INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

LUNG CANCER DETECTION AND CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS

YANDRATI VENKATA RANI
Research Scholar
Department of Computer Science and Engineering
J.S. UNIVERSITY, Shikohabad, UP

Dr Vikram Singh Rathore
Professor
Department of Computer Science and Engineering
J.S. UNIVERSITY, Shikohabad, UP

Dr VittapuMani Sarma
Professor
Department of Computer Science and Engineering
MRCM&MS,JNTUH,Hyderabad

Abstract -Cancer-related mortality is mostly caused by lung cancer, which is the main cause of death. The identification of lung cancer at an early stage is essential for having better treatment choices and a better prognosis. When it comes to the diagnosis of lung cancer, histopathological pictures of lung tissue are often what are used. In spite of this, the use of histology images for the purpose of detecting lung cancer calls for a significant amount of skill and experience. As a result of their remarkable picture categorization skills, Convolutional Neural Networks (CNNs) have the potential to significantly improve the identification of lung cancer in biopsy pictures. In this study, we offer a convolutional neural network (CNN) method to diagnosing lung cancer using structural analysis of biopsy pictures. This technique seeks to identify lung cancer. Should it be put into action, the technique that we have described has the potential to significantly enhance the precision of lung cancer diagnosis. Around the world, lung cancer is one of the top causes of mortality that may be attributed to cancer. The prognosis and prognosis of a patient are significantly enhanced when

they get medical assistance in a timely manner. For the purpose of establishing a diagnosis, medical professionals evaluate histopathological pictures of lung tissue that has been biopsied from locations that may be affected. It might be difficult to differentiate between the many subtypes of lung cancer caused by the fact that there are lots of them. The use of Convolutional Neural Networks allows for the detection and classification of a wide range of lung tumors with greater precision. Healthcare providers are able to determine the most effective treatment choices and the chances of a patient's survival with the use of this information. The purpose of this inquiry is to examine the similarities and differences between healthy tissue and squamous cell cancer and adenocarcinoma. It has been determined that the CNN model has a training accuracy of 96.11% and a confirmation accuracy of 97.20%.

Key words: Convolutional Neural Networks, lung cancer, diagnosis.

1.INTRODUCTION

Lung cancer is the most common kind of cancer that affects humans. Around twenty-five percent of all cancer cases are lung cancer. passes away. In approximately eighty percent of instances, smoking was the primary cause of lung cancer in the individuals tested. It is possible for people who do not smoke to get lung cancer due to exposure to radon, secondhand smoking, air pollution, or other factors such as being exposed to asbestos at work, truck exhaust, or certain chemicals. Even persons who don't smoke have a chance of developing lung cancer. Biopsies, sputum cytology, tissue samples, imaging sets (x-rays, CT scans), and other diagnostic procedures are some of the tests that are used in the process of identifying malignant cells and eliminating other potential disorders. Pathologists with the necessary expertise are required to examine the tiny histopathological slides that were created after the biopsy in order to identify lung cancer and determine the various types and subtypes of the disease. In a short amount of time, it may be challenging for medical professionals, including physicians, to identify all of the many varieties of lung cancer. As a result of incorrect cancer diagnosis, patients are at a significant risk of receiving the incorrect therapy, which also has the potential to be deadly.

Machine learning (ML) is a subfield of artificial intelligence that enables computers to acquire knowledge without being explicitly taught. This is accomplished by providing them with data sets and a variety of tasks to complete. The majority of the authors of prior research publications considered the possibility of diagnosing and locating lung cancer via the utilization of CT scans and x-ray images in conjunction with machine learning techniques such as Support Vector Machine (SVM), Random Forest (RF), Bayesian Networks (BN), and Convolutional Neural Network (CNN). Additionally, histopathological images were used in a few of the investigations; however, these images are not as effective as histopathological images in distinguishing between images of carcinomas and images of tissue that does not contain cancer. The objective of this research is to evaluate the extent to which Support Vector Machine (SVM) and Convolutional Neural Network (CNN) are able to detect lung cancer via the use of tissue scan analyses. CNN, which is a kind of deep learning, has shown to be effective in activities involving the categorization of images. On the other hand, support vector machines (SVM) are a well-known machine learning technology that has been used during the process of solving classification issues. In a number of research, tools from the field of machine learning have been used to assist in the locate of individuals who have lung cancer. A machine learning technique that was built on a deep learning algorithm known as ResNet-50 was used by Zhang et al. (2019) in order to categorize images of lung cancer tissue into different categories. Ninety-two percent of the time, the research was correct. In a second piece of research, Li et al. (2020) used a machine learning technique that was founded on Support Vector Machine (SVM) in order to categorize images of lung cancer tissue into several categories. The research was accurate 89.6 percent of the time.

