International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

Adaptive Handcrafted Features Convolutional Neural Network for Lung Cancer Detection Using CT Images

Kapila Moon
Computer Science
Pacific university- Research Scholar, Udaipur,India
RAIT-Faculty,Navi Mumbai
e-mail: kapila.moon@gmail.com

Dr.Ashok Jetawat
Computer science
Pacific university-Professor
Udaipur,India
e-mail: drashokjetawat@gmail.com

Abstract-In this study, an Adaptive Handcrafted Features Convolutional Neural Network (AHFCNN) for lung cancer detection from CT images is developed. The database is first built using data from web resources. After that, a pre-processing step is created to purge the photos of undesirable information. The pre-processing procedure is also taken into account while enhancing the photos. Because they work on various images with excellence and simplicity, the pixel intensity assessment and histogram techniques are used to improve the image quality. The Grey level cooccurrence matrix (GLCM) and local binary pattern (LBP) are then used to extract the necessary features. Finally, the CT image of the lung cancer is classified using the retrieved features. Convolutional neural networks (CNN) and hybrid meta-heuristic approaches (HMHA) are combined in the suggested classifier. The HMHA was applied to the CNN to select the best gain values. The Coati Optimisation Algorithm (COA) and Honey Badger Optimisation (HBO) are combined to create the HMHA. The HBO was used in the COA to improve the coatis' updating procedure. The proposed methodology was put into practise in Python, and its effectiveness was assessed by taking into account performance indicators including sensitivity, specificity, recall, sensitivity, and F-Score. Recurrent Neural Network- Whale Optimisation Algorithm (RNN-WOA), Deep Belief Neural Network- Remora Optimisation Algorithm (DBNN-ROA), and CNN- Grey Wolf Optimisation (CNN-GWO) are traditional methodologies that are compared to the proposed methodology.
.

Keywords- handcrafted features, convolutional neural network, CT image, Local binary pattern, gray level cooccurrence matrix, and coati optimization algorithm**.**

1.INTRODUCTION

The World Health Organization reports that cancer is the second largest cause of mortality worldwide. In terms of overall mortality, lung cancer ranks second, with 1.8 million annual fatalities. Early lung cancer detection (LCD) is crucial for medical professionals to undertake prognostic estimates and customize the treatment plan. Given the shortcomings of medical technology and the overwhelming workload, LCD utilizing artificial intelligence is receiving more and more attention in academia and practice. Reducing the time spent on medical diagnosis gives clinicians more time to focus on expert surgery and consultation, improving the quality of healthcare [1]. In contrast to a developing method that uses breath analysis via an electronic nose, we consider the conventional method of lung cancer screening using biomedical imaging in this work [2].

Cancer is seen as a major issue that may significantly increase mortality in both women and men due to confusing clinical testing and non-invasive treatments [3]. Patients with lung cancer have a comparatively low survival rate when compared to those with other cancers. In the early stages of lung cancer, it can be difficult to find the nodule regions that are present in the soft lung tissues [4,5]. Lung cancer can be identified using computerised tomography (CT) and chest radiographs (CXR) to detect pulmonary nodules [6]. Additionally, cutting-edge technology is employed to scan the entire chest region in a single breath hold and guarantees minimal noise images. When oncologists decide how to treat cancer, radiologists frequently find lung nodules. The size of the nodule is modest in the advance state of the disease, so even with the help of skilled radiologists, the doctor must take extra time to look for lung cancer. Radiologists give a variety of viewpoints when modest

morphological distinctions between benign and malignant nodules are present [7].

Recent advancements in radiography have been made possible by the use of convolutional neural networks (CNN), a subset of deep learning (DL). Deep learning-based [8] models have also shown potential in the diagnosis of chest radio graph nodules and masses, with reported sensitivities in the range of 0.51 to 0.84 and a mean number of false positive indications per image (mFPI) of 0.02-0.34. The ability of radiologists to spot nodules was also enhanced by these CAD models as compared to working without them[9]. Radiologists may have difficulty locating nodules and distinguishing benign from malignant nodules in clinical practise. Because nodules frequently resemble normal anatomical features, radiologists must pay close attention to the shape and marginal characteristics of nodules. Even skilled radiologists may incorrectly identify a patient since these issues are brought on by the conditions rather than the radiologists' abilities [10].

The remaining portions of this study are structured as follows. To highlight the significance of applying this subject in medical science, Section 2 will examine what has been achieved so far in terms of using machine learning algorithms to identify disorders. Three techniques and all of their phases are fully described in Section 3. The methods are used step-by-step to demonstrate their outcomes in Section 4. In Section 5, the advantages and disadvantages of each approach are discussed, along with a method comparison and method sensitivity analysis. Following that, the optimal strategy will be discussed, and Section 6 will feature upcoming projects.

2. Related Works

The use of deep learning techniques has improved the ability to identify lung cancer at its most advanced stage. By diagnosing the illness early on and taking the appropriate precautions, these strategies are utilized to control disease outbreaks. However, because these methods are applied singly, gaining reliability and more uniform results is more difficult. Some recent works related to lung cancer prediction are reviewed.

The diagnosis of lung cancer has been described by Kanchan Sitaram Pradhan et al., [11], using an inventive, clever method. Downloading two benchmark datasets that include attribute information from several patients' health records was the first step in the data collection procedure. Two widely used techniques, "Principal Component Analysis (PCA) and t-distributed Stochastic Neighbour Embedding (t-SNE)", have been used to extract features. The deep features were also recovered using "the pooling layer of Convolutional Neural Network (CNN)". The key features were subsequently identified using the Best Fitness-based Squirrel Search Algorithm (BF-SSA), also known as optimal feature selection. For successfully navigating the search space and enhancing feature selection performance, this hybrid optimization technique was recognized as being superior in several sectors.

In order to detect lung cancer on chest radio graphs, Akitoshi Shimazaki et al. [12] suggested a deep learning (DL)-based method using the segmentation approach. From January 2006 to June 2018, our hospital separately collected chest radio graphs for a training dataset and a test dataset. The DL-based model was trained and validated using fourfold cross-validation on the training dataset. Using the independent test dataset, the model's sensitivity and mean false positive indicators per picture (mFPI) were evaluated. The test dataset consists of 151 radiographs and 159 nodules/masses as opposed to 629 radio graphs and 652 nodules/masses in the training dataset.

A framework for ensemble learning has been presented by Sara A. Althubiti *et al.*, [13] for the early diagnosis of lung cancer. Choose between median, Gaussian, 2D convolution, and mean as the optimal filter to use on medical CT images during the preprocessing stage. It was then determined that the median filter was the best choice. Next, apply adaptive histogram equalization to increase image contrast. The improved pre-processed image was next submitted to fuzzy c-means and k-means clustering, two optimization techniques.

