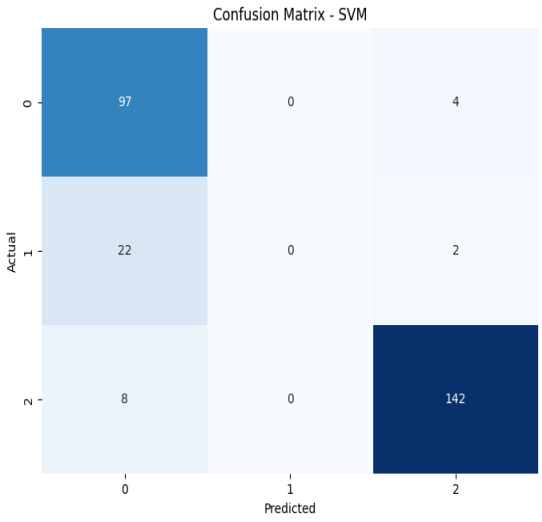
b) Training and Testing results of DT model



c) ROC curve for KNN

Figure 13: Overall DT results

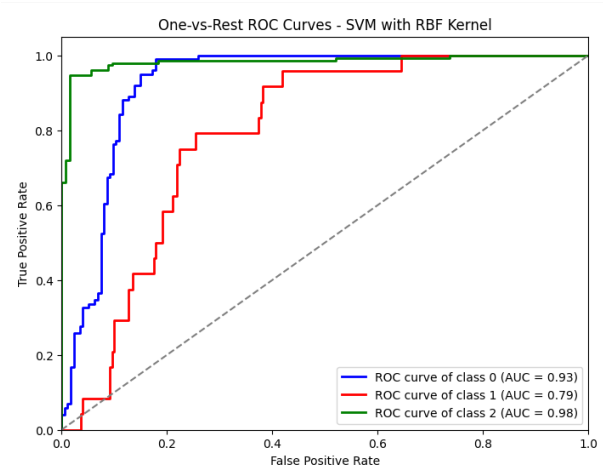
SVM: The data points that are closest to the decision border are known as support vectors, and they are essential in determining the margin. SVMs are memory-efficient for big datasets because they only use a subset of training data points (called support vectors) that are close to the decision border. Figure 14 a) demonstrate the confusion matrix score, b)train data set result, had a lower accuracy of 0.831 compared to KNN. The MCC of 0.715 and F1 score of 0.783 also indicate comparatively lower performance in terms of both correlation and balance between precision and recall and test dataset result had an accuracy of 0.869, an MCC of 0.769, and an F1 score of 0.832. While the accuracy is decent, the MCC and F1 score are slightly lower compared to KNN. This ROC curve indicates performance across all metrics, especially in terms of accuracy, F1-score and MCC.



a) SVM Confusion matrix

```
Model performance for Training set
- Accuracy: 0.8309002433090025
- MCC: 0.7148634893641425
- F1 score: 0.7829141092355728
---------------------------------
Model performance for Test set
- Accuracy: 0.8690909090909091
- MCC: 0.7693888948793326
- F1 score: 0.8323331513652955
```

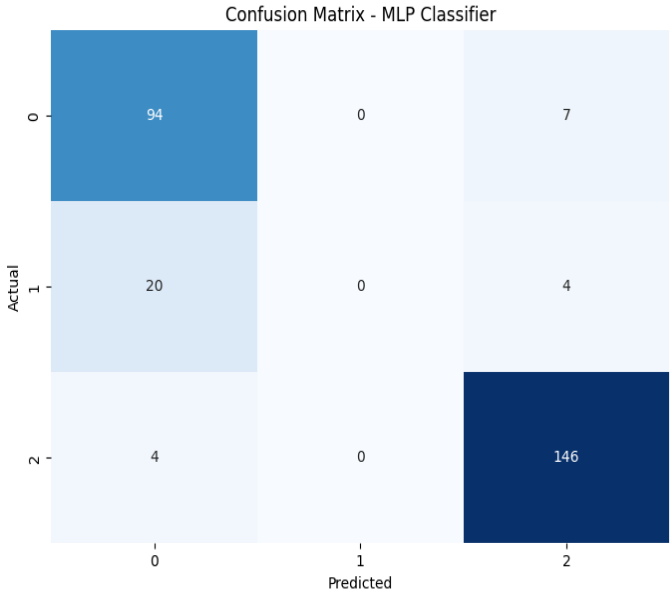
b) Training and Testing results of SVM model



