J. Electrical Systems 20-3 (2024): 1375-1396

[1] **Chintha Vishnu Vardhana Reddy**

[2]**Malikarjuna Reddy A**

Lrfe: A Novel Local Response Feature Elimination Process for Identification of Lung Cancer Cells

Journal of Electrical Systems

Abstract: - One of the main causes of cancer-related mortality across the globe is lung cancer. Early-stage lung cancer frequently exhibits no symptoms, which delays diagnosis until the illness has progressed. Before symptoms manifest, screening and early detection techniques can aid in the early diagnosis and treatment of lung cancer. Many olden research papers have implemented image processing and few latest papers have implemented computer vision techniques to detect lung cancer. Particularly when dealing with minor or subtle anomalies, image processing algorithms may not be able to detect lung cancer lesions with sufficient sensitivity and specificity. It is still difficult to increase the detection algorithms' accuracy and dependability, especially when dealing with early-stage lesions or situations where attributes overlap. It takes a lot of processing power, such as high-performance GPUs and enormous memory capacities, to train deep learning models, particularly large-scale convolutional neural networks (CNNs). In this proposed research, the model pre-processes the images using the ostu and sober filter mechanisms because Otsu's approach adjusts to the features of the input image, including noise, contrast, and lighting fluctuations. It is capable of handling images with varying dynamic ranges and intensity distributions without depending on pre-established threshold settings. When it comes to image noise, the Sobel filter is more resilient than other edge detection methods. It produces clearer edge maps and fewer false detections by determining the gradient magnitude, which amplifies edge information while suppressing noise. The features are extracted using the tuned AlexNet pre-trained model, in AlexNet there is a layer known as "Layer-wise Relevance Propagation". By giving each pixel or feature in the input image a relevance score, the LRP layer offers fine-grained feature attribution. This makes it possible to analyze in great depth which particular elements or areas of the input image are most important for the network to forecast, which helps to clarify the underlying patterns that the network has learned. At last, the extracted features are further reduced using the enhanced feature elimination method. By iteratively selecting subsets of features based on their importance, RFE helps to identify the most relevant features for the classification task. Integrating RFE with SVM classifiers can lead to improved model performance by focusing on the subset of features that are most discriminative and informative for the classification problem.

Keywords: Ostu Method, Sober Filter, Layer-wise Relevance, Fine Grained Features, Feature Elimination.

I. INTRODUCTION

The pixel values of an image might give it a high dimensionality. In order to make images easier to handle for further processing and analysis, feature extraction techniques minimize the dimensionality of the image while maintaining crucial visual information. In many cases, extracted characteristics are easier to read than raw data, which facilitates comprehension of the underlying relationships and patterns in the information. CNNs are a subset of deep neural network architecture that are frequently employed in computer vision tasks to extract features. Convolutional layers and pooling layers make up its structure; the purpose of the former is to learn the spatial hierarchies of features from images. Learnable filters are utilized by convolutional layers to extract local patterns, and information is aggregated by pooling layers to decrease spatial dimensions and improve translation invariance. Figure 1 presents the feature extraction using neural networks.

[1] Research Scholar, Department of CSE, Anurag University, Ch.vishnuvardhana@gmail.com,

[2] Head and Associate Professor Department of Artificial Intelligence, Anurag University, Mallikarjunreddycse@cvsr.ac.in

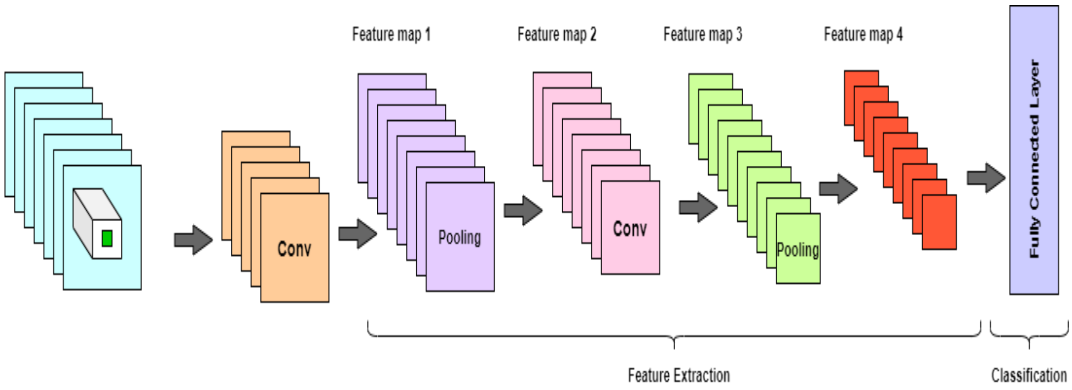

Figure 1: Feature Extraction using Neural Networks

When working with huge datasets and intricate network designs, training deep neural networks for feature extraction can be computationally costly. Some applications may find this to be impracticable due to the substantial computational resources and time required. To overcome this problem, in this paper it divided the process into two phases. The initial phase uses feature extraction using pre-trained model using “AlexNet”

(a) AlexNet as a feature extractor: One can evaluate the significance of input feature gradients in affecting the model's output by computing them. The predictions of the model are usually more affected by features with larger gradient magnitudes. This method is known as saliency mapping or gradient-based feature attribution. AlexNet has completely linked layers after pooling and convolutional layering. It learnt to extract hierarchy characteristics from raw photos, including edges, textures, & object sections and forms, via ImageNet training. The convolutional layers' learnt features effectively represent input pictures. Transfer learning applies training from one task to another similar task. AlexNet is useful for jobs with little information or computing resources since its pre-trained weights may be applied to different tasks with minimum alterations. By employing AlexNet as a extraction, one may use learned representations for object identification, segmentation, and picture retrieval. New deep neural network training demands a lot of processing power and labeled data. Because the AlexNet model has already learnt generic attributes from a big dataset like ImageNet, employing it as a feature extractor saves a lot of time and computing resources. In situations with little labeled data or quick prototyping, this efficiency is very useful.

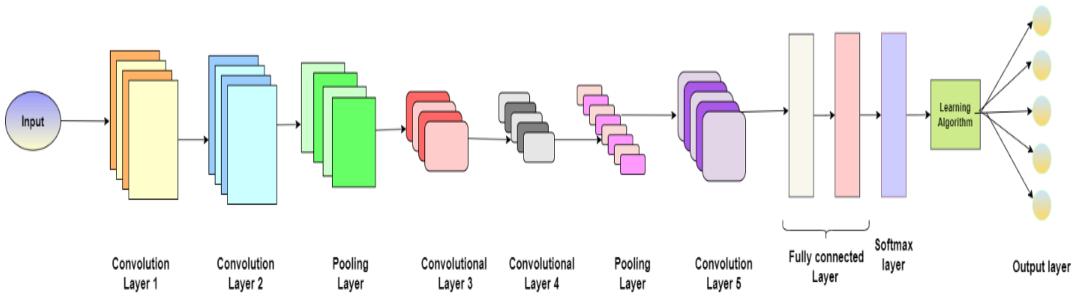

Figure 2: Architecture of AlexNet as Feature Extractor

(b) The second phase uses the concept of feature ranking mechanism for highlighting the characteristics that are most informative or discriminative for the job at hand, feature ranking helps guide feature engineering efforts. With this information, new features may be developed or current ones can be modified to enhance the performance of the model. In the proposed model, it uses the enhanced pipeline feature elimination mechanism. Especially for large datasets with plenty of features, neural network training can be computationally demanding. By removing some of the input elements, training times can be shortened and resource needs can be minimized. The process feature elimination is shown in figure 3

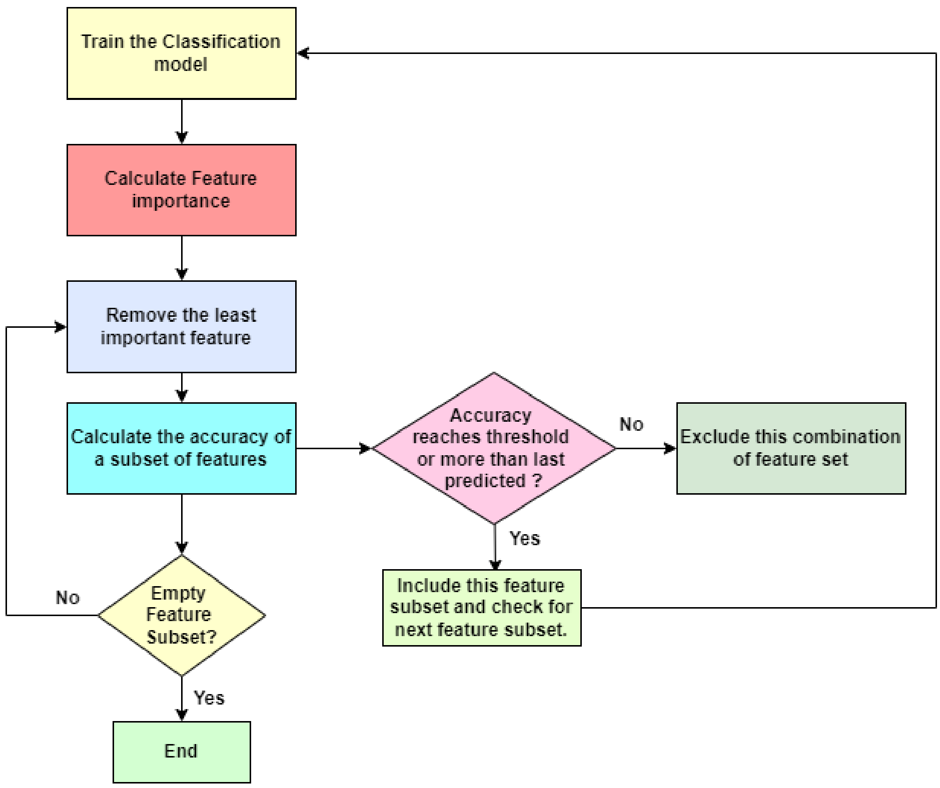

Figure 3: Feature Ranking using Elimination Process

The proposed research uses embedded regulation technique because integrates feature selection directly into the model training process, rather than treating it as a separate step. Features are selected directly during model training in embedded approaches. These techniques use the model's inherent processes to discover significant characteristics throughout the training process, hence selecting the most pertinent ones. A kind of linear regression known as Lasso involves enhancing the loss function with an L1 penalty term. The penalty term in question penalizes the absolute value of the coefficients, hence promoting sparse feature weights. DT, RF, & GBM separate features according to their significance in minimizing impurity or mistake. Tree-based models determine the value of each characteristic by evaluating their frequency of usage for splitting and the extent to which they decrease impurity. The significance scores may be used for the purpose of picking features. GBM are machine learning algorithms that enhance forecast accuracy by aggregating the predictions of numerous weak learners. ElasticNet regression model training: identify relevant features with non-zero coefficients. Utilize regularized tree models to train and assess the significance of attributes in order to identify the most essential ones.