Machine learning was suggested by Sharmila Nageswaran et al. [14] to predict lung cancer. This study demonstrates how machine learning and image processing technology can be used to effectively classify and forecast lung cancer. Getting pictures was the first step. The dataset for the experimental investigation included 83 CT images from 70 different patients. The geometric mean filter was applied before the images were processed. As a result, the picture quality increased. The images were then divided up using the -means method. With the help of this segmentation, the area of the image can be located. Then, machine learning-based categorization algorithms were applied. ANN, KNN, and RF were among the machine-learning methods employed for the classification.

To aid in the early identification of lung cancer, Negar Maleki et al. [15] have provided Computerised Tomography (CT) scans of patients. In the first method, artificial neural networks (ANN) and convolutional neural networks (CNN) were used to analyse and classify the images. The second method involved pre-processing and segmenting the images before using them with CNN and ANN. All the pre-processed and segmented images have to be converted into numerical data using feature extraction as the final step in the third technique.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
***Article Received:** 25 November 2023 **Revised:** 12 December 2023 **Accepted:** 30 January 2024*

3. Proposed Architecture

The use of machine learning algorithms in medicine has expanded quickly, according to published studies. Deep learning algorithms are less flexible than machine learning algorithms when it comes to changing their characteristics. Deep learning algorithms, on the other hand, can swiftly assess and learn from raw data since they can automatically alter the features and collect them. Although deep learning algorithms are strong, they require a sufficient amount of data to demonstrate their strength. This research develops a deep learning-based method for predicting lung cancer. Figure 1 depicts the architecture.

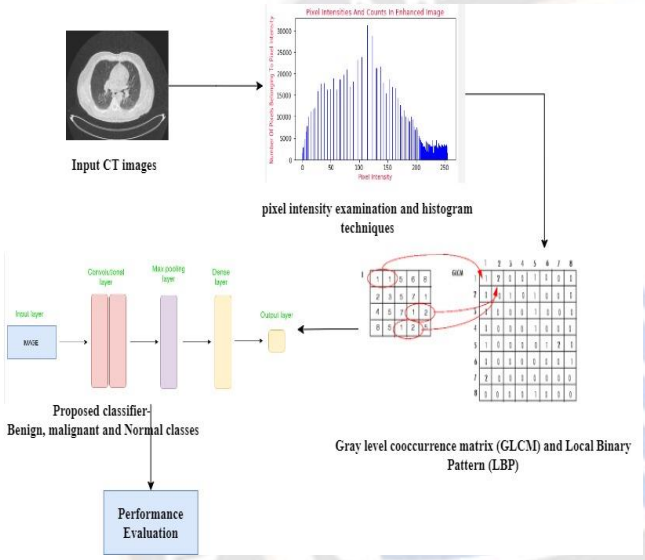


Figure 1. Proposed Architecture

3.1. Pre-processing

In lung cancer detection, pre-processing is a required part of empowering classification accuracy. The acquired photos are then put through steps to identify pixel noise and contrast information to improve the image quality. The obtained photos have a variety of inconsistent data and low-quality pixels, which reduce the prediction's accuracy for lung cancer. The histogram approach and pixel intensity analysis are taken into consideration to improve the quality of the supplied CT image.

3.1.1. Histogram Technique and Pixel Intensity Examination

The use of image histogram techniques is thought to improve image quality since they produce a variety of excellent and simple images. In order to authenticate the images by dividing them into two divisions, the created approach [16] uses the histogram methodology in the acquired CT lung image. The photos are also examined using this strategy in terms of quantity and quality. Each pixel is measured against the restricted parameter to improve the quality of the image, and the noise contained within the pixel is divided by the middle parameter. This method is being used to delete the picture clamour part in the past. After that, measures of splitting the image histogram related to weighted mean function and controlling sub histogram for empowering picture contrast are used to validate the noise-removed CT lung image. The input CT image contains G discrete gray level pixels defined as $p_0, p_1, \dots, p_{g-1}$. Once initialized, the pixel value of the image is computed based on their pixel cumulative distribution, and pixel density is calculated. For controlling how the pixels are connected [17] to a certain range of pixels, the pixel density is used. The pixel quality is estimated in relation to the density parameter, and then the density function is computed using,

$$PDF(p_n) = \frac{n^h}{n}; h = 0,1,2,3 \dots g-1 \qquad (1)$$

Here, n is defined as the complete pixels present in the image and n^h is a pixel parameter related to the parameter h. The weighted mean parameter of each pixel is calculated using the following formula after the pixel distribution parameter calculation.

$$x_t = \frac{\sum_{L=a}^{b} L * CDF}{\sum_{L=a}^{b} CDF\,(L)} \qquad (2)$$

where, a, b is defined as the sub-interval parameter of a histogram that is generated up to (0,255). This formulation improves the image quality while lowering noise. Sending the preprocessed picture to feature extraction.

3.2. Feature Extraction

In this approach, from the pre-processed image, the required features are extracted. The feature contains information that should to variation between classes. These classes may be sensitive to nonrelated changes in the pre-processed image and it is limited in count, to manage, effectual calculation of differential functions, and to limit the quantity of required training information. This approach computes the parameter pairs that manage the shape of a character uniquely and precisely [18]. In this stage, each image is defined by a feature vector and its identity. Its target is to avoid the original information by managing affirmative characteristics or features that vary one input image from another image.

3.2.1. Gray Level Cooccurrence Matrix (GLCM)

The statistical method for obtaining texture features must use the GLCM methodology. By calculating the degree of roughness, granularity, and contrast of the interaction between

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

neighbouring pixels in the image, the mathematical calculation takes a grey degree distribution into consideration. The average characteristics of the blue, green, and red pixels are taken into consideration during the first stage of GLCM computation in order to save the grayscale and produce a decent grey approximation. The weighted total of the blue, green, and red channels is then used with the coefficient parameters according to the BR.6001 standard. The co-occurrence matrix is calculated next using a large number of grayscale pixel parameters. once the co-occurrence matrix has been computed. The co-occurrence matrix is multiplied by this duplicate of the matrix to create a symmetric matrix. The following step involves splitting each symmetrical GLCM by the total number of parameters to normalise the symmetrical GLCM [19]. Calculating the textural features from the generalised GLCM matrix is the final step. Each GLCM feature is used to extract the textural features. The following formula is used to calculate the GLCM feature:

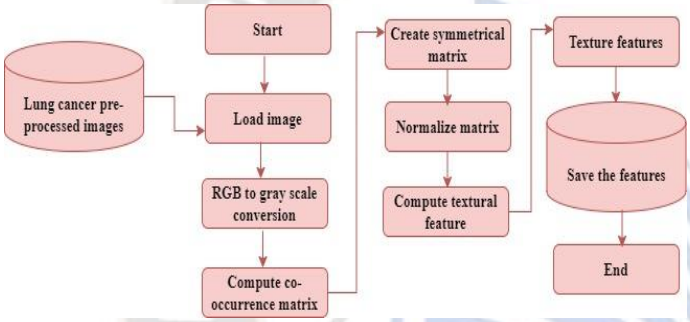


Figure 2. Flow chart of the GLCM features

$$sum\ of\ square\ variance = \sum_{i}\sum_{u}(i-\mu)^2 p(i,j) \quad (3)$$

$$Correlation = \frac{\sum_i \sum_j (ij)p(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \quad (4)$$

$$Energy = \sum_{i}\sum_{i} p(i-j)^2 \quad (5)$$

$$Angular\ second\ moment = \sum_{i}\sum_{j}\{p(i,j)\}^2 \quad (6)$$

$$Inverse\ different\ Moment = \sum_{i}\sum_{j}\frac{1}{1+(i-j)^2}.p(i,j) \quad (7)$$

$$Dissimilarity = \sum_{i}\sum_{j}|i-j|\,.p(i,j) \quad (8)$$

$$Contrast = \sum_{i,j=0}^{n_g-1} p(i,j)|i-j|^2 \quad (9)$$

Here, $\sigma_x\sigma_y$ is a statistical moment of the image that is second-order, $\mu_x\mu_y$ is referred to as the image's first-order statistical moments, (x,y) is defined as the pictorial information of the two-variable function, n_g is characterised as being the number of distinct grey levels in the quantized image, $p(i,j)$ the normalised grey time spatial dependency matrix and (i,j) is defined as the GLCM coordinates of every ranging.

3.2.2. Local Binary Pattern (LBP)

Normally, the LBP is utilized in one radius on eight directional matrix coordinates of the matrix parameter. In this paper, a kernel matrix is applied to compute the central pixel of the cell. Initially, in the first stage of the image [20], it divides into sub-cells. Based on this process, the first rotation kernel of the LBP is presented as follows,

$$g(x,y) = \sum_{i=-1}^{1}\sum_{j=-1}^{1} k_1(i,j) \times s_1(x-i, y-j) \quad (10)$$

Here, $g(x,y)$ is a central pixel parameter and it is a pixel matrix for the first kernel, k_1 is an eight-rotational kernel. Based on this process, the features are extracted.

3.3. Adaptive Handcrafted Features Convolutional Neural Network

To extract the necessary features for lung cancer detection and classification, the GLCM and LBP are used. The features are used in the proposed classifier to train the network. For CNN to work at its best, a hybrid meta-heuristic technique is devised. In the sections following, a thorough explanation of this suggested strategy is provided.

3.3.1. Convolutional Neural Network

The recommended method creates a weight map of pixel activity data from various source photos using CNN for lung cancer detection. This research uses a hybrid metaheuristic approach to increase the efficacy of CNN training. The CNN consists of three convolutional layers and one max pooling layer. The first two layers are the convolutional layer. The first layer consumes the features from the input image. In the layer after that [21], there are more feature maps. The features of the output map are extracted by CNN's convolutional layer. The term "max-pooling layer" refers to the subsequent layer. In order to reduce the number of variables, unnecessary samples

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

from the features are avoided. The output map from the pooling layer is used to extract more complicated information from the convolution layer using the fourth layer. It is regarded as a lightweight design to lessen the learning complexity and conserve memory. In addition, each branch's feature maps are first concatenated. The concatenated ones are then linked to a two-dimensional vector by adding a fully connected layer on top of them. The probability distribution of several parameters is used while diagnosing lung cancer.It is done by forward mapping it to a bi-directional SoftMax layer and categorizing it based on this probability value. The SoftMax classifier is constructed as follows to produce the CNN network's classification for lung cancer:

$$f(p_i) = \frac{e^{pi}}{\sum_{j=1}^{n} e^{pi}} \qquad (11)$$

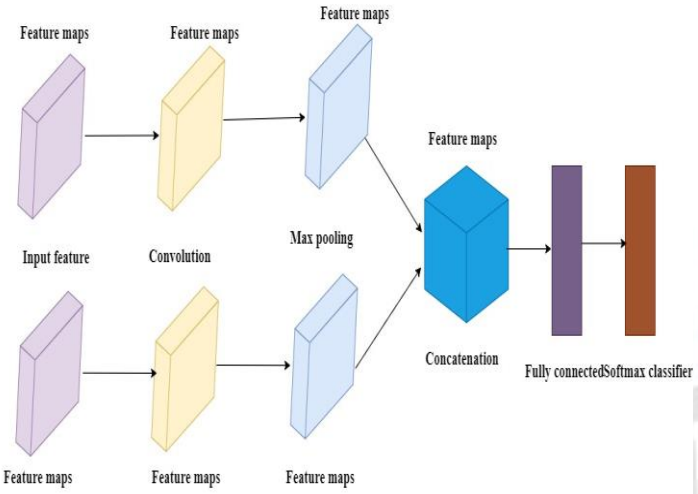


Figure 3. CNN Design Architecture

Complete input vectors are normalised if one p_i is greater than the remaining p, in which case its connected parameter is close to 1 and the others are close to 0.The batch size is paired to 128, therefore the equation below yields the SoftMax loss function.

$$l = \sum_{i=0}^{batchsize} -logf(p_i) \qquad (12)$$

The loss function is reduced using a hybrid meta-heuristic technique with the SoftMax loss function as the optimisation aim. The weight decay and momentum are initially set as pairs of 0.9 and 0.0005, respectively. Hence, the weight updating process is formulated as follows,

$$v_{i+1} = 0.9.v_i - 0.0005.\alpha.\omega_i - \alpha.\frac{\partial l}{\partial w_i} \qquad (13)$$

$$w_{i+1} = w_i + v_{i+1} \qquad (14)$$

Here, $\frac{\partial l}{\partial w_i}$ is defined as the loss derivative of weight, v_i is characterized as the dynamic variable and l is characterized as the loss function.

3.3.2. Hybrid meta-heuristic approach

In the suggested method, the CNN's ideal hyperparameters are chosen using a hybrid meta-heuristic method. The hybrid meta-heuristic method combines HBA and COA. Utilizing HBA improves the updating process in the COA algorithm. A detailed explanation of the COA and HBA is presented in the below section.

3.3.2.1. Coati Optimization Algorithm

The coati's behaviour when chasing after iguanas and their traits when fleeing from and fending against predators were used to construct this algorithm. These natural coati [22]' characteristics are a general inspiration for developing this algorithm.

