INTERNATIONAL JOURNAL OF RESEARCH CULTURE SOCIETY
Monthly Peer-Reviewed, Refereed, Indexed Journal
Volume - 7, Issue - 5, May - 2023

ISSN(O): 2456-6683
[Impact Factor: 6.834]
Publication Date:31/05/2023

IJRCS

DOIs:10.2017/IJRCS/202305016 --:-- Research Paper / Article / Review

A Modified Neural Network for Lung Cancer detection Using Machine Learning Approach

[1] Akhilesh Kumar Dubey, [2]Dr. Arvind Kourav, [3]Prof. Nehul Mathur

[1]M.Tech Student, Department of ECE, Bhopal Institute of Technology, Bhopal (M.P.),INDIA
[2]Professor, Department of ECE, Bhopal Institute of Technology, Bhopal (M.P.),INDIA
[3]Assistant Professor, Department of ECE, Bhopal Institute of Technology, Bhopal (M.P.),INDIA
Email - [1]akhilesh.dubey16@gmail.com, [3]mathurnehul@gmail.com

Abstract: *Cancer is a disease that is becoming increasingly prevalent around the world. Researchers have conducted numerous studies to determine where on the human body cancer is most common. This research work is proposed to find out the early stage of lung cancer and explore the accuracy levels of various machine learning algorithms. Due to improper handling of DICOM images, there is low accuracy and a high implementation cost. For medical image processing, many different types of images are used, but computer tomography (CT) scans are generally preferred because of less noise. The proposed approached performances are evaluated based on their accuracy, sensitivity and classification. Machine learning based lung cancer prediction models have been proposed to assist clinicians in managing incidental or screen detected indeterminate pulmonary nodules. Such systems may be able to reduce variability in nodule classification, improve decision making and ultimately reduce the number of benign nodules that are needlessly followed or worked-up.*

Key Words: *Lung Cancer, Machine Learning, Feed Forward network, Neural Network and deep learning.*

1. INTRODUCTION:

Lung cancer is diagnosed in the United States at a rate second only to that of breast cancer. Lung cancer patients have a survival rate of only 15% five years after their diagnosis. Survival analysis is a common topic in medical research. The survival rate of cancer patients can be predicted using a predictor variable that indicates whether or not certain events, such as death or recurrence of a disease, have occurred over a specified time period. A patient's prognosis after a cancer diagnosis must be predicted by the predictor models. You have two sponge-like organs in your chest, the lungs, which are responsible for breathing. He has three lobes in each of his proper lungs. The left lung has two lobes on each side. The left lung is reduced in size to compensate for the increased size of the heart. Air enters your lungs when you inhale through your nose or mouth and travels down your trachea (windpipe). The trachea in the lungs divides into several smaller bronchi.

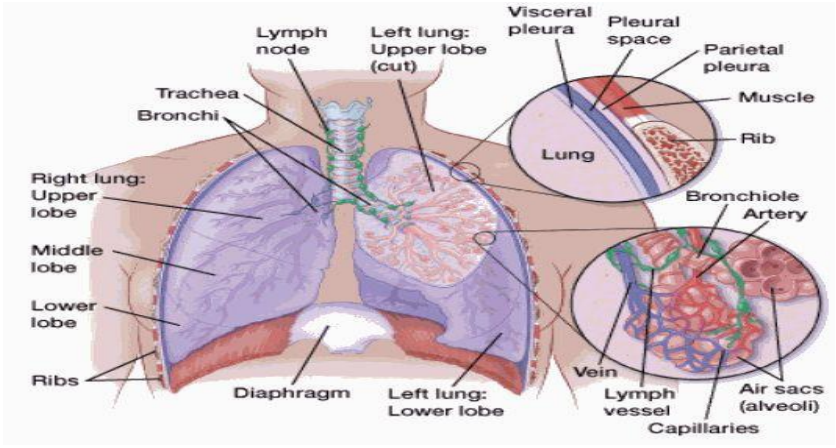

Fig.1. Normal structure and function of the lungs [1]

Smaller branches, known as bronchioles, branch out from the main bronchial tree and are called bronchial branches. The bronchial tubes culminate in little air sacs called alveoli. Carbon dioxide is expelled from the blood as

you inhale oxygen via the alveoli. Your lungs are responsible for taking in oxygen and exchanging carbon dioxide. The lining of the bronchi and other parts of the lung, such as the bronchioles or alveoli, are common places for lung cancer to begin.