- **LITERATURE SURVEY**

S. Borkar and N. S. Nadkarni had the responsibility of writing it. It is possible to detect lung cancer with the use of CT scans since these scans provide a clear view of the tumor that is located inside the body and indicate how it is developing. The process of looking at the images on a computer screen may take a significant amount of time, increase the likelihood of making errors, and cause a delay in the detection of lung cancer. This is despite the fact that CT scans are superior to other imaging modalities.

M. Saric, M. Stella, and M. Sikora are the authors of this scholarly work. An examination of the tissue by a pathologist is the standard method for analyzing it; nevertheless, this method is not only time-consuming but also prone to errors. An automatic cancer area detection system would be of great assistance to the physician and would significantly accelerate the procedure. This article presents a completely automated method for detecting lung cancer in whole slide photographs of lung tissue samples. The method was developed by us. Image classification at the patch level is accomplished by the use of convolutional neural networks (CNN). It examines the effectiveness of two different CNN architectures, namely VGG and ResNet. According to the findings, the strategy that is based on CNN may assist medical professionals in detecting lung cancer.

C. Venkatesh and R. Vinoth are one of the authors of this piece.During the sorting phase, the kernels are included in the process. Moreover, the paper investigated the ways in which force uniformity might be used as a first step. In spite of the fact that it is not often used in CNN-based division procedures, it performed exceptionally effectively when paired with information expansion in order to differentiate between brain cancers in MRI images. The job is made more extensive by determining a number of characteristics that are associated with photographs. determined the characteristics of the cells and successfully located tumor cells in regions that had a high number of infections. Identifying the characteristics allows you to determine the severity of the sickness as well as the stage it is currently in. As part of the SVM classification process, the data that was acquired is used.

It was written by W. Zhao and Q. Wu.If individuals desire to survive, it is possible that it is better to detect lung cancer at an early stage. An innovative approach to use neural networks for the purpose of locating small cell lung cancer (SCLC) in computed tomography (CT) images is presented here in the form of the entropy degradation method (EDM). There is a possibility that this research may assist in the early detection of lung cancer. High-resolution computed tomography (CT) images of the lungs are provided by the National Cancer Institute. These images are used for testing and training purposes. A total of twelve CT images of the lungs were selected from the collection. Six of them depict healthy lungs, whereas the other six depict patients who have small cell lung cancer. The training of our model is accomplished by selecting five scans at random from each group. A test of the model is performed using the most recent two scans. We have a 77.8 percent success rate with our strategies.

- **EXISTING SYSTEM**

In the field of machine learning, Support Vector Machines (SVM) are a typical method that is used to categorize items into categories. Finding the hyperplane in the feature space that most effectively divides the classes is the method that support vector machines (SVM) function. The photos of lung cancer tissue have been sorted into several categories using a technique known as support vector machine (SVM). For the purpose of its input, SVM uses features that were meticulously crafted from images. The characteristics were selected on the basis of how effectively they differentiate between the groups. The support vector machine (SVM) algorithm then selects the hyperplane that separates the groups based on these characteristics. The existing support vector machine (SVM) technique not only requires various extraction processes, but it also requires more time to train the model, and it does not perform well with extremely big datasets. Additionally, it is not exact enough; the correct range is between 80 and 85 percent.