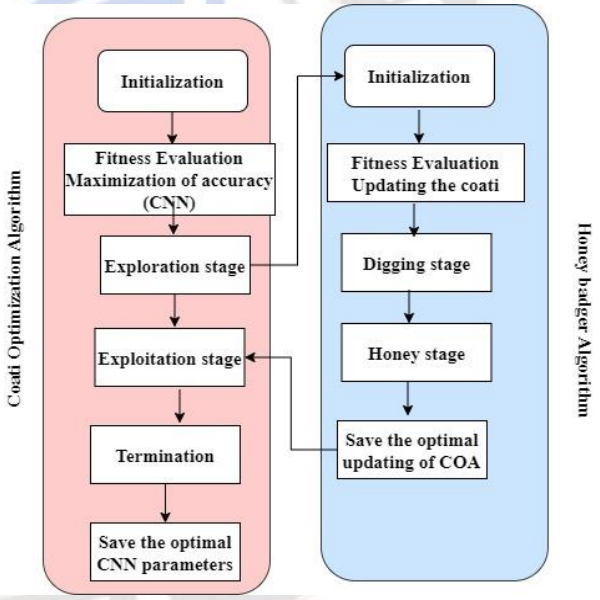


Figure 4. flowchart of the hybrid metaheuristic approach

The step-by-step process is presented as follows,

Step 1: Initialization process

The coatis are defined as population members of the COA approach, which is a population algorithm[24]. The parameters for the decision parameter are computed for each coati point in the search space. The coati's stance defined a potential resolution to the problem as a result. The following equation is used to initialise the coati position of the search space in the initial COA implementation,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
***Article Received:** 25 November 2023 **Revised:** 12 December 2023 **Accepted:** 30 January 2024*

$$x_i: x_{i,j} = LB_j + R.\left(UB_j - LB_j\right), i = 1,2,..,n, \quad j = 1,2,\dots,m \qquad (15)$$

Here, $UB_j and\ LB_j$ is defined as the upper bound and lower bound of the decision variable, R is defined as the random real number, m is described as the quantity of decision factors. [23], n is characterized as the definite quantity of coatis, $x_{i,j}$ is settled as the invariable of the decision variable and x_i is characterized as the coati attitude in the search space. The coati's population is mathematically defined as the following matrix,

$$x = \begin{bmatrix} x_1 \\ \dots \\ x_i \\ \dots \\ x_n \end{bmatrix}_{n\times m} = \begin{bmatrix} x_{1,1} & \dots & x_{1,j} & \dots & x_{1,m} \\ \dots & \dots. & \dots & \dots & \dots. \\ x_{i,1} & \dots & x_{i,j} & \dots. & x_{i,m} \\ \dots. & \dots. & \dots & \dots. & \dots \\ x_{n,1} & \dots. & x_{n,j} & \dots. & x_{n,m} \end{bmatrix}_{n\times m} \qquad (16)$$

The positioning of potential solutions in the decision parameters results in the validation of numerous parameters of the issue's objective function.

Step 2: Fitness Evaluation

In this algorithm, CNN hyperparameters are selected. Related to this hyperparameter selection, the fitness function is formulated as follows,

$$FF = Max(Accuracy) \qquad (17)$$

Here, the FF is created to obtain the optimal classification measure by creating its accuracy.

Step 3: Exploration stage

In this stage, the coati position is updated in the search space related to validating their approach when attacking iguanas. The iguana posture is thought to represent the position of the ideal population member in this design. The following is how this coatis position update is presented:

$$x_i^{p1}: x_{ij}^{p1} = x_{ij} + R\left(Iguana_j - i.x_{i,j}\right), fot\ i = 1,2,\dots,\left\lfloor\frac{n}{2}\right\rfloor\ and\ j = 1,2,\dots,m \qquad (18)$$

The iguana is discovered at a random location inside the search area once it has fallen to the ground. It is formulated as follows in relation to this arbitrary position, coatis on the ground move in the search space:

$$Iguana^g: Iguana_j^g = LB_j + R.\left(UB_j - LB_j\right), j = 1,2,\dots,m \qquad (19)$$

$$x_i^{p1}: x_{ij}^{p1} = \begin{cases} x_{ij} + R\left(Iguana_j - i.x_{i,j}\right) & f_{iguana}^g < f_i \\ x_{ij} + R\left(x_{i,j} - Iguana_j^g\right), & Else \end{cases} \quad for\ i = \left\lfloor\frac{n}{2}\right\rfloor + 1, \left\lfloor\frac{n}{2}\right\rfloor + 2, \dots, n\ and\ j = 1,2,\dots m \qquad (20)$$

Every coati's new position is calculated, and if it improves the objective function's parameter, the updating method is controllable. The coati, though, makes a save at the final position. The formulation of this update procedure is as follows:

$$x_i = \begin{cases} x_i^{p1} & f_i^{p1} < f_i \\ x_i & Else \end{cases} \qquad (21)$$

Where, $\lfloor . \rfloor$ is a floor function, f_{iguana}^g is a value of an objective function, $Iguana_j^g$ is the position of the iguana on the ground, i is characterised as the integer that is picked at random from the pair {1,2},

$Iguana_j$ is the search space's iguana position, which is typically seen of as the best member position, R is described as the interval real number generated at random, f_i^{p1} considered to be the objective function parameter, x_{ij}^{p1} is characterized as the dimension, x_i^{p1} is defined as the coati's new position as computed.

Step 4: Exploitation stage

The process for improving the position of coatis in the search space is created at this stage based on the typical coatis characteristics. A coati is attacked by a predator, but the animal manages to get away. In order to model this behaviour, a random position is constructed nearby each coati's location in relation to the formulation below.

$$LB_j^{local} = \frac{LB_j}{T}, UB_j^{local} = \frac{UB_j}{T}, \quad here\ t = 1,2,\dots,T \qquad (22)$$

$$x_i^{p2}: x_{ij}^{p2} = x_{ij} + (1-2R)\left(LB_j^{local} + R.\left(UB_j^{local} - LB_j^{local}\right)\right) for\ i = 1,2,\dots,n\ and\ j = 1,2,\dots,m \qquad (23)$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
***Article Received:** 25 November 2023 **Revised:** 12 December 2023 **Accepted:** 30 January 2024*

If the goal function's parameter is improved by the newly computed position, which is the condition that is written as follows,

$$x_i = \begin{cases} x_i^{p2} & f_i^{p2} < f_i \\ x_i & Else \end{cases} \quad (24)$$

Here, R is defined as the random number, f_i^{p2} is an objective function parameter, x_i^{p2} is a novel location computed for the coati related to the exploitation phase, t is defined as the iteration counter.

Step 5: Termination condition

It is the final step of the COA. Based on this process, the optimal hyperparameter of CNN is optimized. Once the iteration is completed, the optimal solution is saved and sent to the CNN for classification of lung cancer. In the COA, the updating process is enhanced by considering the HBA. The section below provides an explanation of this algorithm procedure.