The lungs are protected from the outside world by the pleura, a thin layer of lining. The pleura act as a cushion between human lungs and the chest wall, allowing them to expand as well as contract when you breathe. This dome-shaped diaphragm divides the upper and lower torsos by forming a barrier between the two areas. When you inhale and exhale, the diaphragm rises and falls, causing the lungs to fill and empty.

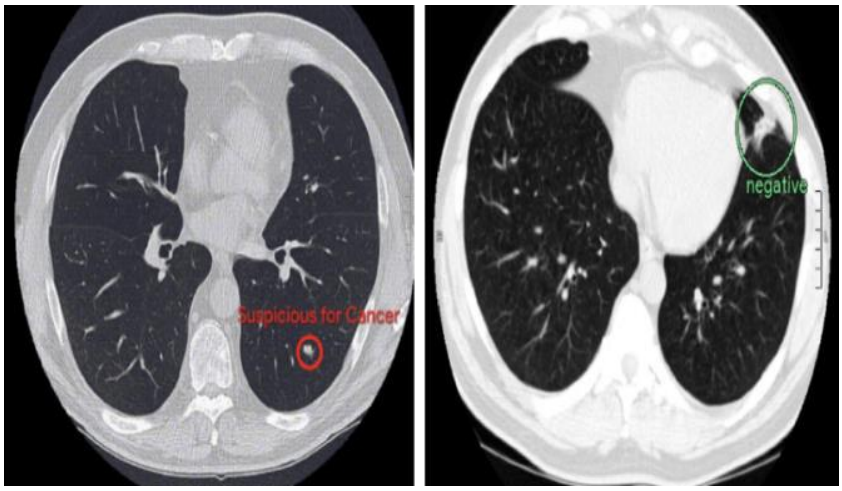

Fig. 2 less noisy as compared to MRI lung cancer

Lung cancer is the leading cause of death in the United States. In 2012, there were 1.6 million deaths and 1.8 million new cases reported. Among both men and women, it is the leading cause of cancer-related death in the United States, accounting for more deaths than all other cancers combined. Cancers of the lungs have a 17.8% five-year survival rate, but that number would be much higher if the disease had been discovered earlier. Only 15% of the cases were caught in the early stages, however. This shows that early detection of lung cancer is critical to the success of treatment. Lung cancer is the leading cause of cancer death in smokers. As a result, it has been suggested that cancer may be a result of an individual genetic disposition inherited from family members. In other words, some people are genetically predisposed to developing lung cancer because of genetic mutations or flaws in a gene.

Machine Learning Artificial Intelligence In Cancer Detection

It is essential to have a solid comprehension of what machine learning is, as well as what it is not, before beginning a detailed analysis of which types of machine learning methods work best for which kinds of scenarios. Specifically, it is important to have an understanding of what machine learning is not. The field of artificial intelligence research known as machine learning employs statistical, probabilistic, and optimization techniques in order to "learn" how to classify new data discover new patterns, or forecast new trends based on previous examples (Mitchell 1997).