One of the issues with the existing system is that the support vector machine (SVM) requires distinct processes for feature extraction. It takes a long time to train SVM, which is one of the reasons why it does not perform well with a large amount of data. The accuracy of SVM ranges between 80% and 85%.

- **PROPOSED SYSTEM**

A form of deep learning system known as convolutional neural networks (CNNs) has shown that it has the potential to be effective in the classification of images. Convolutional layers are used by CNN in order to extract characteristics from captured images. The network acquires the characteristics as it goes through the training phase. When all of the steps that are entirely tied to the qualities have been completed, the final sorting will be finished. CNN has been used in the process of categorizing cell tissue images of lung cancer patients. In the photographs, CNN takes in the raw pixel counts that are there. Following this, the CNN software will educate itself on the essential characteristics as part of the sorting process. When it comes to the algorithms that are used for machine learning, CNNs are the ones that have received the most attention. CNNs have the ability to maintain spatial associations even while they are changing the images that they are given as input. It has previously been mentioned that radiography is heavily dependent on the way in which objects interact to one another in space. The point at where malignant and healthy lung tissue meet is an example of an important example. Other examples include the border of a bone where it meets a muscle. Through the use of Convolutional, Rectified Linear Unit (RELU), and Pooling Layers, CNN is able to transform an image that is composed of pixels. By doing so, the information is sent to a Fully Connected Layer, which then assigns the information to the category that has the best probability of being correct. This results in the class scores or odds being calculated. There is no need to do any further feature extraction procedures, which is a significant advantage. It is a pretty simple process that requires very little time and effort to complete. An accuracy of 95% to 98% is considered satisfactory.

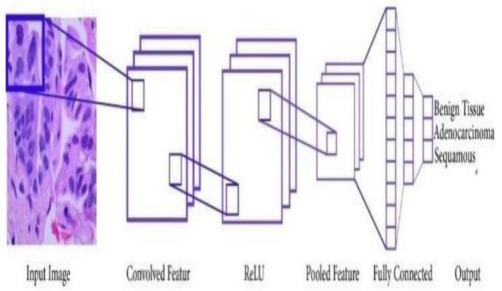
- **METHODOLOGY**

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

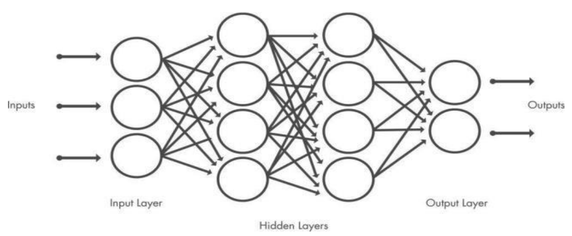
In the sections that follow, we will discuss the elements that comprise our proposed system, which are the following: gathering data, organizing data, training a model, testing the model, and producing predictions.

The histology photographs were obtained via the usage of the LC25000 Lung and Colon Histopathological Image Dataset, which was employed for the data collection process. The lung squamous cancer cells, the adenocarcinoma, and the benign tissue are the three forms of benign tissue that we investigate.

With regard to the formatting of the data, the collection consisted of RGB color histology photos saved as.jpeg files. In order to ensure that the photographs were suitable for the CNN operation, they were altered to ensure that they always had an aspect ratio of one to one and pixel sizes of (180, 180). For the purpose of accelerating the



convergence process, the values of each pixel in the picture were altered such that they were all between 0 and 1. For the purpose of increasing the quantity of images and making the data pattern more diverse, we used several strategies for capturing pictures, such as zooming in and out on photographs and flipping them both vertically and horizontally inside the frame. The neural network has a tendency to become too proficient at what it does when it is taught with a greater number of epochs but a smaller number of training data sets.



A liner stack of layers was used in the construction of the Convolutional Neural Networks (CNNs or ConvNets) that were used for the classification and recognition of still images. Training, testing, and prediction were all accomplished with its assistance.

On both the training and the test pictures, convolutional layers with fully connected layers, max pooling, and kernel filters were used. The item was placed in the appropriate category with the assistance of the softmax tool. During the training and testing of the model, Virtual Studio Code was used.