(c) Types of Feature Ranking Mechanisms: Feature selection filtering approaches assess the significance of features using certain criteria, independent of any particular machine learning model. These techniques are very efficient in terms of CPU resources and may be used as a preliminary step before training a model. Choose characteristics for the target variable that have strong absolute correlation coefficients. Features with a weak association may be eliminated. Choose characteristics with strong target variable shared data. Target-specific features are more important. Choose features with strong ANOVA F-values, suggesting substantial mean differences across the variable of interest categories. Use categorical characteristics with strong chi-squared values to identify meaningful associations with the desired variable. Identify continuous characteristics that exhibit large disparities in means across multiple categories of the target variables. Discard low-variance characteristics since they may not improve model prediction. Figure 4 represents the feature ranking classification

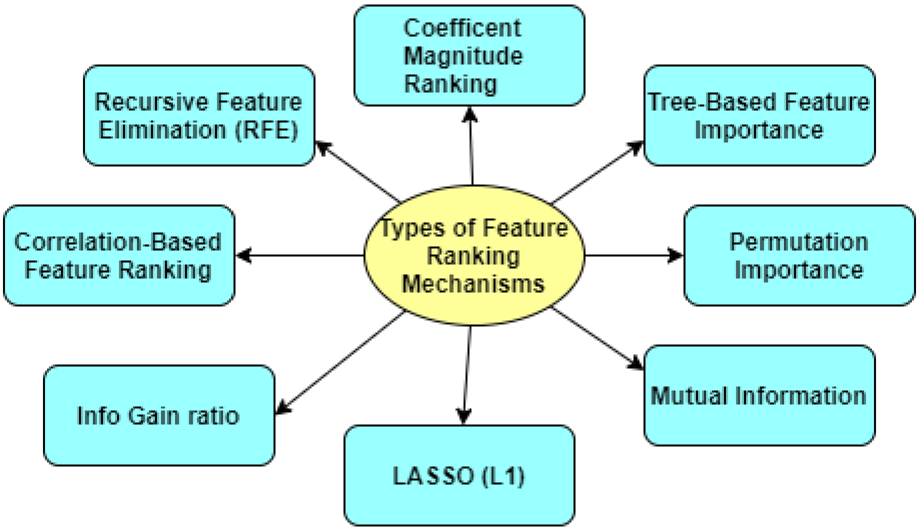

Figure 4: Classification of Feature Ranking

Wrapper techniques for feature selection include the selection of subsets of features depending on their performance in training a particular machine learning model. Beginning with a blank set of features, forward selection adds features repeatedly, evaluating each one's performance alone and in combination with previously chosen characteristics. Backward elimination begins with the whole collection of features and progressively eliminates a single attribute at a time, taking into account its individual impact on the vehicle's efficiency. Recursively removing features from the whole collection of features, RFE evaluates model performance. After training the model with all characteristics at first, the feature or features that are least significant are determined and eliminated. Bidirectional search starts with a selected group of features and repeatedly adds and removes features depending on their respective contributions to model efficiency.

Features are selected directly during model training in embedded approaches. These techniques use the model's inherent processes to discover significant characteristics throughout the training process, hence selecting the most pertinent ones. A kind of linear regression known as Lasso involves enhancing the loss function with an L1 penalty term. The penalty term in question penalizes the absolute value of the coefficients, hence promoting sparse feature weights. DT, RF, & GBM separate features according to their significance in minimizing impurity or mistake. Tree-based models determine the value of each characteristic by evaluating their frequency of usage for splitting and the extent to which they decrease impurity. The significance scores may be used for the purpose of picking features. GBM are machine learning algorithms that enhance forecast accuracy by aggregating the predictions of numerous weak learners. ElasticNet regression model training: identify relevant features with non-zero coefficients. Utilize regularized tree models to train and assess the significance of attributes in order to identify the most essential ones.

II. LITERATURE REVIEW:

In [1], Volkan Çetin et al attempted to develop an automated model using data pertaining to lungs and sophisticated artificial neural networks. A 5-fold layered crossover validation was conducted to the lung data training dataset in order to assess the models' performance. The complete training dataset was trained and evaluated on the test dataset after the best model was chosen using k-fold cross-validation. The logistic activation function back propagation algorithm was used, and the cost function—which is computed as the sum of squared errors—was applied. 26 features in the dataset were used to classify lung cancer into low, middle, and high categories. The data underwent a number of modification and preprocessing techniques to improve its appropriateness for predictive analysis. The model's inputs were the first 26 features. The data were categorized and normalized in the first stage. Later, during multilayer Perceptron training using Back propagation, a supervised learning algorithm—more precisely, the One Step Secant Back propagation algorithm—was used to categorize the risk of acquiring lung cancer.

In [2], Kwok Tai Chui et al created the modified generative adversarial network (MTL-MGAN) and multi-round transfer learning method for liquid crystal displays (LCDs). The MTL component eliminates the need for time-consuming searches of dataset prioritization across multiple data sets by transferring information between the target area and prioritized source domains. This is accomplished by customizing loss functions with regard to domain, instance, and feature in order to prevent negative transfer and maximize transferability through a multi-round transfer learning process. Ablation experiments were employed by the researchers to evaluate the effectiveness of each component inside the MTL-MGAN design. These studies aimed to assess the contributions of loss functions, multi-round transfer learning, the prioritization method, and the modified generative adversarial network (MGAN) to the prevention of negative transfers. The results of these studies attest to the efficacy of each element. The research has importance as it validates the feasibility of utilizing multi-round transfer learning to enhance the best solution of the target model. Moreover, the MTL-MGAN technique provides a generic way to connect the target and source domains, suggesting that it may have wide applications in the LCD context.

In [3], Umamaheswaran Subashchandrabose et al. investigated the diagnostic efficacy of an ensemble federated learning strategy for multi-order lung cancer. The suggested approach combines many machine learning methods trained on various datasets to classify multi-order lung cancer using ensemble federated learning. This method aims to increase generality and accuracy by protecting the privacy and confidentiality of the data. The technique achieved an astonishing 89.63% classification accuracy for lung cancer. The validation of distributed operational local neural networks—which are used to categorize lung cancer—was successfully finished. The potential application of this method to multidimensional medical models and electronic health records is highlighted, with the aim of developing a reliable recommendation and decision assistance system. The benefits and drawbacks of the proposed ensemble federated learning technique are discussed in the backdrop of health care uses with a focus on lung cancer diagnosis and classification. The authors discuss potential directions for further study, including enhancing optimization methods and exploring other local neural network topologies. This study demonstrates the use of distributed computation, federated learning, and complicated mathematical representations to protect patient privacy, facilitate group decision-making, and manage large-scale medical records.

In [4], Ruina Sun et al outlined the authors' method for classifying and segmenting pictures of lung cancer using an enhanced Swin Transformer model. The study found that the Swin Transformer model outperformed other models on tasks like as lung cancer segmentation and classification. The results showed that pre-training is a helpful strategy to increase the overall efficacy of the Swin Transformer model. Focusing on lung cancer pictures, the scientists specifically used their recently discovered segmentation technique—which is based on an efficient transformer—for medical image processing. In this medical imaging application, pre-training was demonstrated to be an effective method for improving the accuracy of the Swin Transformer model. The research also contrasted it with ViT (Vision Transformer) to show the former's efficacy and promise in the field of healthcare imaging. Notably, the study involved four seasoned thoracic radiologists, demonstrating a comprehensive and expert-driven evaluation of the recommended approach.

In [5], Silvia D'Ambrosi et al gave a presentation on a study that described a combinatory signature of circRNA and mRNA derived from blood platelets for the early identification of lung cancer. Researchers used a computational framework that leverages nCounter analysis of data and machine learning to investigate platelet mRNA and circRNA for the diagnosis of lung cancer. When blood platelet-derived circRNA and mRNA signatures are combined, they function as biomarkers and enhance the accuracy of lung cancer detection with a higher area under the curve (AUC) than whenever RNA type signatures are employed exclusively. Out of all the techniques, the random forest classifier with greatest ROC area under the curve and accuracy was based on a 28-mRNA signature. 48 ng of total RNA produced the highest counts, according to the examination of RNA concentrations. Using a sample of thirty patients, the researchers investigated how combining the measurement of platelets circRNA and mRNA could improve the identification of lung cancer. Because CircRNA is stable, abundant, and expresses specifically in NSCLC (non-small cell lung cancer), it has become a viable biomarker for liquid biopsy testing. Although the small number of platelet specimens in the study is a restriction, the encouraging outcomes show the promise of this methodology. Remarkably, every prediction model identified nodules in the lungs as noncancerous with accuracy. If the observed mRNA-circRNA signatures are to be considered effective, more validation using a larger sample size is required.

In [6], Lijun Hao et al presented ImAdaBoost, an enhanced AdaBoost technique for creating a classifier capable of differentiating between the breathing patterns of individuals with lung tumors and those in good condition. By fusing the well-known AdaBoost algorithm with the voting approach and K-fold cross-validation, the researchers creatively created sub-classifiers. Weights were assigned to a number of different sub-classifiers to generate the final integrated and improved classifier. For feature optimization, Principal Component Analysis (PCA) and the Genetic Algorithm (GA) were used to solve dimensionality concerns and improve classifier performance. The utilization of distinct training data in the two levels of the stacking model enhanced the robustness of the results. On the other hand, the AdaBoost method showed decreases in underfitting and increases in training accuracy. By analyzing the fluctuation of changes in performance throughout 100 tests, the study assessed the classifier's generalization and stability performance. The robustness and dependability of the classifier were demonstrated by the random and independent division of the training samples in each test.