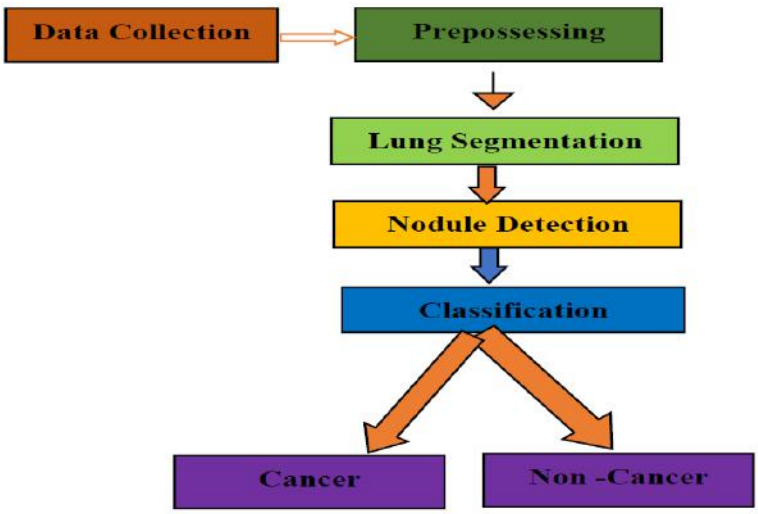

Fig. 3 Block Diagram for Machine Learning based Lung Cancer Detection

In the beginning, natural language processing methods are used to convert the collected input text into a binary format. However, machine learning methods are then used to analyse this binary data in order to produce the appropriate output and decisions.

- The procedure for identifying variations together in tumours is going to be detailed down below.
- "Images of hearts are categorised based on the shape of the heart itself."
- The method is proposed for use in predicting issues related to the heart.
- Dermatologists would be able to make accurate diagnoses of tumours if they used artificial intelligence.
- A framework for the application of artificial intelligence in the intensive care unit.
- Identifying individuals who are at risk of developing cervical cancer using algorithms is currently being researched and developed.
- It has been shown that AI can be helpful in the diagnosis of breast cancer.

2. PROPOSED METHOD FOR FEATURE SELECTION:

COLOR AUTO CORRELOGRAM

The correlation of a signal with a delayed copy of itself is known as autocorrelation. In the case of discrete time, this type of correlation is referred to as serial correlation. In its most fundamental form, it can be understood as the correlation between observations in terms of the time lags that separate them. Analyzing autocorrelation patterns, for instance, enables one to determine the presence of a periodic signal that is obscured by noise as well as its fundamental frequency. This can be done in either of two ways. The examination of functions and sequences of data, such as time domain signals and temperature readings, is one of its many applications. Autocorrelation can be defined in a number of different ways, and no two of these definitions are identical to one another. The terms "autocorrelation" and "auto covariance" are frequently used interchangeably in the scientific community when talking about autocorrelation. There are many different kinds of autocorrelation processes; some examples include moving average processes, autoregressive processes, trend-stationary processes, and unit root processes.

Auto-correlation of stochastic processes Real or sophisticated random processes can be described as having an autocorrelation, in which case the Pearson correlation is used to compare the process's values across time rather than as a constant. As examples, a random process and any point in time (whether an integer or a real number in continuous time) are taken as examples. The value (or realization) that results from a particular process conducted at a given point in time. The mean and variance for each time point in the process are assumed to be constant.

Where denotes the expected value operator and the bar is an illustration of complex conjugation. Take into consideration that the expectation might not be clearly defined. The auto-covariance function between times and values is obtained by first subtracting the mean and then multiplying the values.

ANN Approach

ANN initial divides training information into many subsets, exploitation ambiguous clustering techniques. After this, it trains different ANNs by exploitation different subsets. Then it determines the membership grade of those subsets and connects them through a replacement ANN to induce the ultimate result. The whole structure of ANN is shown in steps. Within the kind of a selected machine learning framework, feed forward ANN covers each the coaching part and therefore the testing part.

Step - I - Featured Selected data training

Step II: Training for training by specific teaching algorithms for training (i = 1, 2, k), ANN model, ANN, (i = 1,2, ... k) for every training set numerous Base ANN model Step III: to scale back the error for every ANN, we tend to simulate the ANN exploitation the complete training set TR and find the results. Then we tend to use the membership grade, which were generated by the ambiguous cluster module to combine the results.

After this, tend to train another new ANN exploitation combined results. Within the testing part, we tend to input directly the take a look at set information into numerous ANN and receive the output. Supported these outputs, final results is achieved by ultimate ambiguous aggregation module.

There are three necessary lawsuits in three phases of the ANN structure-

- Produce totally different training subsets from the initial training dataset TR;
- Produce different base models ANN with different coaching subsets;
- A way to collect numerous results made by numerous base models ANN.