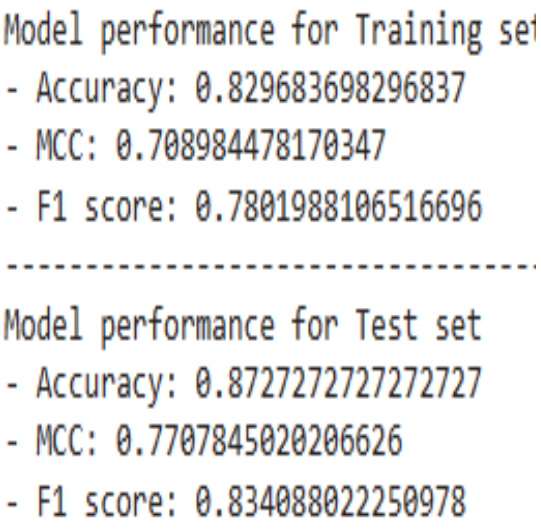
c) ROC curve for SVM

Figure 14: Overall SVM results

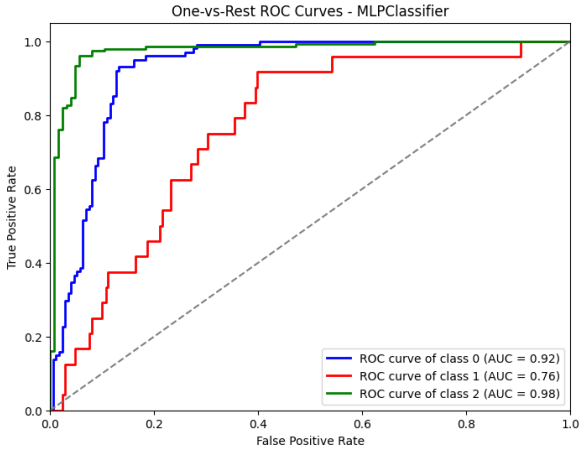
MLP: An artificial neural network with many layers of nodes (neurons) comprising an input layer, one or more hidden layers, and an output layer is called a multilayer perceptron (MLP). Figure 15 a) demonstrate the confusion matrix score, b)train data set result, r performance to SVM RBF, with an accuracy of 0.830, an MCC of 0.709, and an F1 score of 0.780. While MLP is a powerful model, it seems to have slightly underperformed in this context compared to other models and test dataset result had an accuracy of 0.873, an MCC of 0.771, and an F1 score of 0.780. It performs well in terms of accuracy but shows a lower F1 score compared to other models.This ROC curve indicates performance across all metrics, especially in terms of accuracy, F1-score and MCC.



a) MLP Confusion matrix



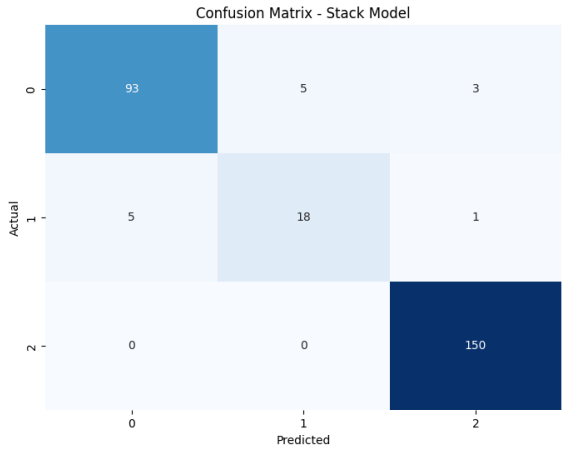
b) Training and Testing results of MLP model



c) ROC curve for MLP

Figure 15:Overall MLP results

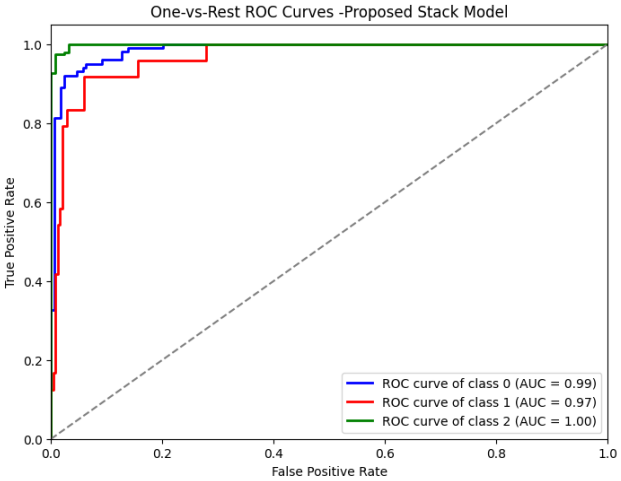
Proposed Model: Proposed stack model: stacked models are a powerful technique in machine learning ensembles, offering enhanced predictive performance by leveraging the strengths of multiple base models and effectively combining their predictions through a meta-model. Figure 16 a) demonstrate the confusion matrix score, b)train data set result demonstrated exceptional performance across all metrics, with an accuracy of 0.998, an MCC of 0.996, and an F1 score of 0.998. This indicates near-perfect classification and a very strong correlation between predicted and actual values and test dataset result outperformed all other models with an accuracy of 0.949, an MCC of 0.909, and an F1 score of 0.948. It demonstrates superior performance across all metrics, indicating the effectiveness of ensemble methods or stacked models in improving predictive accuracy and robustness. This ROC curve indicates performance across all metrics, especially in terms of accuracy, F1-score and MCC.