In [7], Guillaume Chassagnon et al stressed the necessity of incorporating artificial intelligence (AI)-based technologies in order to digitize medical imaging, into clinical practice. Numerous studies have demonstrated that AI systems are as sensitive as radiologists in identifying lung nodules, but less specific. Importantly, deep learning-based computer-aided detection (CAD) technologies have proven to be able to reduce false positive outcomes while increasing radiologists' sensitivity. Using deep learning techniques or radiomics, artificial intelligence (AI) systems have been developed to forecast the cancer risk in lung nodules. One important aspect that affects how broadly radiomic-based CAD techniques for lung cancer detection may be applied is the effect of radiation exposure. The likelihood of malignancy in lung nodules has also been predicted using deep learning techniques; commercially available products primarily fall into the CAD-type software for detection group. The general consensus is that AI-based solutions will significantly contribute to reducing the medico-economic effect and expediting screening procedures. 26,309 participants, a sizable cohort, were included in the analysis.

In [8], Hassaan Malik et al created CDC_Net, a deep learning network that uses chest X-rays to classify COVID-19 along with other chest disorders. In terms of accuracy, CDC_Net performed better than other pre-trained models. By assessing lung opaqueness in chest X-rays, the Chest Disease Classifier Network obtains an impressive accuracy rate of 99.39% in distinguishing COVID-19 from various chest-related illnesses. For medical practitioners who depend on subjective diagnostic features from medical imaging, this computerized viewing and detecting technology is extremely valuable. CDC_Net performed better than other pre-trained classifiers, demonstrating its effectiveness in correctly identifying chest illnesses. Globally improving COVID-19 diagnosis is possible by the incorporation of artificial intelligence into radiological systems, as demonstrated by CDC_Net. A sizable dataset of 284,992,606 coronavirus-infected patients was used in the investigation.

In [9], Iftikhar Naseer et al discussed the importance of lung cancer, a dangerous malignancy that has a high death rate. Based on an altered version of the AlexNet architecture, the researcherss presented the LungNet-SVM model, which showed impressive lung nodule detection accuracy. Even though CT scans are the most common diagnostic technique for finding lung nodules, they take a long time. Current methods for feature extraction using convolutional neural networks (CNNs) frequently lack accuracy and efficiency. When it came to identifying pulmonary nodules from CT scans, the suggested LungNet-SVM model outperformed other cutting-edge research in terms of accuracy. Interestingly, in contrast to earlier studies that used more complicated models, the research concentrated on resolving asymmetries in the dataset in order to improve detection accuracy. One dataset was used for training, and just three CT picture sizes and optimizers were used in the development of LungNet-SVM, the lung cancer classification system. To improve lung cancer diagnosis, the authors propose that its adaptability be used to deploy it on different datasets with different neural network topologies and optimizers. In table 1, each method presents a trade-off between advantages and drawbacks, emphasizing the need for careful consideration based on specific project requirements and constraints.

Table 1: Comparative Analysis on Existing Approaches

S. No	**Author**	**Method**	**Merits**	**Demerits**	**Accuracy**
1.	Volkan Çetin	ANN	Uses back propagation method, automated	Should work on large datasets, costly	98.75 %
2.	Kwok Tai Chui	MTL-MGAN	Follows prioritization, feasible, reduced complexity	More number of iterations, impact of negative transfer	98.7 %
3.	Umamaheswaran Subashchandrabose	ensemble Federated Learning-based approach	Threshold based filtration, HoG and basic computations are implemented	Breach in reliability, no privacy of data.	89.63 %
4.	Ruina Sun	Swin Transformer	Uses hierarchical feature representations, UperNet for segmentation of image.	Only has 2D image processing, prone to over fitting.	99.91 %
5.	Silvia D'Ambrosi	bioinformatics pipeline	Identified 5 different biomarkers, RFECV with random forest has been introduced.	Less number of samples, more execution time	81%
6.	Lijun Hao	ImAdaBoost	Integrated classifiers improved performance, K-fold cross validation.	Impact of imbalance in classes, should apply feature optimization	92.14 %
7.	Guillaume Chassagnon	AI	Less computational time, limited impact on the economy.	High cost, prone to error	71.88 %
8.	Hassaan Malik	CDC_Net	3 base classifiers, even works on low resolution	Need to improve quality and lessen the cost	99.39 %
9.	Iftikhar Naseer	LungNet-SVM model	Removes asymmetries, adaptability	Trained and validated on same dataset, need to improve performance	97.64 %

II.1. RESEARCH GAPS IDENTIFIED:

A. Artificial neural networks:-

Which are renowned for being black-box systems—are used. To improve the interpretability of the model's predictions, techniques such feature importance analysis, SHAP (SHapley Additive exPlanations) values, and LRP (Layer-wise Relevance Propagation) should be investigated..

B. MLT-MGAN framework:-

While the MTL-MGAN framework may offer promising results, its scalability and computational efficiency need to be evaluated, particularly when dealing with large-scale datasets or real-time applications. Assessing the computational resources required and optimizing the framework for efficiency would enhance its practical utility and adoption.

C. .Principal Component Analysis (PCA):-

Although Principal Component Analysis (PCA) and Genetic Algorithm (GA) are used for feature optimization, the study does not discuss the rationale behind the choice of these techniques or their impact on classifier performance. Providing insights into the feature selection process and evaluating alternative feature optimization methods could improve the understanding of ImAdaBoost's behavior and effectiveness.

III. PROPOSED METHODOLOGY:

In the proposed methodology, the process starts with pre-processing to improve the quality of the images and then it uses the AlexNet to extract the features using the different layers. Then at last it uses the feature ranking mechanisms to reduce the features and the entire process is shown in figure 5

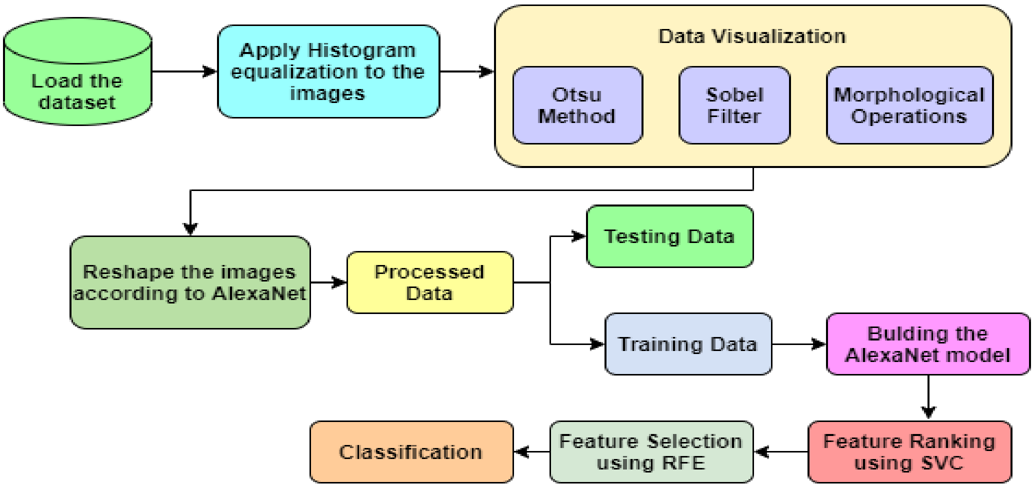

Figure 5: Proposed Architecture using Enhanced Feature Elimination

3.1. *Pre-processing Operations*: Pre-processing pictures before inputting them to neural network models is crucial for optimizing the performance of models, improving the caliber of input data, promoting learning, and guaranteeing compliance with the model design. Normalization aids in stabilizing the training procedure by guaranteeing that the input characteristics possess comparable scales. Performing mean subtraction and standard deviation normalization is an alternative method of normalization that centers the values of pixels around zero and scales them to have a variance of one. Data augmentation enhances the model's ability to handle certain transformations and enhances its resilience to changes in the input data by producing more versions of the input pictures. A common initial step is to resize photographs to a standard size, especially when working with datasets including images of different sizes. Transforming images from one color space to another can improve the model's generalization capabilities while reducing its computational complexity. Filters and histogram equalization are integrated throughout the process.

(a) Histogram Equalization: One technique to improve a picture's contrast is histogram equalization, which involves rearranging the pixel intensities in the image. The procedure entails calculating the image's histogram, which displays the pattern of pixel intensities, and then transforming the intensity values to distribute them throughout the whole dynamic range. The goal of this modification is to produce a picture with a more equal distribution of pixel values, which should increase contrast and make details easier to see. In essence, this leads to a more equitable split of pixel values by expanding the areas that were initially low contrast and compressing the parts that were initially high contrast. Histogram equalization significantly improves the look of pictures, making them more aesthetically pleasing and simplifying analysis for a variety of computer vision applications, including object identification, segmentation, and recognition. Histogram equalization may greatly enhance image quality, but it's important to remember that it also has the ability to magnify noise and add distortions from the original photo. Thus, careful consideration and additional processing steps may be required based on the specific application context.

(b) Ostu Method: This method is known as binarization or threshold otsu which is mainly utilized for the automation of threshold images, which separates both the objects that may be from back and front ground in the form of gray scale. The threshold holds primary request from the object with some value where it can segment the image is separated by the regions and required part is considered and remaining was removed or considered as object. Generally, an image holds bimodal categories with the histogram phase in the phase of two, this obtained by clear observation of separations. For every methodology from the image a optimized value was considered from those min and maximum values are considered where min holds the intra variances and the max holds inter variances with the related pixels in the image. This is necessary because by finding the threshold values the images back and front are differenced in the form of pixels. The computation of binarization is shown in equation (x) and calculations are shown in figure 6.

$$Optimal\_Threhold(\sigma^2) = argmax_t(P\left(\frac{i}{t}\right) * P\left(\frac{j}{t}\right) * (\mu_i - \mu_j)^2)) - (\mathrm{x})$$

Where,

$P\left(\frac{i}{t}\right), P(\frac{j}{t})$ represents probabilities of threshold with respect to class i & j