Forward Neural Network

A directly proportional relationship between input and output occurs in perception, whereas a connection between input and output occurs in FFNNs. The activation function in the hidden layer creates a nonlinear connection. There is a network with a direct connection between the input layer and the output layer that is formed by combining the connection form based on perception with a multilayer network. Cascade Forward Neural Networks (CFNN) are the result of this connection pattern (CFNN). The following are the equations that can be derived from the CFNN model.

$$y= \sum_{i=1}^{n} f^{i}\omega_{t}^{i}\chi_i + f^{0}\left(\sum_{j=1}^{n}\omega_j^0 f_j^h\left(\sum_{i=1}^{n}\omega_{ji}^h\chi_i\right)\right) \quad 1$$

Where $f^i$the activation function from the input is layer to the output layer and ω_{ji}^h is weight from the input layer to the output layer. If a bias is added to the input layer and the activation function of each neuron in the hidden layer is then equation (5.2) becomes

$$y = \sum_{i=1}^{n} f^i\,\omega_i^i\chi_i + f^0\left(\omega^b + \sum_{j=1}^{k}\omega_j^0 f^h\left(\omega_j^b + \sum_{i=1}^{n}\omega_{ji}^h\chi_i\right)\right) \quad 2$$

In this research, the CFNN model is applied in time series data. Thereby, the neurons in the input layer are the lags of time series data $Xt-1, Xt-2, \ldots, Xt-p$, whereas the output is the current data Xt . The architecture of CFNN model in predicting time series is shown at figure 4.

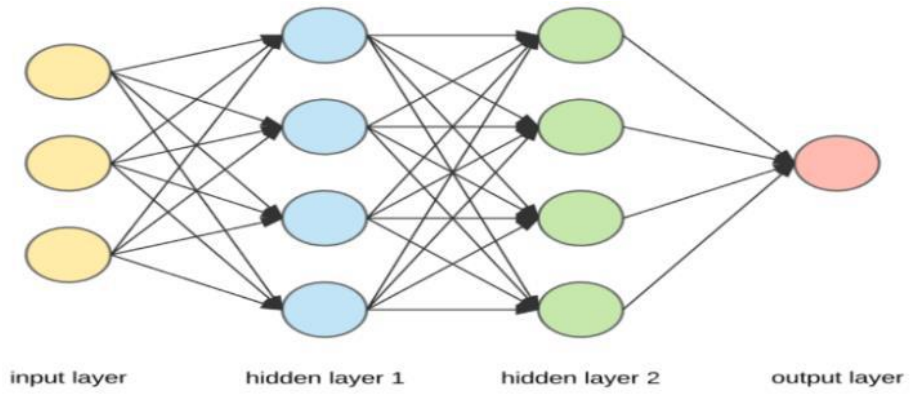

Fig 4 Different layers of back proportion method

3.5 Training Back Propagation

Back-propagation is a technique employed in artificial neural networks to calculate a gradient that's required within the calculation of the weights to be utilized in the network. it's ordinarily accustomed train deep neural networks, a term bearing on neural networks with quite one hidden layer. There are two main kind of neural network feed forward and back forward network. There are three main layers in neural network.

- Input Nodes – The Input nodes give data from the surface world to the network and are along noted because the “Input Layer”.
- Hidden Nodes – The hidden nodes don't have any direct association to the surface of the planet (hence the name is "hidden"). They calculate and transfer information from the input nodes within the output nodes. A group of hidden nodes creates a "hidden layer". Whereas a feed forward network can exclusively have one input layer and one output layer, it'll have zero or multiple Hidden layer.

Output Nodes – The Output nodes are together noted because the “Output Layer” and are accountable for computations and transferring data from the network to the outside world.

3. SIMULATION AND RESULTS:

The simulation model and its results will be discussed in this chapter along with the proposed algorithm. Matrix laboratory should be utilised for the process of implementing the proposed algorithm. The Matrix Laboratory is a well-known piece of software that can be used for various calculations related to the implementation of algorithms for data analysis. The number of data analysis tools available in MATLAB is quite extensive.

The results of the simulation of the proposed method for developing middleware using machine learning technique for histopathological image-based lung cancer detection are shown in this section, along with the result calculation and the simulation of the proposed method.

Result Parameters

There are different results parameters are analysis in the presented method. These parameters are shown in below.