a) Proposed Model Confusion matrix

```
Model performance for Training set
- Accuracy: 0.9975669099756691
- MCC: 0.9958787687103409
- F1 score: 0.9975579540646868
----------------------------------
Model performance for Test set
- Accuracy: 0.9490909090909091
- MCC: 0.9087985273920542
- F1 score: 0.9484047438659242
```

b) Training and Testing results of Proposed Model



c) ROC curve for Proposed Model

Figure 16: Overall proposed model results

The results analysis reveals in figure 17, the proposed model significantly outperforms individual machine learning models in terms of accuracy, Matthews Correlation Coefficient (MCC), and F1 score. The proposed model achieved an outstanding accuracy of 0.998, a high MCC of 0.996, and an impressive F1 score of 0.998, indicating near-perfect classification and a strong correlation between predicted and actual values. In comparison, while KNN showed high accuracy and MCC, the proposed model surpassed it significantly. Similarly, SVM and MLP exhibited lower performance metrics, highlighting the superior predictive capabilities of the proposed model. These results underscore the effectiveness of ensemble methods, particularly the hybrid stacked model approach, in enhancing predictive accuracy and robustness in lung cancer classification tasks.

	KNN	SVM	DT	MLP	Raza et. al. []	proposed model
Accuracy	0.959854015	0.830900243	0.874695864	0.829683698	0.99	0.99756691
MCC	0.931629617	0.714863489	0.786323479	0.708984478	0	0.995878769
F1	0.959102292	0.782914109	0.871563472	0.780198811	0	0.997557954

Figure 17:comparative analysis of train data

The analysis of the results demonstrates in figure 18, that the proposed model outperforms individual machine learning models. The proposed model achieved a high accuracy of 0.949, a substantial MCC of 0.909, and an impressive F1 score of 0.948, indicating its capability for accurate classification and strong correlation between predicted and actual values. Compared to the baseline models, including KNN, SVM, and MLP, the proposed model demonstrates superior performance across all metrics, highlighting the effectiveness of ensemble methods

and the hybrid stacked model approach in enhancing predictive accuracy and robustness in lung cancer classification tasks.