μ_i, μ_j represents mean values of class i & j

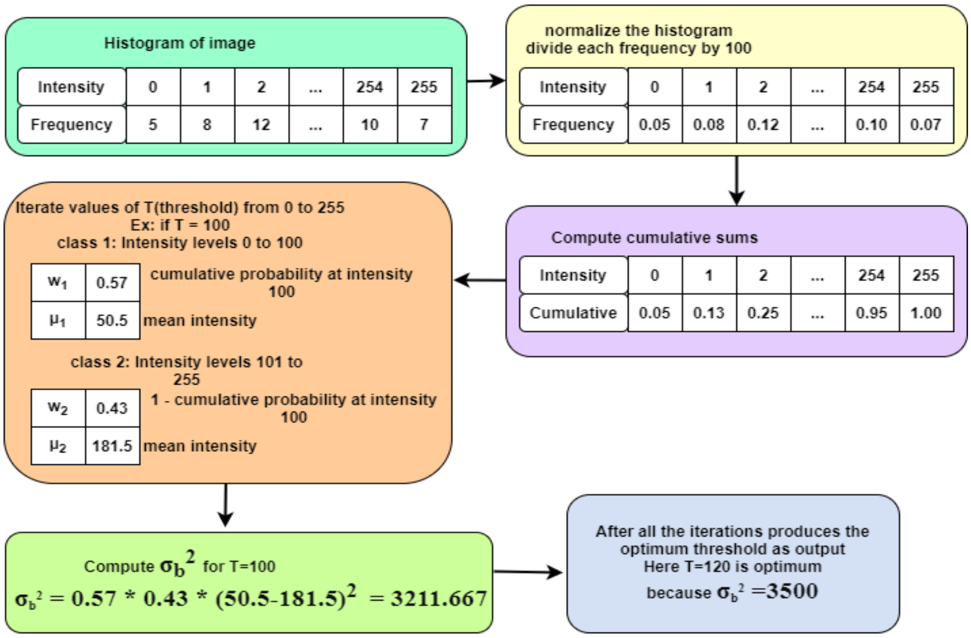

Figure 6: Computation of Optimal Threshold using OSTU

(c) Sobel Filter: The filter has got its name from the inventor, it is mainly utilized for the identification of edges from the image. This method can be applied from processing of image and computer vision technologies. The calculation can be employed by considering boundaries related to image with gradient magnitudes & related to direction of images with its pixel range. The structure of the image holds 2 3*3 conv kernels, this both filters execute on the horizontal and vertical axis of the directions. From all the sides the gradients are calculated by the computation through convolution model. This filter is applied on image with simple steps initially the image is convert to grayscale by applying edge detection. Now the image was gradient with horizontal component which has sliding with the sum of elements which access the couple of phases in kernel with related images in pixels this was processed by sobelX and parallelly worked on y-phase. The both pixels are powered by 2 which increases the edge pixels, then the outcome is generated to gradient direction. Finally, the thresholds are detected by setting and the values are considered above the threshold rate. This is how the filters are worked and the edges are detected. The gradient magnitude (G) at each pixel (x, y) in the image following the Sobel filter's application can be calculated using the equation (x) and sample calculations are shown in figure (7)

$$G(X,Y) = \sqrt{G_X(X,Y)^2 + G_Y(X,Y)^2} \text{ – (x)}$$

Where

$G_X(X,Y), G_Y(X,Y)$ are the gradient magnitudes in X & Y directions

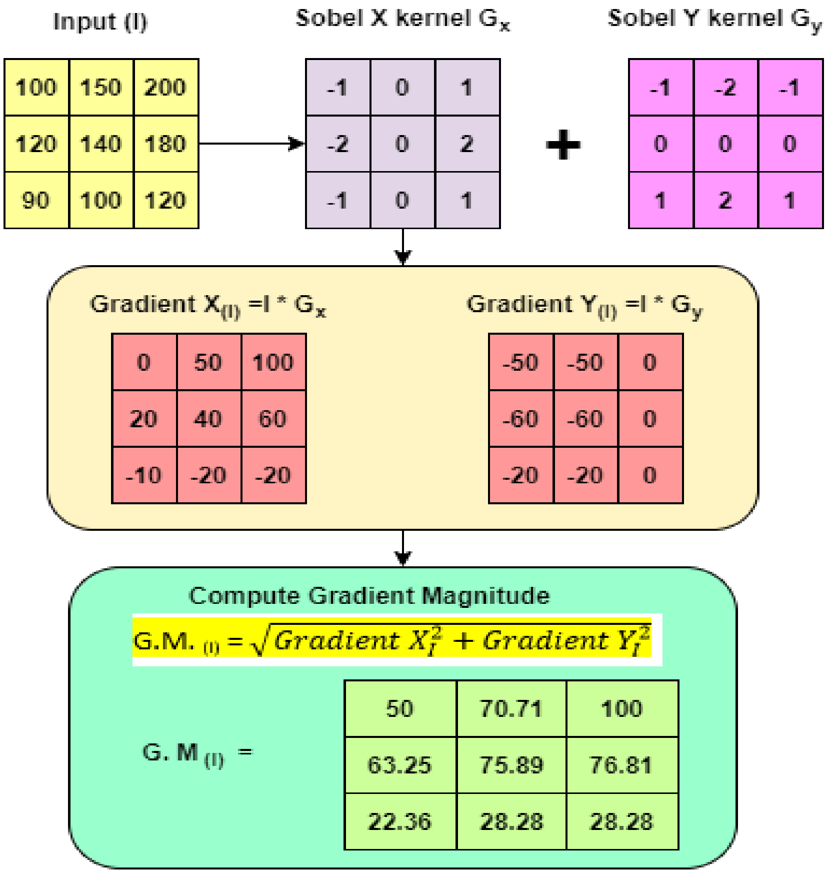

Figure 7: Sobel Filter Computations

(d) Morphological Operations: Morphological operations refer to a collection of image processing methods used to analyze and manipulate geometric structures present in pictures. These processes include altering the form or morphology of items in an image by using a structuring component, which is a tiny, predefined shape or kernel. These are often used for applications such as noise reduction, picture improvement, categorization, and extraction of features. Two surgeries, erosion and dilation, are undertaken. These foundational operations provide the basis for other ones, including opening, closing, and many permutations of erosion and dilation. Erosion operates by reducing the boundaries of items in a picture. Dilation is a morphological operation that enlarges the borders of items in a picture, serving as the reverse of erosion. Erosion and dilatation work together to create opening. Closing is the sequential application of dilatation and erosion. By subtracting the image's erosion from its dilation, we may get the morphological gradient. It delineates the perimeters of each item in the picture and is beneficial for detecting edges. Figure 8 presents the operations performed on the image.

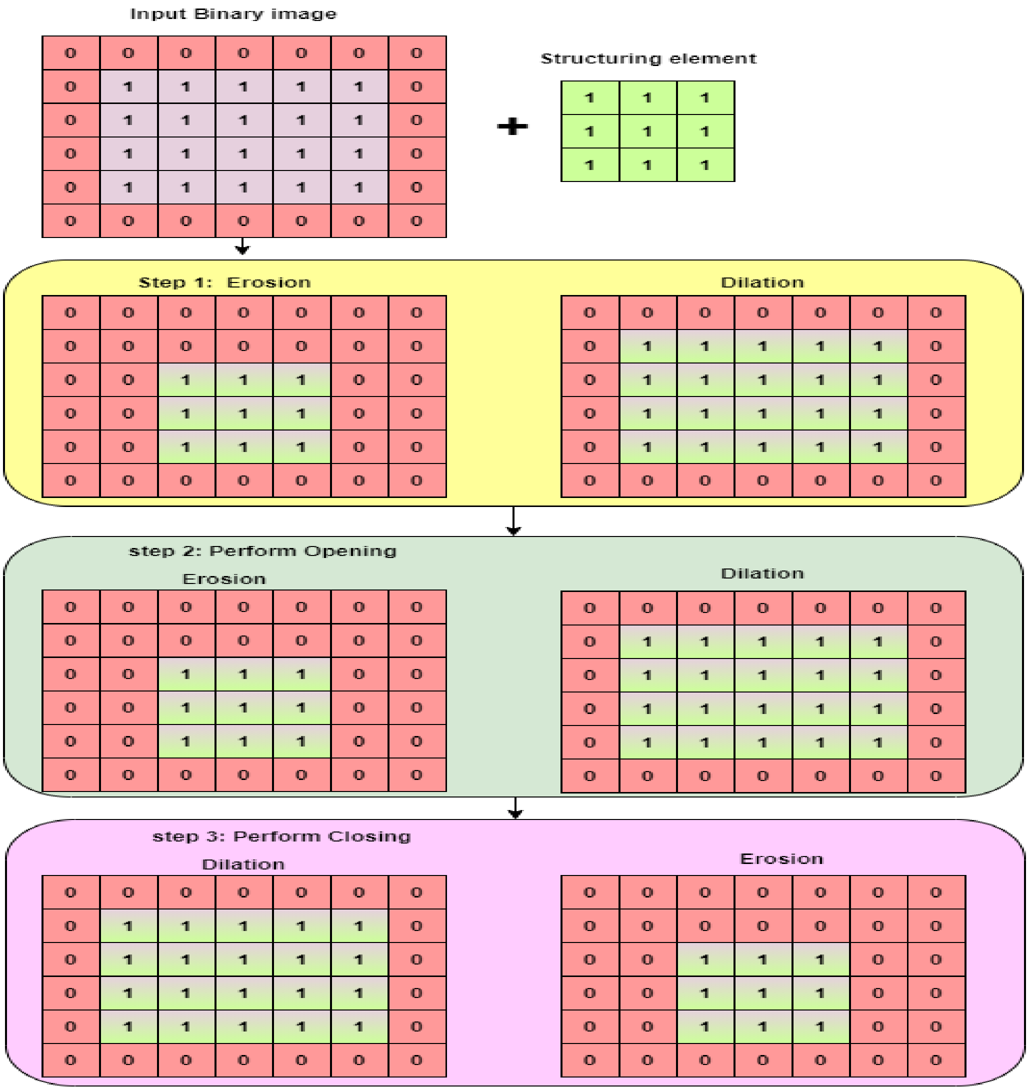

Figure 8: Morphological Operations

3.1.2. Need of Pre-trained Model: Normal initialization is an efficient technique utilized for the weight initialization process in DL. Assigning weight was processed during training which mainly impacts the processing speed and performance of the model. When training a network, it is important to avoid having activations disappear or explode due to unexpected changes in the network's parameters, which is why weight initiation is used. Normal initialization uses a distribution that is Gaussian with mean 0 and modest variation. Normal distributions are symmetric around their means and have a variance that defines their spread. Normal initialization maintains network activating variance. Sigmoid and tanh saturation may result from excessive weights and activations. The activities, on the other present, may become too tiny if the amount of weight are too little, which will make it impossible for the system to learn in an efficient manner. It is more probable that the activations will remain within an optimal range throughout the training process when the weights are initialized using a distribution that is normal with an appropriate variance. Different types of standard setup may be utilized depending on the job and the way the network is set up. For instance, networks containing ReLUs, which die if initialized incorrectly, may need a minor scaling factor change.