Accuracy(ACC)

$$\text{Accuracy} = (TP + TN) / S$$

When S is the total number of samples in the test set, FP is the number of false positives and FNR is the false negative rate. Sensitivity is the probability of a positive test, given the plant in view is the decease.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

Where: TP = True positive; FP = False positive; TN = True negative; FN = False negative.

Selectivity (S)

A challenge of estimating the size of query results, known as "selectivity estimate," is common one in database management systems. Query selectivity estimation with numerous associated characteristics is very difficult to get right. The query processor may choose a suboptimal plan if it is unable to accurately estimate the size of the database. Deep learning has now been used to tackle this issue and the results are encouraging. When it comes to multi-attribute queries with multiple predicates and low selectivity, several of the offered methods fail to provide reliable answers.

$$\text{Selectivity} = \text{mean}((tp ./ (tp + fp)))*100;$$

Simulation outcomes and Results

In the above discuss the GUI design for simulation of proposed method shows discusses the simulation outcomes of proposed method. The proposed method simulation divided into three parts. In the first part discuss the simulation outcomes on 70% training data and remaining 30% data for testing as well as validation. Similar that apply training with 80% data for testing use remaining 20% data.

Training testing outcomes on 70% data

In the training of proposed method use GUI interface with modified feed forward neural network, number of epochs, 15 and other training parameters as an input. For initialization of training process follow these steps,

1. Using browse button to select the target input folder data.
2. After that in the next enter training ratio in the editable text window. Give the file name in the for training data.

Fig. 5 GUI of training data with 70%

3. Now click on train button.
4. Start training, after completion of training, the outcomes of proposed method in 70% training accuracy is 98.5556% .
5. The confusion matrix of proposed method also shown in the GUI.

In the above figure 6.7 shows the training outcome of proposed method.

ISSN(O): 2456-6683
[Impact Factor: 6.834]
Publication Date:31/05/2023

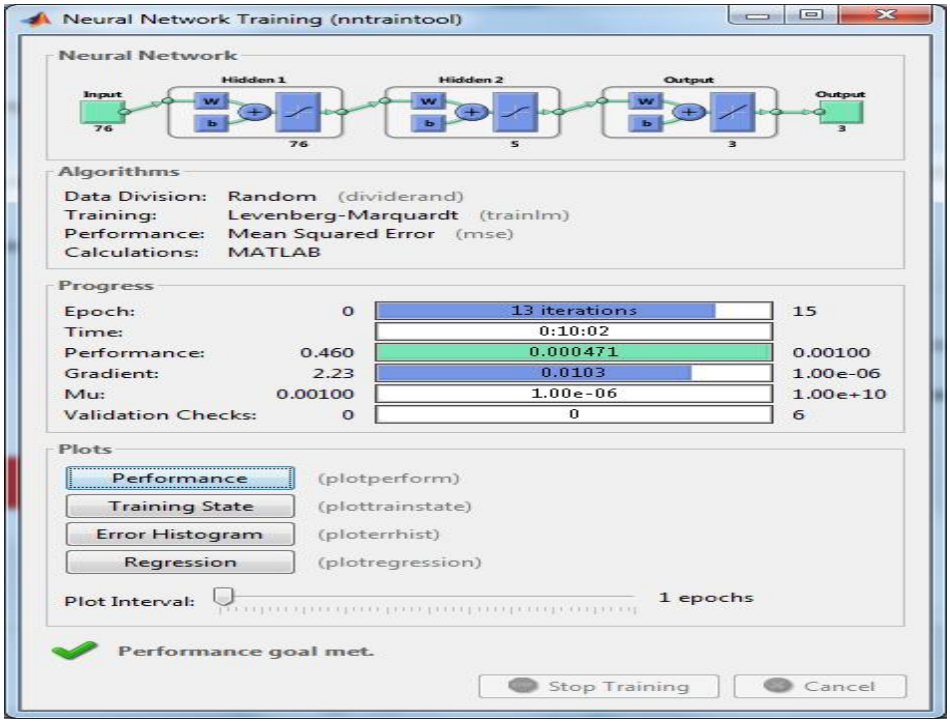

Fig. 6 Neural Network and Input Parameters

In the above figure 6 shows the neural network input parameters. The proposed method is based on feed forward neural network.