➢ ARCHITECTURE-

A Convolutional Neural Network, often known as a CNN, is a sort of Deep Learning that is capable of distinguishing between objects and features in an image. This is accomplished by assigning each and every one of these characteristics and objects relevant weights and biases that may be learnt. There is a significant reduction in the amount of pre-processing that is required for a ConvNet in comparison to other classification techniques. Convolutional neural networks are capable of learning these filters

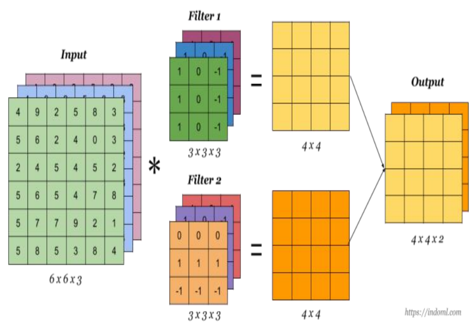


Fig 1. Convolutional layer

and features with sufficient training, but simple

methods need filters that are created by hand. It is not necessary for CNNs to have humans extract characteristics since they are able to learn the features on their own.

The data that CNNs provide for recognition is quite accurate. CNNs are capable of being retrained to do new types of recognition tasks,

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

which facilitates the addition of additional networks to those that are currently installed. In a convolutional neural network, there may be tens or even hundreds of layers, and each one of those layers could be trained to identify distinct aspects of an image. Every single training picture is subjected to filters at varying resolutions, and the output of every single convolved image is then taken into consideration for the subsequent step. It is possible that the filters will begin with the most fundamental ones, such as color and edges, and then go to the traits that distinguish the object from others. The input layer, the output layer, and multiple hidden layers in between are the components that make up a CNN. In this regard, it is identical to other neural networks. These layers perform operations on the data in order to identify characteristics that are specific to the data. The activation layer, also known as the ReLU layer, the convolution layer, and the pooling layer are the three layers that are used the most often. If you apply the appropriate filters, a ConvNet could be able to accurately capture the temporal and spatial connections that are present in an image. Because there are fewer criteria that need to be considered and weights are repeated, the design performs a better job of fitting the photo collection altogether. In other words, the network might be equipped with the knowledge necessary to better comprehend how intricate the image is. It is the responsibility of the ConvNet to reduce the size of the images while preserving the specific components that are required for effective forecasting.

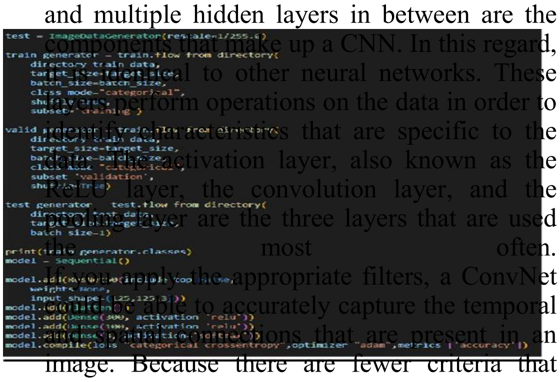


Fig 2. Architecture

Convolutional layer filters are used in order to create a feature map that is dependant on the location. A feature map is created by the filters in the convolutional layers, and the location of the map is determined by the filters. In the event that an item in a photograph has shifted slightly, for instance, the convolutional layer may not be able to locate it. Next, the feature map will make a note of the precise locations of the features that are included inside the input. In pooling layers, the concept of "translational invariance" makes it feasible for the CNN to maintain its original state whenever translations are performed. Even after the input has been translated, the CNN is still able to detect the characteristics that are present in the input by virtue of this.

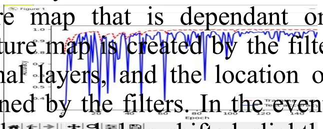
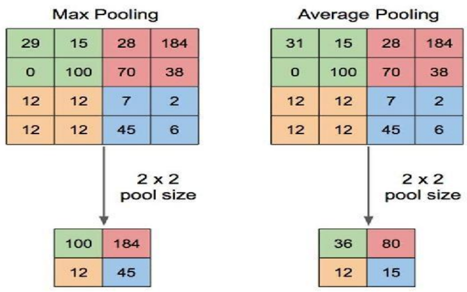


Fig 3. Pooling layer

- **WORKING**



The model was completely trained using CNN layers, and its sizes were then adjusted.