Pre-trained activation simplifies DNN training, particularly for small target datasets or restricted processing resources. Pre-trained initialization uses information from training with an extensive set of data to initialize model parameters, improving convergence and generalization on the target task. The pre-trained model captures multiple degrees of abstractions in the dataset by learning generic characteristics. The weights from the model that was previously trained can be used to start a new model that is meant for a different job or dataset after it has been trained on the big dataset and met performance standards. The new model is fine-tuned on the desired dataset after initializing with weights that have been trained. Fine-tuning adjusts pre-trained model weights to match the intended task or dataset. Most pre-trained models fine-tune just the last few layers, freezing the older layers to maintain learnt representations. The source dataset taught the pre-trained model general traits that can be used for the goal job. This may greatly minimize the data needed to train the desired model. Using pre-trained weights to initialize the model speeds up fine-tuning. TL helps transfer information from the original domain onto the target domain, improving generalization performance, particularly for limited datasets.

3.2. Feature Extraction using AlexNet: In AlexNet there is a peculiar layer known as “Local Response Normalization”. LRN is a method employed by neural network models to standardize the activations within a nearby region around each individual neuron. The AlexNet architecture, which triumphed in the ILSVR competition, had a significant role in popularizing it. The LRN technique is used after the ReLU activation function in many layers of the AlexNet framework. The goal of LRN is to improve the neurons' capacity of competing for activation in their immediate vicinity. The LRN in AlexNet is applied to the result of the ReLU activation function. This scaling factor is meant to reduce the effect of neurons that are too active and increase the rate of activity of neurons that are relatively strong in their immediate proximity. Each neuron's activation value in a feature map is subjected to a normalization process by LRN.

In order to divide the activation value in this procedure, a scaling factor that accounts for the activations of nearby neurons within a certain local region is used. The phrase "local response" describes the normalization process operating in a geographical area close to each particular neuron. As a result, each feature map undergoes a distinct normalization procedure that accounts for the activations of nearby map elements at the relevant spatial position. Within AlexNet, a given neighborhood is made up of many feature maps that are situated at the same geographic point. By normalizing responses, LRN improves the network's capacity for generalization and reduces the likelihood of over fitting. It also helps to draw attention to stronger activations and reduce weaker activations in the vicinity of those stronger activations, which improves the network's capacity to differentiate between the features it has learned.

Local Response Normalization (LRN) constituted a component of the AlexNet architecture, serving to normalize activations across adjacent channels. The mathematical expression for LRN implemented in AlexNet is depicted as follows:

Given an activation map $a_{i, j, k}$ at position (i, j) of channel k, LRN computes the normalized output $y_{i,, j ,k}$ as:

$$Y_{i, j, k} = \frac{a_{i,j,k}}{(k+\alpha \sum_{l=\max(0,k-\frac{n}{2})}^{\min(N-1,k+\frac{n}{2})} (a_{ijk})^2)^{\beta}} \quad \text{— (x)}$$

Where:

- N is the total number of channels.

- n is the size of the normalization window.

- α is a scaling parameter.

- β is an exponent.

- (i, j) represent the spatial dimensions of the activation map.

This equation computes the normalized activation yi,j,k for each location in the activation map ai,j,k, by dividing the activation by the sum of squares of activations within a local region around it, with size n. The scaling parameter α and the exponent β control the strength and shape of the normalization. This operation helps to enhance the relative activation of neurons with respect to their neighbors across different channels and shown in figure 9.

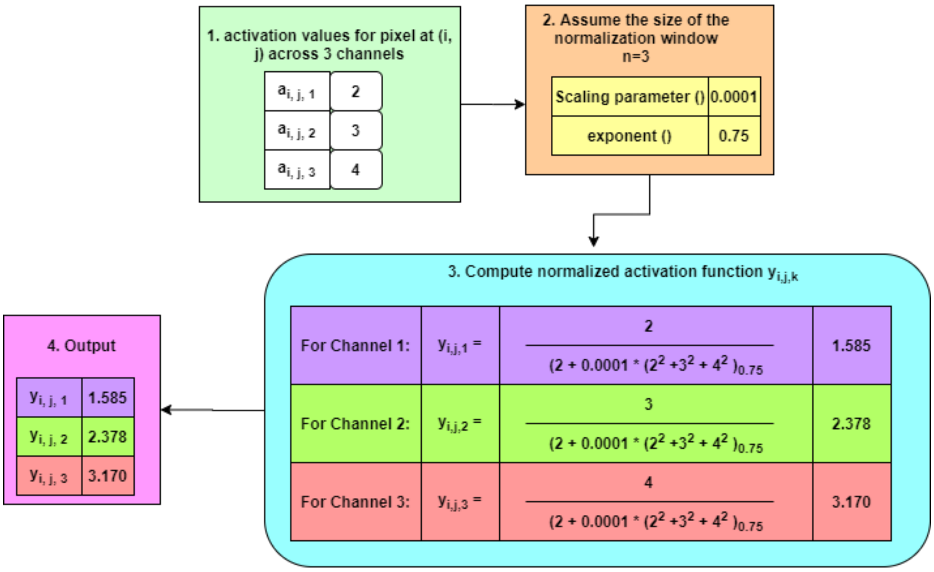

Figure 9: Working of Responsive Layer in AlexNet

3.2.1. Tuning of AlexNet Learning Rate: It is very important to pick the right learning rate when training neural networks because it has a direct effect on how fast they converge and how well the model they learn is. Learning rate scheduling adjusts training learning rates according to rules or heuristics. Training rate scheduling may minimize overshooting and bounce about local minima by progressively lowering the learning rate, helping the model settle to the global minimum. During training rounds, learning rate annealing progressively lowers the learning rate. These solutions gently lower the rate of learning, enabling the optimization process to delve into the loss landscape while avoiding local minima. Cyclical learning rates range between lowest and maximum values throughout training rounds. By regularly boosting its learning rate, the optimization process may avoid local minima and better explore the loss terrain. This method finds sharper and broader local minima, improving generalization. A technique called gradient clipping limits the size of slopes during training so they don't get too big, which could make the optimization process wobble or split. Gradient clipping stabilizes training and lets the optimizer explore the loss scenery, perhaps avoid local minima.

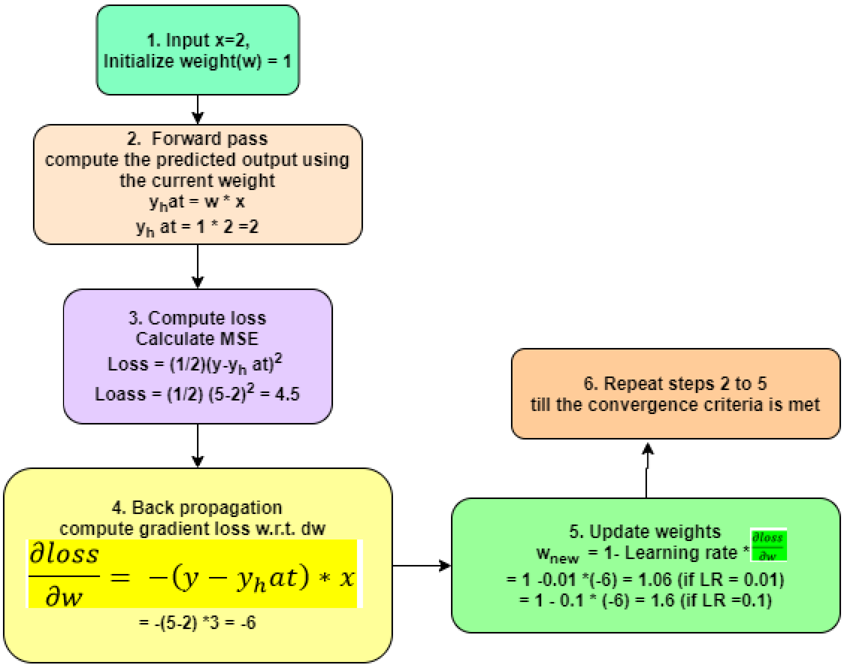

Figure 9: Working of Learning Rate

3.2.2. Tuning of Loss function: Cross-entropy loss, generally referred to as log deficit, is a widely utilized for the function of loss in classification problems, especially in the domain of neural networks. Calculated here is the quantification of the discordance between the target class labels' actual and anticipated probability distributions. The binary loss of cross-entropy is employed in binary classification problems, which are tasks in which each sample pertains to one of two classes, either 0 or 1. For simplicity, let's call the sample's true name s (either 0 or 1) and the chance that it belongs to class 1 q. Binary L, the binary cross-entropy loss, given a singular sample and shown in equation (x)

$$Binary\_Loss = \sum_{i=1}^{n} P_i \log(P_i) + (1 - P_i) \log(1 - P_i) \quad \text{– (x)}$$

Categorical type of loss is used when there are more than two categories and each sample needs to be put into one of them. Denote the actual probability distribution across C classes as y, and the calculated probability distribution as p. For a single sample, this (L_categorical) measures the disparity between the expected and real probability distributions.