- There are 76 input parameter are selected as an input parameters.
- There are two input hidden layers are apply.
- Data division is in random process.
- For training process use levenbergmarquardt (LBM).

These are major input parameters which are apply in the input of the proposed method. Now discuss the testing outcome plots. There are three major plot for the performance analysis of proposed method. First is performance validation, second one is training state, third one is error histogram and last one is regression plot.
shows best validation performance output **13** epoch and the mean square error **0.0085855**.

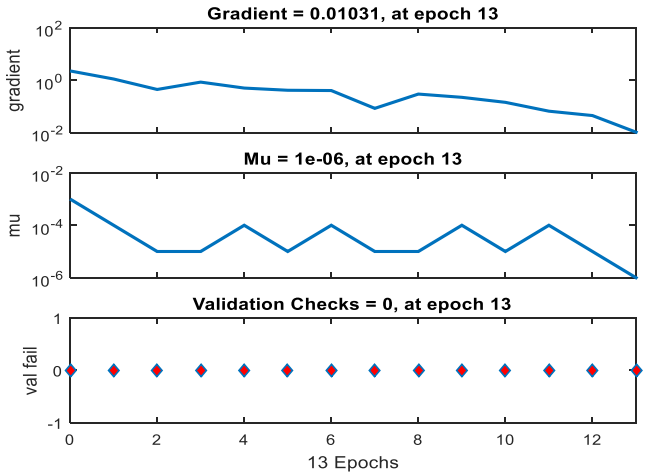

Fig. 7 Shows the proposed method gradient, Mu, and validation

In the above figure 7 shows the proposed outcomes in terms of gradient, mu, and validation checks. In the above divided into four subplots., in the first subplots shows the gradient outcome plot of proposed work, in second subplot shows the Mu plot with optimum error value in the last shows validation checks. In the X axis shows the number of epochs in the simulation and Y axis in the first plot denote gradient value, second plot denote mu value and third plot shows the validation fail.

4. CONCLUSION AND FUTURE WORK:

The Cancer is a disease that is becoming increasingly prevalent around the world. Numerous researchers have conducted a variety of studies in an effort to determine the areas of the human body that are most commonly affected by cancer. The results of one study like this one motivated us to carry out this research in the field of detecting lung cancer. The most common cause of death attributable to cancer in both men and women is lung cancer. The likelihood of a favourable prognosis for lung cancer patients who undergo early detection is increased. The utilisation of image processing systems makes it possible to detect and diagnose abnormalities earlier and more quickly than is possible with the use of other screening tests. When developing a method for the early diagnosis and treatment of disease, taking into account the passage of time is one of the considerations that goes into the process.

REFERENCES:

1. Bharti, Meenakshi, Jaytrilok Choudhary, and Dhirendra Pratap Singh. "Detection and Classification of Pulmonary Lung Nodules in CT Images Using 3D Convolutional Neural Networks." In *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1, pp. 1319-1324. IEEE, 2022.
2. Halder, Amitava, Saptarshi Chatterjee, and Debangshu Dey. "Adaptive morphology aided 2-pathway convolutional neural network for lung nodule classification." *Biomedical Signal Processing and Control* 72 (2022): 103347.
3. Gnanavel, S. "Identification and Classification of Lung Nodules Using Neural Networks." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12.6 (2021): 1956-1961. https://turcomat.org/index.php/turkbilmat/article/view/4799
4. Abdullah, Dakhaz Mustafa, Adnan Mohsin Abdulazeez, and Amira Bibo Sallow. "Lung cancer prediction and classification based on correlation selection method using machine learning techniques."Qubahan Academic Journal1.2 (2021): 141-149.https://journal.qubahan.com/index.php/qaj/article/view/58
5. Sujitha, R., and V. Seenivasagam. "Classification of lung cancer stages with machine learning over big data healthcare framework." Journal of Ambient Intelligence and Humanized Computing 12 (2021): 5639-5649.https://link.springer.com/article/10.1007%2Fs12652-020-02071-2
6. Shanthi, S., and N. Rajkumar. "Lung cancer prediction using stochastic diffusion search (SDS) based feature selection and machine learning methods." Neural Processing Letters 53.4 (2021): 2617-2630. https://link.springer.com/article/10.1007/s11063-020-10192-0
7. Doppalapudi, Shreyesh, Robin G. Qiu, and YouakimBadr. "Lung cancer survival period prediction and understanding: Deep learning approaches." International Journal of Medical Informatics 148 (2021): 104371.
8. Naik, Amrita, and Damodar Reddy Edla. "Lung nodule classification on computed tomography images using deep learning." Wireless Personal Communications 116.1 (2021): 655-690.
9. Patel, Divek, Connor Cowan, and Prateek Prasanna. "Predicting Mutation Status and Recurrence Free Survival in Non-Small Cell Lung Cancer: A Hierarchical ct Radiomics–Deep Learning Approach." 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE, 2021.
10. Kumar, Alok, Sandeep Kumar Shukla, Archana Sharma, and Pranay Yadav. "A Robust Approach for Image Super-Resolution using Modified Very Deep Convolution Networks."In 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), pp. 259-265. IEEE, 2022.
11. Mishra, Akhil, Ritu Shrivastava, and Pranay Yadav."A Modified Cascaded Feed Froward Neural Network Distributed Denial of Service Attack Detection using Improved Regression based Machine Leaning Approach."In 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1292-1299. IEEE, 2022.
12. Tiwari, Sandeep, Nitesh Gupta, and Pranay Yadav. "Diabetes Type2 Patient Detection Using LASSO Based CFFNN Machine Learning Approach."In 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN), pp. 602-608. IEEE, 2021.
13. Tiwary, Abhigyan, M. Kumar, and Pranay Yadav."Prediction of Covid-19 Patient in United States of America Using Prophet Model." In 2021 International Conference on Advances in Technology, Management & Education (ICATME), pp. 94-99. IEEE, 2021.
14. Tiwari, Prayag, Pranay Yadav, Sachin Kumar, Brojo Kishore Mishra, Gia Nhu Nguyen, Sarada Prasad Gochhayat, Jagendra Singhk, and Mukesh Prasad. "Sentiment analysis for airlines services based on Twitter dataset." Social Network Analytics: Computational Research Methods and Techniques 149 (2018).
15. Singh, Jagendra, Mukesh Prasad, Yousef Awwad Daraghmi, Prayag Tiwari, Pranay Yadav, Neha Bharill, Mahardhika Pratama, and Amit Saxena. "Fuzzy logic hybrid model with semantic filtering approach for pseudo relevance feedback-based query expansion." In 2017 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1-7. Ieee, 2017.

16. Chavate, Shrikant, Ravi Mishra, and Pranay Yadav. "A Comparative Analysis of Video Shot Boundary Detection using Different Approaches." In 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), pp. 1-7. IEEE, 2021.
17. Yadav, Pranay. "Color image noise removal by modified adaptive threshold median filter for RVIN." In 2015 International Conference on Electronic Design, Computer Networks & Automated Verification (EDCAV), pp. 175-180. IEEE, 2015.
18. Sharma, Shachi, and Pranay Yadav. "Removal of fixed valued impulse noise by improved Trimmed Mean Median filter." In 2014 IEEE International Conference on Computational Intelligence and Computing Research, pp. 1-8. IEEE, 2014.
19. Yadav, Pranay, and Parool Singh. "Color impulse noise removal by modified alpha trimmed median mean filter for FVIN." In 2014 IEEE International Conference on Computational Intelligence and Computing Research, pp. 1-8. IEEE, 2014.
20. Gupta, Vikas, Dilip Kumar Gandhi, and Pranay Yadav."Removal of fixed value impulse noise using improved mean filter for image enhancement."In 2013 Nirma University International Conference on Engineering (NUiCONE), pp. 1-5. IEEE, 2013.
21. Sharma, Bharti, Sachin Kumar, Prayag Tiwari, Pranay Yadav, and Marina I. Nezhurina. "ANN based short-term traffic flow forecasting in undivided two lane highway." Journal of Big Data 5, no. 1 (2018): 1-16.