Fig 4. Prediction code

- **Training and Validation Accuracy:**
  When compared to the total number of epochs, the graph illustrates the degree of accuracy that the model achieved in both the tests and the validations. In order to train the model, layers were used.

Fig 5. Training and Testing Accuracy

RESULTS

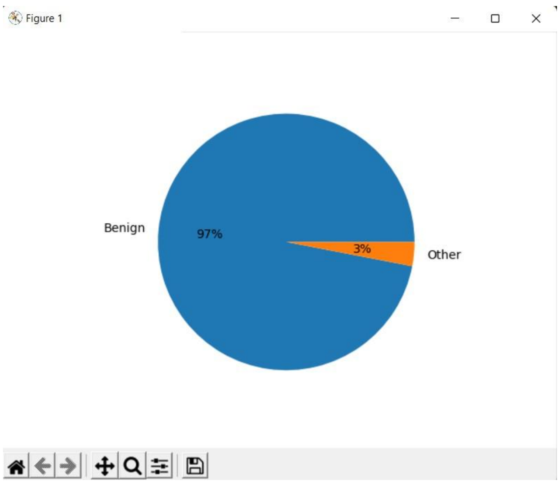
Based on the findings that we obtained, it is clear that the approach that was recommended is quite effective in locating lung cancer cells. When everything was

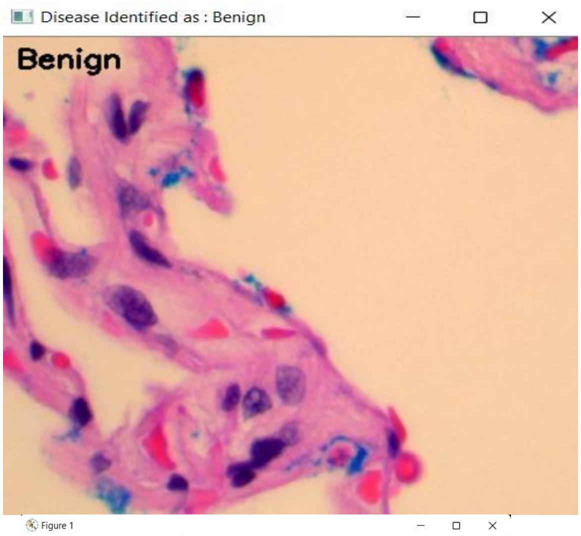
ISSN: 2057-5688

said and done, the CNN model was successful in answering 93% of the questions on the test set. It had a sensitivity of 89% and a specificity of 97%. Furthermore, these findings demonstrate that CNNs are capable of reliably locating lung cancer cells on histopathological images.

Fig 6. Benign

Fig 7. Squamous Cell Carcinoma





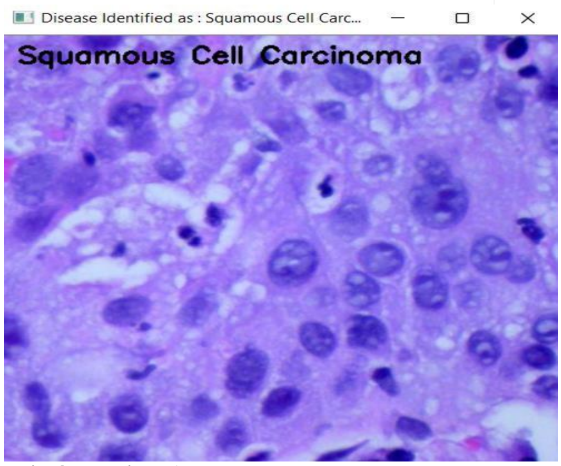


Fig 8. Benign Accuracy

Fig 9. Squamous Cell Carcinoma Accuracy

- **CONCLUSION**
- In this particular investigation, histopathological scans were used