3.3. Enhanced Feature Ranking Mechanism for Computation of Reduced Features: In feature selection approaches, feature ranking determines the value or usefulness of each dataset feature for a machine learning job. Different feature selection methods rank features differently. Linear regression and linear SVM use coefficients or weights to rank features. In tree-based algorithms like random forests, decision trees, and gradient-boosting machines, feature significance scores are based on how often a feature splits data or reduces impurity. Decision tree models & random forests employ information gain to assess data splitting features. Mutual information measures the knowledge gained from one random variable by another. Regularization methods like Lasso regression use the L1 penalty to reduce less significant feature coefficients to zero. RFE ranks features by their ML model performance contribution. qualities that correlate better with the variable of interest or other qualities may be more relevant. The process is shown in figure 10

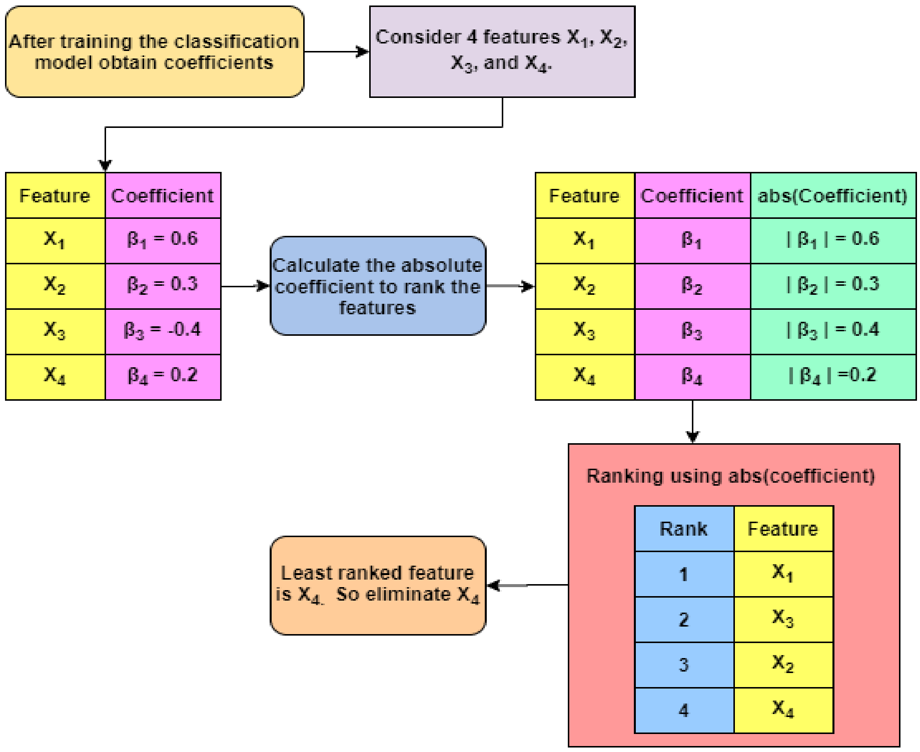

Figure 10: Working of Feature Ranking Mechanism

A popular way to choose the most important features from a dataset in machine learning is to use RFE. When combined with a linear SVM, RFE assists in identifying the group of features that are most useful for classifying. Start with all of the traits that are in the collection. Use all of the data to develop a linear SVM classifier. Figure out how important or useful each trait is to the classifier's work. A common way to do this is to look at the weights that the SVM model gives to each trait. Bigger weights mean that a feature is more important than a smaller one. Ignore the features that aren't important to the dataset. Predetermined or cross-validated hyperparameters define the number of features to delete at each iteration. Iterate through second until a stopping requirement is fulfilled. Possible limits include a set number of features left, a certain level of speed, or some other factor. Check how well the SVM algorithm works with the smaller set of traits.

The Support Vector Machine related to ML methodology which is often utilized for several classifications & many regression issues in different applications. The Support Vector Classifier is a specific sort of SVM that is specifically developed for the task of classification. The main goal of SVC has its major requirement to select the hyperplane with a high-dimensionality shades for effectively distinguishes between distinct classes in the given input data. The SVC is a ML methodology related to SVM family. SVM are very effective tools for performing classification and regression problems. The primary concept behind SVC is to identify an ideal hyperplane that can efficiently distinguish between distinct classes in the input data. The SVC seeks to determine a hyperplane that optimizes the separation between both groups in a classification that is binary situation. The margin, also known as the vector of support, is the measurement of the distance between planes that connects the hyperplane to the nearest information phases in each section. Finding the hyperplane that maximizes the margin and minimizes the classification error is the optimization strategy. SVC is very useful for handling complex and high-dimensional data since it uses kernel functions to efficiently handle non-linear relationships. With the help of these parameters, the algorithm may transfer the input data to a region with more dimensions, which makes it easier to find a hyperplane that might successfully separate the classes. SVM is a flexible technique that may be applied to a variety of tasks, including classification, especially when working with information that may not have linear separation. This tool is primarily useful in a variety of machine learning settings due to its capacity to handle large-scale areas and complicated interactions with efficiency. The steps are displayed in Figure 11.

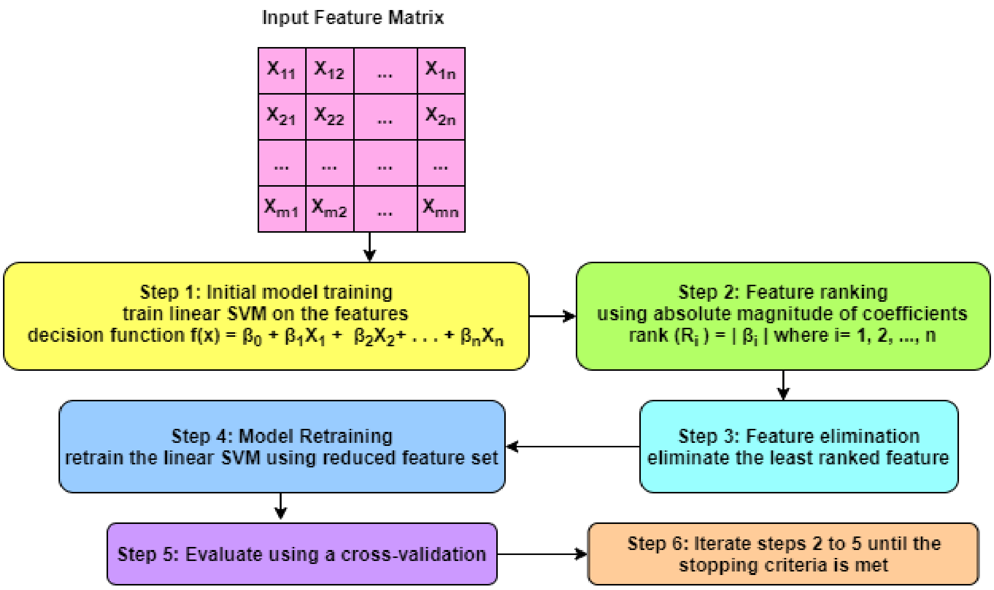

Figure 11: Working of Feature Ranking using Enhanced Feature Elimination

Algorithm for SVM with RFE:

Input Initial gene subst, G={1,2,3 . . . , n}

Output: Rank the list based on the criterion of smallest weight, denoted as R.

Set R={}

Repeat steps 3 to 8 until G is not empty

Train the SVM using G

Compute the weight vector using the equation W= ∑_(i=1)^n▒〖β_i x_i y〗_i

Compute the ranking criteria, Rank = W2

Rank the features as in sorted manner

Newrank = sort(Rank)

Rearrange the Feature Rank list based on the ranking criteria.

Update R = R + G(Newrank)

Decrease the feature subset by removing the feature with the lowest rank.

Update G= G-G(Newrank)

End

IV. RESULTS AND DISCUSSION:

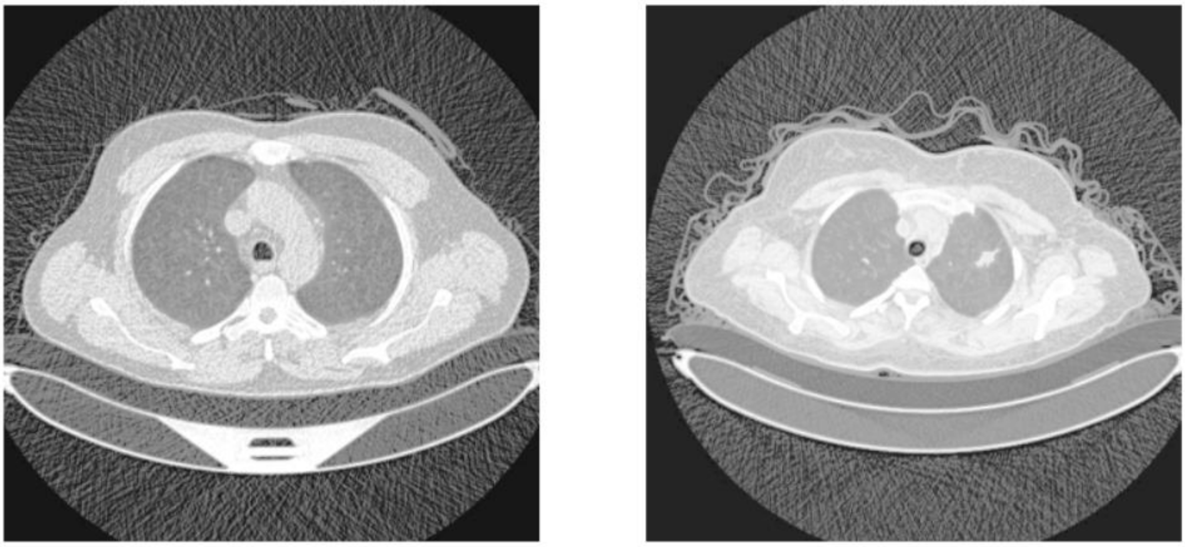

Figure 12: Histogram Equalization

The pictures shown in Figure 12 show the outcome of applying histogram equalization. First, the image's histogram is calculated. The histogram illustrates how frequently certain intensities appear in a picture. The converted image replaces each pixel in the original picture with its corresponding value, dispersing the intensity values to produce a more uniform histogram. The brightness levels of the original picture are then mapped to new intensity values to generate a transformation function.

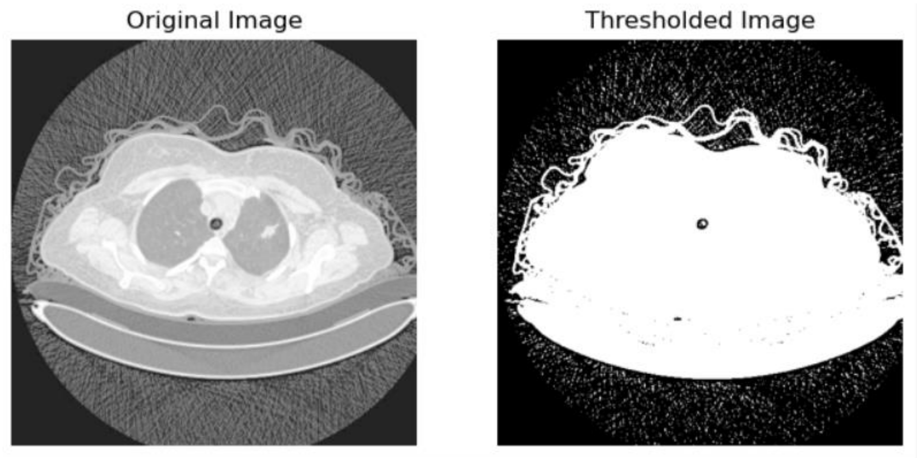

Figure 13: Optimal Threshold Image using OSTU

The original picture (on the left) and the resultant image (on the right) following the application of the ideal threshold via OTSU's approach are shown in figure 13 above. Otsu's approach computes the input image's histogram. The distribution of the pixel intensities within the picture is represented by the histogram, which also displays the number of pixels at each intensity level. Subsequently, Otsu's technique determines the pixel intensities' probability distribution function (PDF). By iterating over every potential threshold value, Otsu's approach determines the difference between both classes of pixels—the background and the foreground—that are divided by each threshold.