3.3.2.2. Honey Badger Algorithm

The honey badgers' foraging traits are started by this algorithm. The honey badger digs or smells, then follows the honeyguide bird to find the food source. The section below presents the HBA's mathematical formulation.

Step 1: Initialization step

In this stage, the population of the candidate solution is initialized with random coatis updating. It is formulated as follows,

$$Candidate\ solution = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1d} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2d} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{nd} \end{bmatrix} \quad (25)$$

$$honey\ badger\ position\ x_i = \left[x_i^1, x_i^2, \dots, x_i^d\right] \quad (26)$$

$$x_i = LB_i + R_1 \times (UB_i - LB_i), R_1\ is\ a\ random\ number\ among\ 0\ and\ 1 \quad (27)$$

Here, x_i is defined as the honey badger position, UB_i and LB_i is defined as a upper and lower bound.

Step 2: Fitness Evaluation

In this evaluation function, the updating process of coati is obtained. The fitness function is formulated as follows,

$$FF = Min\ (Coati\ updating\ time) \quad (28)$$

It reduces the updating time and enhances the outcome of the CNN hyperparameter optimization.

Step 3: Describing Intensity

It is associated with the prey's concentration level and the separation between the prey and the honey badger. The following is how the defining intensity is expressed:

$$I_i = R_2 \times \frac{s}{4\pi d_i^2}, R_2\ is\ a\ random\ number\ among\ 0\ and\ 1 \quad (29)$$

$$s = (x_i - x_{i+1})^2 \quad (30)$$

$$d_i = x_{prey} - x_i \quad (31)$$

Here, d_i is outlined as the separation between the badger and its prey.and s is referred as the source and concentration strength.

Step 4: Upgrade density factor

To facilitate the shift from exploration to exploitation, it controls time-varying randomness. This decreasing factor which reduced [25] with iterations to decrease randomization with time. This density factor is presented as follows,

$$\alpha = c \times exp\left(\frac{-T}{T_{max}}\right), T_{max}\ is\ a\ maximum\ number\ of\ iterations \quad (32)$$

Here, c is defined as the constant.

Step 5**:** Local optima escaping condition and updating process

The three steps are used to get away from the regionally advantageous areas. This algorithm takes into account a flag that controls search direction for reserving chances for agents to search the search location during this operation. Phases of digging and honey are used to acquire the updated process of HBA.

$$x_{new} = x_{prey} + f \times \beta \times i \times x_{prey} + f \times r_3 \times \alpha \times d_i \times |cos(2\pi r_4) \times [1 - cos(2\pi r_5)]| \quad (33)$$

Here, f is defined as the flag and it manages search direction, r_3, r_4, r_5 is defined as the random numbers between 0 and 1, d_i is characterised as the separation between honey badgers and

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
***Article Received:** 25 November 2023 **Revised:** 12 December 2023 **Accepted:** 30 January 2024*

their prey and x_{prey} is defined as the prey position that is the optimal position.

$$f = \begin{cases} 1 & if\ r_6 \le 0.5 \\ -1 & else \end{cases} \quad r_6\ is\ defined\ as\ random\ number\ among\ 0\ and\ 1 \qquad (34)$$

The case when a honey badger manages a honeyguide bird to obtain a beehive is simulated by using the below equation,

$$x_{new} = x_{prey} + f \times R_7 \times \alpha \times d_i, R_7\ is\ a\ random\ number\ among\ 0\ and\ 1 \qquad (35)$$

Here, x_{prey} is characterized as the prey location, x_{new} is characterized as a new position of honey badger. The search in this stage is affected by time-varying search characteristics. A honey badger is calculating the disruption, though. Because of the exploitation and exploration phases, HBA theoretically has a connection to a global optimisation algorithm. Based on this process, the HBA is utilized to update the coatis in the population. Based on the hybrid meta-heuristic approach, the optimal CNN parameters are optimized. Additionally, it enhances the lung cancer classification.

4. Performance Evaluation

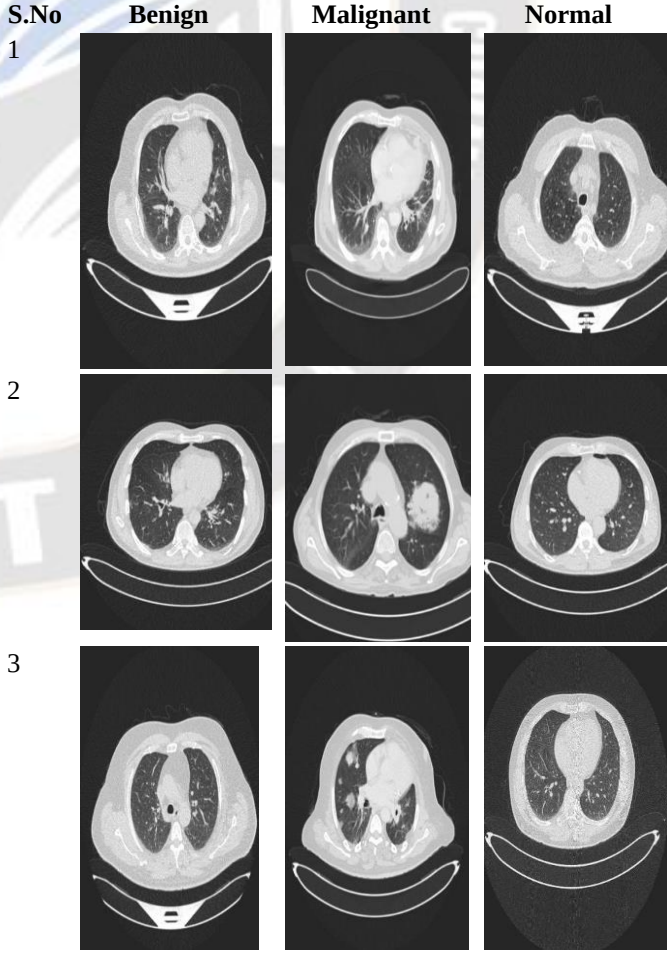
In this section, the validation of the suggested method is examined and presented. The suggested technique is designed to recognise lung cancer from the image database. The deep learning model aims to improve the precision of detection. For fine-tuning the deep learning model, a hybrid metaheuristic technique is also devised. Utilising common performance metrics like F1-score, recall, precision, mean square error, mean absolute error, and accuracy, the suggested deep learning model is assessed. The usefulness and efficiency of these measures in predicting lung cancer are evaluated. The performance of lung cancer detection is typically calculated using precision as a parameter. In Table 1, the simulation metrics are listed.