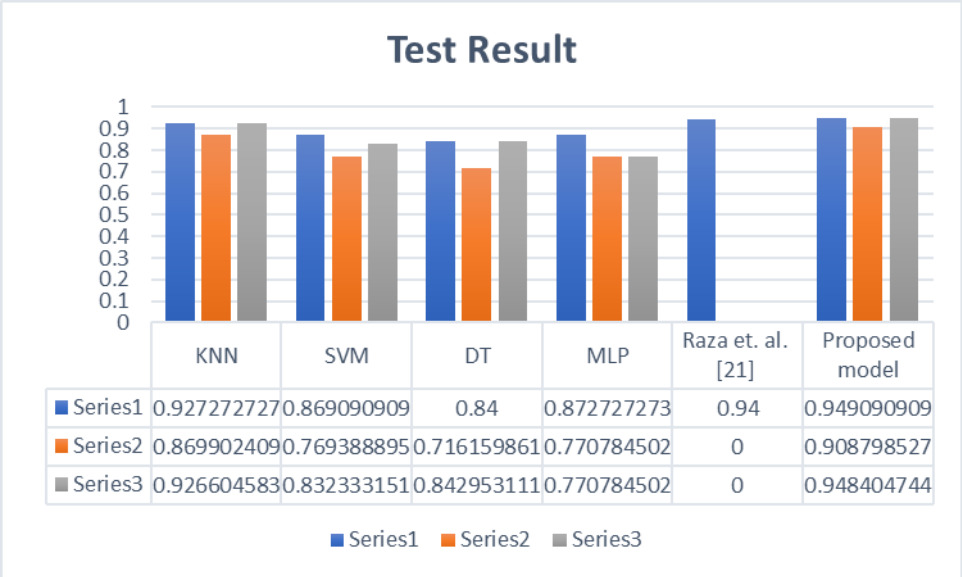


Figure 18:Comparative analysis of test data

Conclusion: The integration of contemporary data science techniques and artificial intelligence into the interpretation of medical imaging scans marks a significant shift towards more reliable and automated diagnostic tools. Unlike traditional methods reliant on subjective visual examination by skilled radiologists, this advancement aligns with the core objective of radiomics, merging personalized medicine with medical imaging. One of the main causes of cancer-related fatalities worldwide, lung cancer, has a difficult environment. However, advancements in lung cancer screening provide hope for improved treatment outcomes. The survivability rates at different stages of lung cancer highlight the criticality of early detection and intervention. Low dose computed tomography (LDCT), the recommended screening test for lung cancer, utilizes X-ray technology to produce detailed images of the lungs, aiding in the detection of abnormal tissue indicative of cancer. While LDCT offers life-saving potential, it also poses risks such as false positives, overdiagnosis, and exposure to X-rays. Therefore, screening is recommended primarily for individuals at high risk of lung cancer. New Generation Information Technologies (New IT) play a pivotal role in driving the evolution of Lung Cancer Classification. Leveraging LDCT in conjunction with AI-based Multilevel Optimization through a hybrid stacked model addresses the aforementioned challenges effectively. The proposed hierarchical reference architecture facilitates the extraction of pertinent features from LDCT scans using GLCM techniques, further refined through IGWO analysis to identify optimal features. This optimized solution, when integrated into a hybrid stacked model, enables accurate classification of input images as normal or abnormal, thereby enhancing diagnostic precision. The comparative analysis of different machine learning models, including KNN, SVM, Decision Tree, MLP, and the proposed stacked model, demonstrates the superiority of ensemble methods in improving predictive accuracy and robustness. The stacked model, with its exceptional performance metrics including accuracy, MCC, and F1 score, showcases the efficacy of leveraging multiple base models and a meta-model for enhanced classification capabilities. The ROC curves further illustrate the strong performance across all metrics, highlighting the stacked model's effectiveness in achieving accurate and reliable predictions. In conclusion, the integration of advanced data science techniques, AI, and ensemble learning methods holds immense potential in revolutionizing lung cancer classification and diagnosis.

CONFLICT OF INTEREST

The authors declare that there are no financial interests, commercial affiliations, or other potential conflicts of interest that could have influenced the objectivity of this research or the writing of this paper.

ACKNOWLEDGMENTS

I would like to express my heartfelt gratitude to my esteemed guide, Dr. Kuldeep Singh Kaswan, Professor at Galgotias University, India, for his invaluable guidance, unwavering support, and insightful feedback throughout this research journey. His expertise, encouragement, and mentorship have been pivotal in shaping this work and have significantly contributed to its successful completion. I am deeply grateful for his constant motivation, patience, and dedication, which have been a source of inspiration throughout my doctoral studies. This work would not have been possible without his profound knowledge and guidance.

REFERENCES:

[1] Xiao, H., Liu, Q., & Li, L. (2023). MFMANet: Multi-feature Multi-attention Network for efficient subtype classification on non-small cell lung cancer CT images. Biomedical Signal Processing and Control, 84, 104768.

[2] Subashchandrabose, U., John, R., Anbazhagu, U. V., Venkatesan, V. K., & Thyluru Ramakrishna, M. (2023). Ensemble Federated learning approach for diagnostics of multi-order lung cancer. Diagnostics, 13(19), 3053.

[3] Mothkur, R., & Veerappa, B. N. (2023). Classification of lung cancer using lightweight deep neural networks. Procedia Computer Science, 218, 1869-1877.

[4] Sun, R., Pang, Y., & Li, W. (2023). Efficient lung cancer image classification and segmentation algorithm based on an improved swin transformer. Electronics, 12(4), 1024.

[5] Barbouchi, K., El Hamdi, D., Elouedi, I., Aïcha, T. B., Echi, A. K., & Slim, I. (2023). A transformer-based deep neural network for detection and classification of lung cancer via PET/CT images. International Journal of Imaging Systems and Technology, 33(4), 1383-1395.

[6] Demiroğlu, U., Şenol, B., Yildirim, M., & Eroğlu, Y. (2023). Classification of computerized tomography images to diagnose non-small cell lung cancer using a hybrid model. Multimedia Tools and Applications, 82(21), 33379-33400.