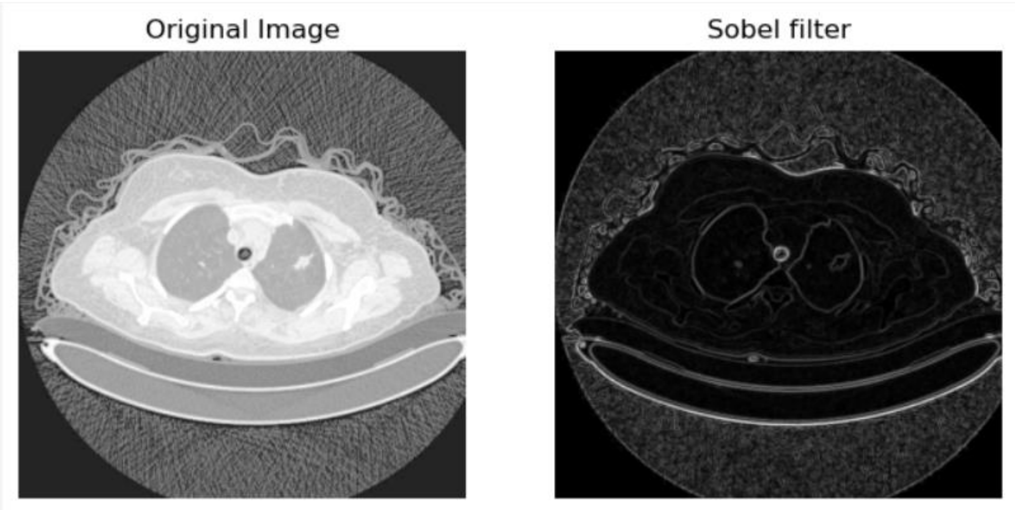

Figure 14: Sobel Filter

The Sobel filter uses two 3x3 kernels—one for horizontal change detection and the other one for vertical change detection—to convolve the picture. These kernels are intended to compute the vertical and horizontal gradient of the picture intensity. Every pixel position's gradient magnitude is determined after the picture has been convolved using both the vertical and horizontal Sobel kernels. The gradient magnitude image that is produced draws attention to areas within the original image that exhibit notable variations in intensity, suggesting the existence of edges. Higher gradient magnitude pixels are probably found near the edges of the picture, whilst lower gradient magnitude pixels are found in the smoother areas. The image's borders are identified by using the Sobel filter. The original picture (on the left) in figure 14 has been subjected to a Sobel filter, allowing its edges to be clearly visible.

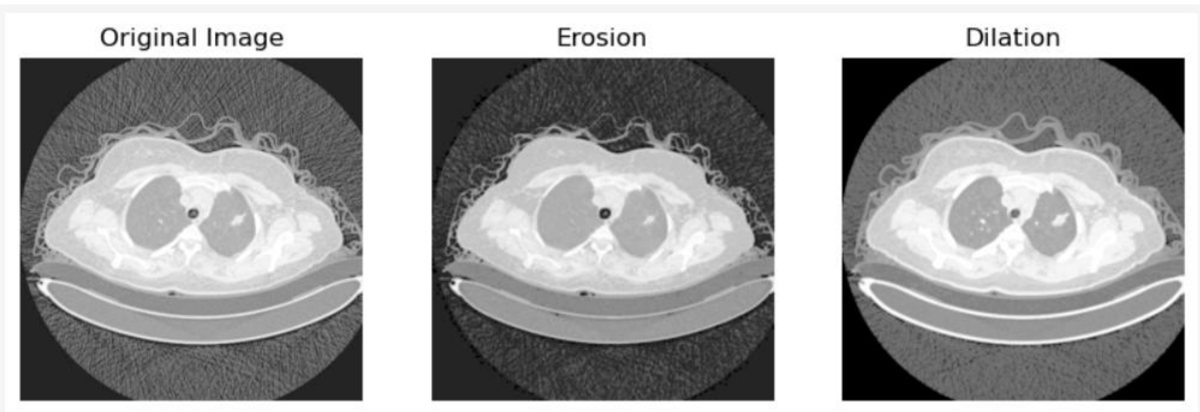

Figure 15: Morphological Operations

Figure 15 shows the original image's transformation following the application of the morphological processes of dilation and erosion, respectively. In these procedures, the layout and connectivity of surrounding pixels are used to modify the values of individual pixels. Opening, closure, dilatation, and erosion are examples of common morphological processes. By reducing the size of an object's borders, erosion may efficiently smooth out edges and remove minute features from a picture. Conversely, dilation enlarges an object's borders, giving them more thickness and prominence. The amount of the morphological alteration is determined by the dimensions and form of the structuring element.

```
 flatten_1 (Flatten)         (None, 9216)              0

 dense_3 (Dense)             (None, 4096)              37752832

 dense_4 (Dense)             (None, 4096)              16781312

 dense_5 (Dense)             (None, 3)                 12291

=================================================================
Total params: 58293635 (222.37 MB)
Trainable params: 58293635 (222.37 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Figure 16: Summary Report of AlexNet

AlexaNet's pre-trained layers are described in figure 16. By using learnable filters in convolutions, the layers of convolution are in charge of extracting characteristics found in the input pictures. The ReLU (rectified linear unit) activation function, which adds non-linearity to the network, comes after each convolutional layer. Between the convolutional layers, max-pooling layers are inserted to extract dominating features and minimize spatial dimensions. Based on the acquired attributes, the fully linked layers carry out sophisticated reasoning and decision-making. This AlexaNet consists of five thick layers.

```
Epoch 25/30
27/27 [==============================] - 52s 2s/step - loss: 0.1349 - accuracy: 0.9112 - val_loss: 0.1780 - val_accuracy: 0.8854
Epoch 26/30
27/27 [==============================] - 51s 2s/step - loss: 0.1354 - accuracy: 0.9231 - val_loss: 0.2566 - val_accuracy: 0.8438
Epoch 27/30
27/27 [==============================] - 50s 2s/step - loss: 0.2305 - accuracy: 0.8604 - val_loss: 0.2245 - val_accuracy: 0.8542
Epoch 28/30
27/27 [==============================] - 50s 2s/step - loss: 0.1524 - accuracy: 0.9101 - val_loss: 0.2213 - val_accuracy: 0.8646
Epoch 29/30
27/27 [==============================] - 51s 2s/step - loss: 0.1060 - accuracy: 0.9325 - val_loss: 0.1478 - val_accuracy: 0.9062
Epoch 30/30
27/27 [==============================] - 50s 2s/step - loss: 0.1265 - accuracy: 0.9266 - val_loss: 0.1886 - val_accuracy: 0.8646
```

Figure 17: Epochs Training of AlexNet

As seen in figure 17, AlexNet usually travels through several epochs during training, where one epoch is one full run through the training dataset. Every epoch entails feeding the training pictures into the network, calculating the loss over the actual and predicted labels, and using back propagation to update the weights and biases of the network. A variety of variables, including dataset dimensions, complexity, and available processing power, can affect how many epochs are needed to train AlexNet. By modifying the model's parameters to reduce training loss and enhance generalization to new data, each epoch aids in the model's learning process.

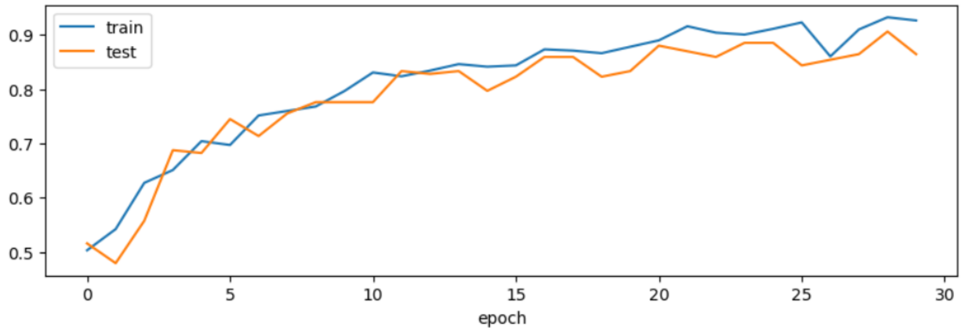

Figure 18: Model Accuracy

Figure 18 presents the epochs-based accuracy computation on both the training and validation data. In case of training accuracy, the graph is gradually increased and it has stability and in case of validation data, it has fluctuation because limited data in the validation. X-axis represents epochs and Y-axis represents the accuracy in terms of percentages. A high accuracy score means that for a sizable chunk of the dataset, the model is producing accurate predictions. It suggests that the algorithm is operating effectively for the job for which it was trained. A low accuracy score indicates that a significant amount of the dataset is being incorrectly predicted by the model, indicating poor performance.

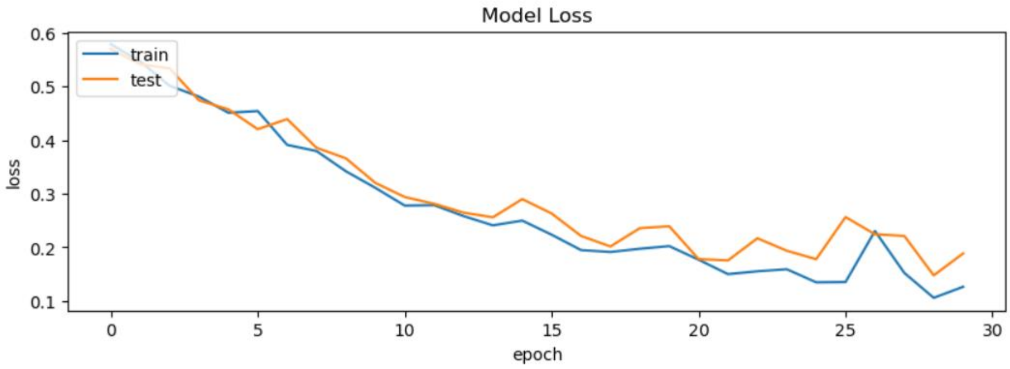

Figure 19: Model Loss

Figure 19 presents the loss of training and validation data. In terms of loss, both the data behaves similarly and it decreases with the increase of epochs. The loss is almost reaches 0 which represents the stability of proposed system in terms of efficiency. One of the most important metrics for evaluating how well the model performed during training is the loss function. It measures the difference between the goal values in the data set used for training and the projected outputs, and it directs the optimization process to reduce this difference. The model learns to generate increasingly exact forecasts on the training set of data by iteratively modifying its parameters so as to reduce the loss function using optimization methods.