Table 1: Simulation Variables

S.No	Parameter	Value
1	Number of search agent	50
2	Momentum	0.9
3	Maximum epochs	15
4	Constant (c)	2
5	Learn rate drop period	5
6	optimizer	hybrid metaheuristic approach
7	Number of iterations	100
8	Learn rate drop factor	0.2
9	Beta	6
10	Mini batch size	500

Dataset description:

In the autumn of 2019, three months were spent in the aforementioned specialised institutions compiling the lung cancer data collection for the Iraq-Oncology Teaching Hospital/National Centre for Cancer Diseases (IQ-OTH/NCCD).It includes CT pictures of both healthy volunteers and lung cancer patients at various stages of the disease. Oncologists and radiologists from these two centres commented IQ-OTH/NCCD slides.The dataset is made up of 1190 images total, which were extracted from slices of CT scans from 110 instances. These cases are broken down into three groups: benign, malignant and normal. One hundred and fifty-five of them have been classified as normal, forty as malignant, and fifteen as benign. It was done in DICOM format for the initial collection of CT images. Figure 5 shows three sample photos with classes.



International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

4

5

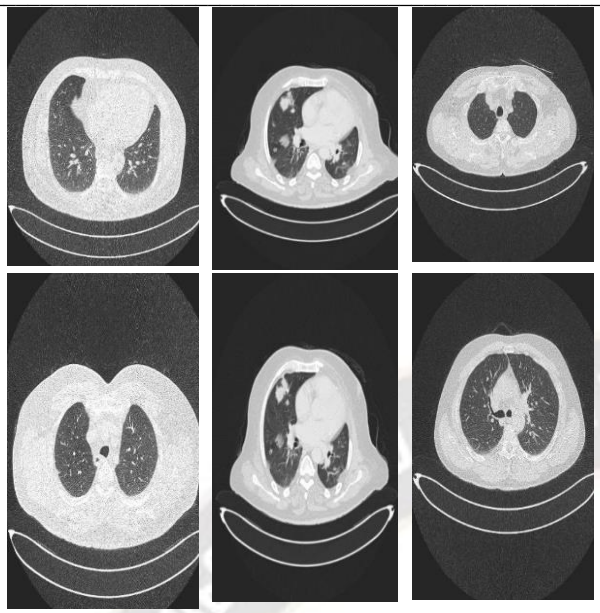


Figure 5. Sample Images

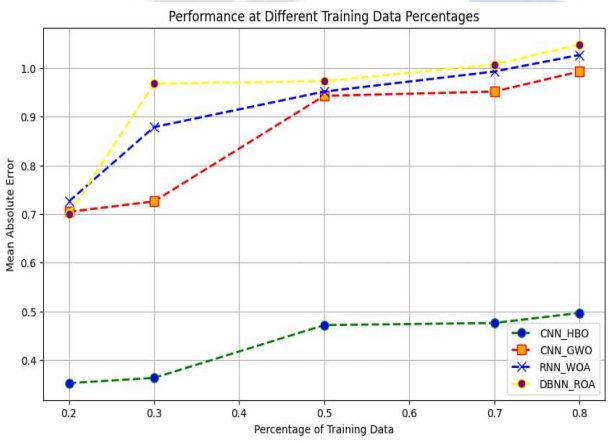


Figure 6. Mean square error

By taking mean square error into account, the proposed methodology for lung cancer prediction is validated. The suggested method is contrasted with the well-known CNN-GWO, RNN-WOA, and DBNN-ROA methods. Based on their training data, CNN-HBO's suggested method produces minimum and maximum error levels of 0.3 to 0.5. The minimum and maximum error values for the standard CNN-GWO, RNN-WOA, and DBNN-ROA approaches are 0.7-1.0, 0.72-1.2, and 0.72-1.3, respectively. The proposed technique has low MAE during the prediction of lung cancer, according to comparison validation. We can therefore draw the conclusion that the suggested methodology produces the best results possible when measuring MAE. By taking the F1-score into account, the proposed methodology for lung cancer prediction is validated and it is given in Figure 6. The suggested method is contrasted with the well-known CNN-GWO, RNN-WOA, and DBNN-ROA methods. Based on their training data, CNN-HBO's suggested method produces a minimum and maximum F1-score of 0.58 to 0.95. The minimum and F1-score for the standard CNN-GWO, RNN-WOA, and DBNN-ROA approaches are 0.32-0.68, 0.38-0.56, and 0.28-0.6, respectively. The proposed technique has a high F1-score during the prediction of lung cancer, according to comparison validation. We can therefore draw the conclusion that the suggested methodology produces the best results possible when measuring the F1 score. By taking precision into account, the proposed methodology for lung cancer prediction is validated and it is given in Figure 7. The suggested method is contrasted with the well-known CNN-GWO, RNN-WOA, and DBNN-ROA methods. Based on their training data, CNN-HBO's suggested method produces minimum and maximum precision of 0.38 to 0.7. The minimum and precision for the standard CNN-GWO, RNN-WOA, and DBNN-ROA approaches are 0.2-0.58, 0.19-0.58, and 0.15-0.28, respectively. The proposed technique has high precision during the prediction of lung cancer, according to comparison validation. We can therefore draw the conclusion that the suggested methodology produces the best results possible when measuring precision.

Figure 7: F1-Score

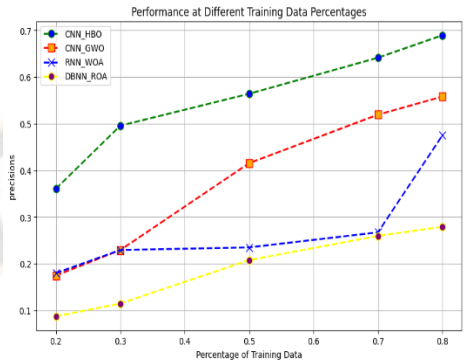


Figure 8. Precision

By taking specificity into account, the proposed methodology for lung cancer prediction is validated and it is given in Figure 9. The suggested method is contrasted with the well-known CNN-GWO, RNN-WOA, and DBNN-ROA methods. Based on their training data, CNN-HBO's suggested method produces

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

minimum and maximum specificity of 0.48 to 0.79. The minimum and maximum specificity for the standard CNN-GWO, RNN-WOA, and DBNN-ROA approaches are 0.48-0.68, 0.43-0.62, and 0.43-0.68, respectively. The proposed technique has high specificity during the prediction of lung cancer, according to comparison validation. We can therefore draw the conclusion that the suggested methodology produces the best results possible when measuring specificity. By taking mean square error into account, the proposed methodology for lung cancer prediction is validated and it is given in Figure 9. The suggested method is contrasted with the well-known CNN-GWO, RNN-WOA, and DBNN-ROA methods. Based on their training data, CNN-HBO's suggested method produces a minimum and maximum mean square error of 0.38 to 0.82. The minimum with maximum mean square error for the standard CNN-GWO, RNN-WOA, and DBNN-ROA approaches are 0.68-1.6, 0.76-1.7, and 0.81-1.98, respectively. The proposed technique has a low mean square error during the prediction of lung cancer, according to comparison validation. We can therefore draw the conclusion that the suggested methodology produces the best results possible when measuring mean square error.