[7] Gugulothu, V. K., & Balaji, S. (2023). An automatic classification of pulmonary nodules for lung cancer diagnosis using novel LLXcepNN classifier. Journal of Cancer Research and Clinical Oncology, 149(9), 6049-6057.

[8] Gugulothu, V. K., & Balaji, S. (2023). An automatic classification of pulmonary nodules for lung cancer diagnosis using novel LLXcepNN classifier. Journal of Cancer Research and Clinical Oncology, 149(9), 6049-6057.

[9] Braveen, M., Nachiyappan, S., Seetha, R., Anusha, K., Ahilan, A., Prasanth, A., & Jeyam, A. (2023). ALBAE feature extraction based lung pneumonia and cancer classification. Soft Computing, 1-14.

[10] Tyagi, S., & Talbar, S. N. (2023). LCSCNet: A multi-level approach for lung cancer stage classification using 3D dense convolutional neural networks with concurrent squeeze-and-excitation module. Biomedical Signal Processing and Control, 80, 104391.

[11] Kuruvilla, J., & Gunavathi, K. (2014). Lung cancer classification using neural networks for CT images. Computer methods and programs in biomedicine, 113(1), 202-209.

[12] Thakur, S. K., Singh, D. P., & Choudhary, J. (2020). Lung cancer identification: a review on detection and classification. Cancer and Metastasis Reviews, 39(3), 989-998.

[13] Inamura, K. (2017). Lung cancer: understanding its molecular pathology and the 2015 WHO classification. Frontiers in oncology, 7, 193.

[14] Teramoto, A., Tsukamoto, T., Kiriyama, Y., & Fujita, H. (2017). Automated classification of lung cancer types from cytological images using deep convolutional neural networks. BioMed research international, 2017.

[15] West, L., Vidwans, S. J., Campbell, N. P., Shrager, J., Simon, G. R., Bueno, R., ... & Salgia, R. (2012). A novel classification of lung cancer into molecular subtypes. PloS one, 7(2), e31906.

[16] Lakshmanaprabu, S. K., Mohanty, S. N., Shankar, K., Arunkumar, N., & Ramirez, G. (2019). Optimal deep learning model for classification of lung cancer on CT images. Future Generation Computer Systems, 92, 374-382.

[17] Gilad, S., Lithwick-Yanai, G., Barshack, I., Benjamin, S., Krivitsky, I., Edmonston, T. B., ... & Aharonov, R. (2012). Classification of the four main types of lung cancer using a microRNA-based diagnostic assay. The Journal of molecular diagnostics, 14(5), 510-517.

[18] Lynch, C. M., Abdollahi, B., Fuqua, J. D., De Carlo, A. R., Bartholomai, J. A., Balgemann, R. N., ... & Frieboes, H. B. (2017). Prediction of lung cancer patient survival via supervised machine learning classification techniques. International journal of medical informatics, 108, 1-8.

[19] Rami-Porta, R., Bolejack, V., Giroux, D. J., Chansky, K., Crowley, J., Asamura, H., & Goldstraw, P. (2014). The IASLC lung cancer staging project: the new database to inform the eighth edition of the TNM classification of lung cancer. Journal of Thoracic Oncology, 9(11), 1618-1624.

[20] Cai, Z., Xu, D., Zhang, Q., Zhang, J., Ngai, S. M., & Shao, J. (2015). Classification of lung cancer using ensemble-based feature selection and machine learning methods. Molecular BioSystems, 11(3), 791-800.

[21] Mohammed, S. H., & Çinar, A. (2021). Lung cancer classification with convolutional neural network architectures. Qubahan Academic Journal, 1(1), 33-39.

[22] Parmigiani, G., Garrett-Mayer, E. S., Anbazhagan, R., & Gabrielson, E. (2004). A cross-study comparison of gene expression studies for the molecular classification of lung cancer. Clinical cancer research, 10(9), 2922-2927.

[23] Kriegsmann, M., Haag, C., Weis, C. A., Steinbuss, G., Warth, A., Zgorzelski, C., ... & Kriegsmann, K. (2020). Deep learning for the classification of small-cell and non-small-cell lung cancer. Cancers, 12(6), 1604.

[24] Nageswaran, S., Arunkumar, G., Bisht, A. K., Mewada, S., Kumar, J. N. V. R., Jawarneh, M., & Asenso, E. (2022). Lung cancer classification and prediction using machine learning and image processing. BioMed Research International, 2022.

[25] Raza, R., Zulfiqar, F., Khan, M. O., Arif, M., Alvi, A., Iftikhar, M. A., & Alam, T. (2023). Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images. Engineering Applications of Artificial Intelligence, 126, 106902.