```
The Selected features :[[  0.          47.24814     0.         ...  82.8028     16.14533
    0.        ]
 [  0.          56.416924    0.         ... 132.85991     5.775679
    0.        ]
 [181.3984    142.44394     0.         ...   0.        145.55185
  197.76926   ]
 ...
 [  6.5135136    9.087408    0.         ...  19.151442   11.540529
   22.336061 ]
 [ 55.479248    0.         41.53575   ...   0.         36.02696
   71.81886   ]
 [ 11.291108   26.117966    0.         ...  50.308365   24.543304
   27.505402 ]]
```

Figure 20: Selected Features

The top-ranked features are chosen for further model development. These features form the subset of attributes that are deemed most informative for predicting the target variable. Feature selection is often an iterative process. After evaluating the model's performance, if necessary, additional feature engineering or selection steps may be undertaken to further refine the model. Feature selection is a critical step in machine learning model development, aiming to improve model performance, reduce over fitting, and enhance interpretability. The figure 20 illustrates the iterative process of selecting relevant features to build an effective predictive model.

V. CONCLUSION

Radiologists can spend less time manually interpreting medical images since deep learning algorithms make it possible to automate jobs like detecting lung cancer. By increasing productivity, automation can let radiologists concentrate on treating patients and handling more complicated cases. The proposed research uses "AlexNet", in which Local Response Normalization (LRN) of feature maps improves contrast and highlights features. For feature extraction tasks, when it's critical to capture small variations between features, this can be helpful. Neighboring neurons compete with one another through LRN, which amplifies the response of highly activated neurons while inhibiting those with lower activations. Improved feature discrimination in the extracted representations may result from this. Data's underlying patterns and relationships can be understood by ranking aspects according to importance. The interpretability of the model is improved, and domain knowledge integration is made easier, when one knows which aspects are most important for the task at hand. The proposed research uses "Enhanced Feature Elimination using the SVM" to reduce the features extracted from the neural network.

REFERENCES

[1] Çetin, Volkan, Hacı Hüsnü Yumrukaya, and Çiğdem Bakır. "Detection of Lung Cancer with Enhanced Feed Forward Backpropagation Artificial Neural Networks." (2023): 2163-2171.

[2] Chui, Kwok Tai, Brij B. Gupta, Rutvij H. Jhaveri, Hao Ran Chi, Varsha Arya, Ammar Almomani, and Ali Nauman. "Multiround transfer learning and modified generative adversarial network for lung cancer detection." International Journal of Intelligent Systems 2023 (2023): 1-14.

[3] 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. https://doi.org/10.3390/diagnostics13193053

[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. https://doi.org/10.3390/electronics12041024

[5] D'Ambrosi, S., Giannoukakos, S., Antunes-Ferreira, M., Pedraz-Valdunciel, C., Bracht, J. W. P., Potie, N., Gimenez-Capitan, A., Hackenberg, M., Fernandez Hilario, A., Molina-Vila, M. A., Rosell, R., Würdinger, T., & Koppers-Lalic, D. (2023). Combinatorial Blood Platelets-Derived circRNA and mRNA Signature for Early-Stage Lung Cancer Detection. International Journal of Molecular Sciences, 24(5), 4881. https://doi.org/10.3390/ijms24054881

[6] Hao, Lijun, and Gang Huang. "An improved AdaBoost algorithm for identification of lung cancer based on electronic nose." Heliyon 9, no. 3 (2023). https://doi.org/10.1016/j.heliyon.2023.e13633

[7] Chassagnon, G., De Margerie-Mellon, C., Vakalopoulou, M., Marini, R., Hoang-Thi, T.-N., Revel, M.-P., & Soyer, P. (2022). Artificial intelligence in lung cancer: current applications and perspectives. Japanese Journal of Radiology. https://doi.org/10.1007/s11604-022-01359-x

[8] Malik, H., Anees, T., Din, M., & Naeem, A. (2022). CDC_Net: multi-classification convolutional neural network model for detection of COVID-19, pneumothorax, pneumonia, lung Cancer, and tuberculosis using chest X-rays. Multimedia Tools and Applications, 82(9), 13855–13880. https://doi.org/10.1007/s11042-022-13843-7

[9] Naseer, I., Masood, T., Akram, S., Jaffar, A., Rashid, M., & Amjad Iqbal, M. (2023). Lung Cancer Detection Using Modified AlexNet Architecture and Support Vector Machine. Computers, Materials & Continua, 74(1), 2039–2054. https://doi.org/10.32604/cmc.2023.032927

[10] Shakeel, P.M., Burhanuddin, M.A. & Desa, M.I. Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier. Neural Comput & Applic 34, 9579–9592 (2022). https://doi.org/10.1007/s00521-020-04842-6

[11] Jianxin Feng, Jun Jiang, "[Retracted] Deep Learning-Based Chest CT Image Features in Diagnosis of Lung Cancer", Computational and Mathematical Methods in Medicine, vol. 2022, Article ID 4153211, 7 pages, 2022. https://doi.org/10.1155/2022/4153211

[12] Shafi, I., Din, S., Khan, A., Díez, I. D. L. T., Casanova, R. del J. P., Pifarre, K. T., & Ashraf, I. (2022). An Effective Method for Lung Cancer Diagnosis from CT Scan Using Deep Learning-Based Support Vector Network. In Cancers (Vol. 14, Issue 21, p. 5457). MDPI AG. https://doi.org/10.3390/cancers14215457

[13] A. Mallikarjuna Reddy, V. Venkata Krishna, L. Sumalatha, "Efficient Face Recognition by Compact Symmetric Elliptical Texture Matrix (CSETM)", Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 4-Regular Issue, 2018.

[14] Mallikarjuna Reddy, A., Rupa Kinnera, G., Chandrasekhara Reddy, T., Vishnu Murthy, G., et al., (2019), "Generating cancelable fingerprint template using triangular structures", Journal of Computational and Theoretical Nanoscience, Volume 16, Numbers 5-6, pp. 1951-1955(5), doi: https://doi.org/10.1166/jctn.2019.7830

[15] Mallikarjuna Reddy, A.,Venkata Krishna, V. and Sumalatha, L." Face recognition approaches: A survey" International Journal of Engineering and Technology (UAE), 4.6 Special Issue 6, volume number 7 , 117-121,2018.

[16] Swarajya Lakshmi V Papineni, Snigdha Yarlagadda, Harita Akkineni, A. Mallikarjuna Reddy. Big Data Analytics Applying the Fusion Approach of Multicriteria Decision Making with Deep Learning Algorithms International Journal of Engineering Trends and Technology, 69(1), 24-28, doi: 10.14445/22315381/IJETT-V69I1P204.

[17] Sudeepthi Govathoti, A Mallikarjuna Reddy, Deepthi Kamidi, G BalaKrishna, Sri Silpa Padmanabhuni and Pradeepini Gera, "Data Augmentation Techniques on Chilly Plants to Classify Healthy and Bacterial Blight Disease Leaves" International Journal of Advanced Computer Science and Applications(ijacsa), 13(6), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130618.

[18] Santhosh Kumar, C.N., Pavan Kumar, V., Reddy, K.S., Similarity matching of pairs of text using CACT algorithm, International Journal of Engineering and Advanced Technology, 2019, 8(6), pp. 2296–2298, DOI:10.35940/ijeat.f8685.088619

[19] Mallikarjuna A. Reddy, Sudheer K. Reddy, Santhosh C.N. Kumar, Srinivasa K. Reddy, "Leveraging bio-maximum inverse rank method for iris and palm recognition", International Journal of Biometrics, 2022 Vol.14 No.3/4, pp.421 - 438, DOI: 10.1504/IJBM.2022.10048978.

[20] Reddy, K.S., Naga Santhosh Kumar, C.H., Effective data analytics on opinion mining, International Journal of Innovative Technology and Exploring Engineering 8(10), pp. 2073-2078, 2019..

[21] A. M. Reddy, K. SubbaReddy and V. V. Krishna, "Classification of child and adulthood using GLCM based on diagonal LBP," 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Davangere, 2015, pp. 857-861, doi: 10.1109/ICATCCT.2015.7457003.

[22] Hosseini, S.H., Monsefi, R. & Shadroo, S. Deep learning applications for lung cancer diagnosis: A systematic review. Multimed Tools Appl 83, 14305–14335 (2024). https://doi.org/10.1007/s11042-023-16046-w.

[23] Mohamed TIA, Oyelade ON, Ezugwu AE (2023) Automatic detection and classification of lung cancer CT scans based on deep learning and ebola optimization search algorithm. PLoS ONE 18(8): e0285796. https://doi.org/10.1371/journal.pone.0285796.

[24] Shah, A.A., Malik, H.A.M., Muhammad, A. et al. Deep learning ensemble 2D CNN approach towards the detection of lung cancer. Sci Rep 13, 2987 (2023). https://doi.org/10.1038/s41598-023-29656-z.

[25] V. NavyaSree, Y. Surarchitha, A. M. Reddy, B. Devi Sree, A. Anuhya and H. Jabeen, "Predicting the Risk Factor of Kidney Disease using Meta Classifiers," 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), 2022, pp. 1-6, doi: 10.1109/MysuruCon55714.2022.9972392.

[26] S. Mamidisetti and A. M. Reddy, "A stacking-based ensemble framework for automatic depression detection using audio signals," International Journal of Advanced Computer Science and Applications, vol. 14, no. 7, pp. 603–612, 2023.

[27] P. S. Silpa *et al*., "Designing of Augmented Breast Cancer Data using Enhanced Firefly Algorithm," *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, 2022, pp. 759-767, doi: 10.1109/ICOSEC54921.2022.9951883

[28] Mamidisetti, Suresh and A. Mallikarjuna Reddy. "Enhancing Depression Prediction Accuracy Using Filter and Wrapper-Based Visual Feature Extraction." *Journal of Advances in Information Technology* (2023),14(6), pp. 1425–1435.

[29] Naik, S., Kamidi, D., Govathoti, S., Cheruku, R., & Mallikarjuna Reddy, A. (2023). Efficient diabetic retinopathy detection using convolutional neural network and data augmentation. *Soft Computing*, 1-12.