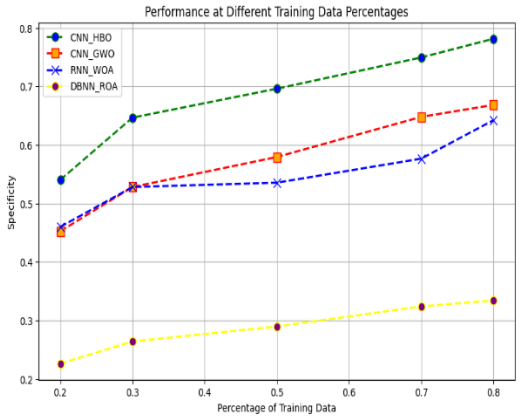


Figure 9. Specificity

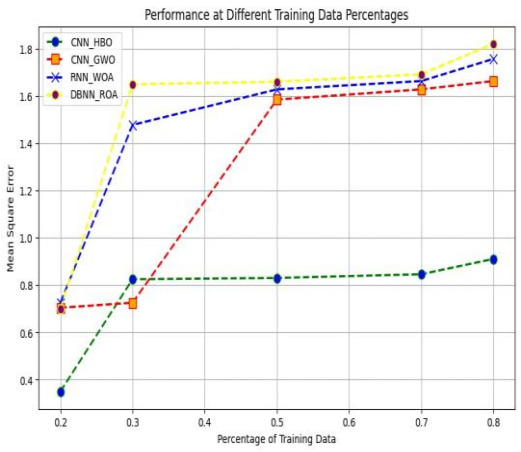


Figure 10.Mean Square error

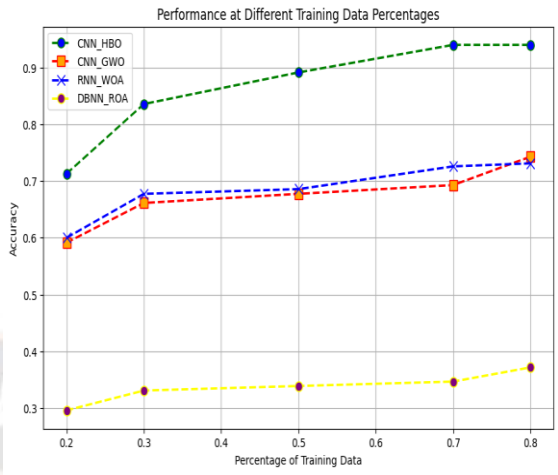


Figure 11.Accuracy

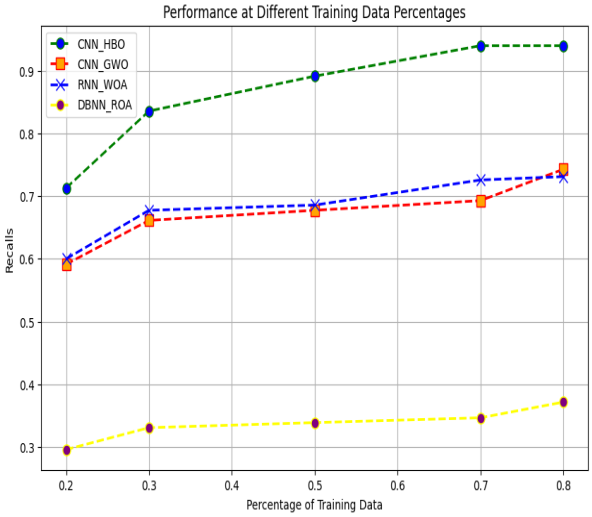


Figure 12. Recall

By taking accuracy into account, the proposed methodology for lung cancer prediction is validated and it is given in Figure 10. The suggested method is contrasted with the well-known CNN-GWO, RNN-WOA, and DBNN-ROA methods. Based on their training data, CNN-HBO's suggested method produces minimum and maximum accuracy of 0.72 to 0.92. The minimum and maximum accuracy for the standard CNN-GWO, RNN-WOA, and DBNN-ROA approaches are 0.62-0.72, 0.65-0.75, and 0.31-0.38, respectively. The proposed technique has high accuracy during the prediction of lung cancer, according to comparison validation. We can therefore draw the conclusion that the suggested methodology produces the best results possible when measuring accuracy. By taking recall into account, the proposed methodology for lung cancer prediction is validated and it is given in Figure 11. The suggested method is contrasted with the well-known CNN-GWO, RNN-WOA, and DBNN-ROA methods. Based on their training data, CNN-HBO's suggested method produces minimum and maximum

recall of 0.72 to 0.95. The minimum and maximum recall for the standard CNN-GWO, RNN-WOA, and DBNN-ROA approaches are 0.62-0.73, 0.65-0.76, and 0.31-0.39, respectively. The proposed technique has high recall during the prediction of lung cancer, according to comparison validation. We can therefore draw the conclusion that the suggested methodology produces the best results possible when measuring recall.

5. Conclusion

An AHFCNN for detecting lung cancer from the CT image has been created in this study. The database was initially compiled from online sources. The pre-processing procedure was devised to remove extraneous information from the photos after that. Additionally, the pre-processing procedure has been taken into account to improve the photos. Because they are effective and straightforward when applied to various images, the pixel intensity evaluation and histogram approaches have been used to enhance image quality. Following that, the GLCM and LBP were used to extract the necessary characteristics. The lung carcinoma from the CT scan is finally classified using the retrieved features. A CNN and an HMHA were combined to create the suggested classifier. The best gain settings for the CNN were selected using the HMHA. It combines the COA and HBO to form the HMHA. The HBO was used in the COA to speed up the coati upgrading process. Accuracy, precision, recall, F-Score, sensitivity, and specificity were performance indicators taken into account while evaluating the proposed methodology's effectiveness once it was implemented in Python. In comparison to the traditional methods RNN-WOA, DBNN-ROA, and CNN-GWO, the proposed methodology is evaluated. The evaluation of lung cancer prediction in the future will take into account real-time data.

References

[1]. Kadir, Timor, and Fergus Gleeson. "Lung cancer prediction using machine learning and advanced imaging techniques." Translational lung cancer research 7, no. 3 (2018): 304.

[2]. Li, Wen, Ji-Bin Liu, Li-Kun Hou, Fei Yu, Jie Zhang, Wei Wu, Xiao-Mei Tang et al. "Liquid biopsy in lung cancer: significance in diagnostics, prediction, and treatment monitoring." Molecular cancer 21, no. 1 (2022): 25.