in order to detect lung cancer. Through the use of a convolutional neural network (CNN), an image of three different forms of cancer—beneficial, adenocarcinoma, and squamous cell carcinoma—was organized into groups. During the validation phase, the model was accurate 96.20% of the time, while during the training phase, it was accurate 96.11% of the time. the approach that is based on deep learning and is completely automated for locating lung cancer in images of entire histology slides. When the accuracy and area under the curve (AUC) of patch classification are compared with VGG16 and ResNet50, its performance is superior. It has been shown via the findings that convolutional neural networks are capable of accurately classifying lung cancer pictures that include the whole slide. Nevertheless, further research is required in order to make this more accurate. In further research, the size of the training set will be increased, and the photos will also be enhanced and made more consistent. On top of that, we are not going to make use of weights that have previously been trained on ImageNet; rather, we are going to attempt training from scratch. The strategy that is based on CNN has been shown to be effective in locating lung cancer in tissue images. There is a significant improvement in accuracy compared to what was discovered in past investigations that used the same information. If you make adjustments to the hyperparameters, you may be able to increase the size of the sample and improve the performance of the model. It is possible that the approach that has been presented will assist in the early identification of cancer, which is essential for a successful result of therapy, and it may be used in conjunction with a variety of other cancer detection methods.

REFERENCES

1) G. A. Silvestri, et al. "Noninvasive staging of non-small cell lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition)." Chest vol. 132, 3 Suppl (2007): 178S-201S. doi:10.1378/chest.07-1360.

2) W. D. Travis, et al. "International association for the study of lung cancer/American thoracic society/European respiratory society international multidisciplinary classification of lung adenocarcinoma." Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer vol. 6, 2 (2011): 244-85. doi:10.1097/JTO.0b013e318206a221

3) L. G. Collins., C. Haines, R. Perkel & R. E. Enck. "Lung cancer: diagnosis and management." American family physician vol. 75, 1 (2007): 56-63.

4) K. Yu, C. Zhang, G. Berry, et al. "Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features." Nat Commun 7, 12474 (2016), doi: 10.1038/ncomms12474

5) D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis," 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), Ras Al Khaimah, 2016, pp. 1-4, doi:

10.1109/ICEDSA.2016.7818560.

6) E.D. Michie, D.J. Spiegelhalter, and C.C. Taylor, "Machine Learning, Neural and Statistical Classification," Proceeding, 1994.

7) W. Ausawalaithong, A. Thirach, S. Marukatat, and T. Wilaiprasitporn, "Automatic Lung Cancer Prediction from Chest X-ray Images Using the Deep Learning Approach," 2018 11th Biomedical Engineering International Conference (BMEiCON), Chiang Mai, 2018, pp. 1-5, doi: 10.1109/BMEiCON.2018.8609997.

8) T. Atsushi, T. Tetsuya, K. Yuka, and F. Hiroshi. (, 2017). “Automated Classification of Lung Cancer Types from Cytological Images Using Deep Convolutional Neural Networks”. BioMed Research International. 2017. 1-6. 10.1155/2017/4067832.

9) W. Rahane, H. Dalvi, Y. Magar, A. Kalane and S. Jondhale, "Lung Cancer Detection Using Image Processing and Machine Learning HealthCare," 2018

10) International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, 2018, pp. 1-5, doi: 10.1109/ICCTCT.2018.8551008.

11) S. Sasikala, M. Bharathi, B. R. Sowmiya. "Lung Cancer Detection and Classification Using Deep CNN." (2019).

12) SRS Chakravarthy and H. Rajaguru. "Lung Cancer Detection using Probabilistic Neural Network with modified Crow-Search Algorithm." Asian Pacific Journal of Cancer Prevention, 20, 7, 2019, 2159-2166, doi: 10.31557/APJCP.2019.20.7.2159.

13) AA. Borkowski, MM. Bui, LB. Thomas, CP. Wilson, LA. DeLand, SM. Mastorides. "Lung and Colon Cancer Histopathological Image Dataset." (LC25000). ArXiv: 1912.12142v1 [eess.IV], 2019.