[3]. Anil Kumar, C., S. Harish, Prabha Ravi, Murthy Svn, B. P. Kumar, V. Mohanavel, Nouf M. Alyami, S. Shanmuga Priya, and Amare Kebede Asfaw. "Lung cancer prediction from text datasets using machine learning." BioMed Research International 2022 (2022).

[4]. Dritsas, Elias, and Maria Trigka. "Lung cancer risk prediction with machine learning models." Big Data and Cognitive Computing 6, no. 4 (2022): 139.

[5]. Sattar, Mohsin, Abdul Majid, Nabeela Kausar, Muhammad Bilal, and Muhammad Kashif. "Lung cancer prediction using multi-gene genetic programming by selecting automatic features from amino acid sequences." Computational Biology and Chemistry 98 (2022): 107638.

[6]. Yang, Yang, Li Xu, Liangdong Sun, Peng Zhang, and Suzanne S. Farid. "Machine learning application in personalised lung cancer recurrence and survivability prediction." Computational and Structural Biotechnology Journal 20 (2022): 1811-1820.

[7]. Liu, Suli, and Wu Yao. "Prediction of lung cancer using gene expression and deep learning with KL divergence gene selection." BMC bioinformatics 23, no. 1 (2022): 175.

[8]. Guo, Lan-Wei, Zhang-Yan Lyu, Qing-Cheng Meng, Li-Yang Zheng, Qiong Chen, Yin Liu, Hui-Fang Xu et al. "A risk prediction model for selecting high-risk population for computed tomography lung cancer screening in China." Lung Cancer 163 (2022): 27-34.

[9]. Zhong, Yifan, Yunlang She, Jiajun Deng, Shouyu Chen, Tingting Wang, Minglei Yang, Minjie Ma et al. "Deep learning for prediction of N2 metastasis and survival for clinical stage I non–small cell lung cancer." Radiology 302, no. 1 (2022): 200-211.

[10]. Puttanawarut, Chanon, Nat Sirirutbunkajorn, Narisara Tawong, Chuleeporn Jiarpinitnun, Suphalak Khachonkham, Poompis Pattaranutaporn, and Yodchanan Wongsawat. "Radiomic and dosiomic features for the prediction of radiation pneumonitis across esophageal cancer and lung cancer." Frontiers in Oncology 12 (2022): 768152.

[11]. Wang, Yulong, Haoxin Zhang, and Guangwei Zhang. "cPSO-CNN: An efficient PSO-based algorithm for fine-tuning hyper-parameters of convolutional neural networks." Swarm and Evolutionary Computation 49 (2019): 114-123.

[12]. Shimazaki, Akitoshi, Daiju Ueda, Antoine Choppin, Akira Yamamoto, Takashi Honjo, Yuki Shimahara, and Yukio Miki. "Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method." Scientific Reports 12, no. 1 (2022): 727.

[13]. Althubiti, Sara A., Sanchita Paul, Rajanikanta Mohanty, Sachi Nandan Mohanty, Fayadh Alenezi, and Kemal Polat. "Ensemble learning framework with GLCM texture extraction for early detection of lung cancer on CT images." Computational and Mathematical Methods in Medicine 2022 (2022).

[14]. Nageswaran, Sharmila, G. Arunkumar, Anil Kumar Bisht, Shivlal Mewada, J. N. V. R. Kumar, Malik Jawarneh, and Evans Asenso. "Lung cancer classification and prediction using machine learning and image processing." BioMed Research International 2022 (2022).

[15]. Maleki, Negar, and Seyed Taghi Akhavan Niaki. "An intelligent algorithm for lung cancer diagnosis using extracted features from Computerized Tomography images." Healthcare Analytics 3 (2023): 100150.

[16]. Ramamoorthy, M., Shamimul Qamar, Ramachandran Manikandan, Noor Zaman Jhanjhi, Mehedi Masud, and Mohammed A. AlZain. "Earlier detection of brain tumor by pre-processing based on histogram equalization with neural network." In Healthcare, vol. 10, no. 7, p. 1218. MDPI, 2022.

[17]. Pour, Asra Momeni, Hadi Seyedarabi, Seyed Hassan Abbasi Jahromi, and Alireza Javadzadeh. "Automatic detection and monitoring of diabetic retinopathy using efficient convolutional neural networks and contrast limited adaptive histogram equalization." IEEE Access 8 (2020): 136668-136673.

[18]. Barburiceanu, Stefania, Romulus Terebes, and Serban Meza. "3D texture feature extraction and classification using GLCM and LBP-based descriptors." Applied Sciences 11, no. 5 (2021): 2332.

[19]. Andono, Pulung Nurtantio, and Eko Hari Rachmawanto. "Evaluasi Ekstraksi Fitur GLCM dan LBP Menggunakan Multikernel SVM untuk Klasifikasi Batik." Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi) 5, no. 1 (2021): 1-9.

[20]. Fauzi, Arthur Ahmad, Fitri Utaminingrum, and Fatwa Ramdani. "Road surface classification based on LBP and GLCM features using kNN classifier." Bulletin of Electrical Engineering and Informatics 9, no. 4 (2020): 1446-1453.

[21]. Han, Zhimeng, Muwei Jian, and Gai-Ge Wang. "ConvUNeXt: An efficient convolution neural network for medical image segmentation." Knowledge-Based Systems 253 (2022): 109512.

[22]. Dehghani, Mohammad, Zeinab Montazeri, Eva Trojovská, and Pavel Trojovský. "Coati Optimization Algorithm: A new bio-inspired metaheuristic algorithm for solving optimization problems." Knowledge-Based Systems 259 (2023): 110011.

[23]. Abou Houran, Mohamad, Syed M. Salman Bukhari, Muhammad Hamza Zafar, Majad Mansoor, and Wenjie Chen. "COA-CNN-LSTM: Coati optimization algorithm-based hybrid deep learning model for PV/wind power forecasting in smart grid applications." Applied Energy 349 (2023): 121638.

[24]. Hashim, Fatma A., Essam H. Houssein, Kashif Hussain, Mai S. Mabrouk, and Walid Al-Atabany. "Honey Badger Algorithm: New metaheuristic algorithm for solving optimization problems." Mathematics and Computers in Simulation 192 (2022): 84-110.

[25]. Khan, Muhammad Haris, Abasin Ulasyar, Abraiz Khattak, Haris Sheh Zad, Mohammad Alsharef, Ahmad Aziz Alahmadi, and Nasim Ullah. "Optimal Sizing and Allocation of Distributed Generation in the Radial Power Distribution System Using Honey Badger Algorithm." Energies 15, no. 16 (2022): 5891.