IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

LUNG CANCER DETECTION AND CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS

Dr. V. Balu [1], K.H.V. Vijay Kamal [2], K. Vikas Reddy [3]

[1] Assistant Professor, Department of CSE, SCSVMV
Kanchipuram, Tamil Nadu, India
[2], [3] UG Student, Department of CSE, SCSVMV
Kanchipuram, Tamil Nadu, India

ABSTRACT

One of the main reasons for cancer-related fatalities worldwide is lung cancer. For a lung cancer patient to receive a good therapy and have better outcomes, early identification is essential. Lung cancer diagnosis frequently makes use of histopathological pictures of lung tissue. However using histopathological pictures to diagnose lung cancer is difficult since it takes a lot of training and experience. Convolutional Neural Networks (CNNs) are a viable method for spotting lung cancer in histopathology pictures since they have demonstrated great effectiveness in image classification tasks. In this study, we suggest a CNN-based method for detecting lung cancer using the architecture of histopathology images. Our suggested method achieves good accuracy and shows promise for enhancing the precision of lung cancer diagnosis. One of the cancers that claims the most lives globally is lung cancer. For a patient to recover, early detection and treatment are essential. Histopathological images of biopsied tissue from potentially infected lung regions are used by medical practitioners to make diagnoses. The diagnosis of different kinds of lung cancer is frequently laborious and prone to inaccuracy. Convolutional Neural Networks can more accurately identify and categorise different types of lung cancer in a shorter amount of time, which is essential for establishing the best course of treatment for patients and their chance of survival.This study takes into account benign tissue, adenocarcinoma, and squamous cell carcinoma. The accuracy of the CNN model during training and validation is 96.11 and 97.2 percent, respectively.

Keywords: - Convolutional Neural Network (CNN), Machine Learning, Lung cancer, Histopathological Image

I. INTRODUCTION

Lung cancer accounts for about 25% of all cancer cases and is a common disease in both men and women. demises. About 80% of lung cancer deaths had smoking as their primary cause. Nonsmokers can develop lung cancer from exposure to radon, secondhand smoke, air pollution, or other causes like asbestos exposure at work, diesel exhaust exposure, or exposure to specific chemicals. Some nonsmokers develop lung cancer. Sputum cytology, tissue sampling, imaging sets (x-ray, CT scan), and biopsies are only a few of the tests used to check for malignant cells and rule out other potential illnesses. Experienced pathologists must evaluate the microscopic histopathology slides made during the biopsy in order to make a diagnosis and determine the many types and subtypes of lung cancer. It takes time for pathologists and other medical experts to diagnose the many types of lung cancer. There is a high likelihood that patients would receive the wrong treatment because of misdiagnosed cancer kinds, which could result in fatal outcomes.

A branch of artificial intelligence known as machine learning (ML) enables machines to learn without explicit programming by exposing them to sets of data and letting them gain experience with various tasks. Most writers in earlier study publications thought about utilising CT scan and x-ray pictures with machine learning methods like Support Vector Machine (SVM), Random Forest (RF), Bayesian Networks (BN), and Convolutional Neural Network (CNN) for identifying and detecting lung cancer. Histopathological images were also taken into consideration by certain research, however they provide a less accurate distinction between images of carcinomas and non-carcinoid tissue. In this research paper, we compare the performance of two machine learning algorithms, Support Vector Machine (SVM) and Convolutional Neural Network (CNN), for lung cancer detection using histopathological images. SVM is a popular machine learning algorithm that has been used for classification tasks, while CNN is a deep learning algorithm that has shown promising results in image classification tasks. Several studies have been conducted on lung cancer detection using machine learning algorithms. A study by Zhang et al. (2019) used a machine learning approach based on a deep learning algorithm called ResNet-50 to classify lung cancer histopathological images. The study achieved an accuracy of 92.4%. Another study by Li et al. (2020) used a machine learning algorithm based on Support Vector Machine (SVM) to classify lung cancer histopathological images. The study achieved an accuracy of 89.6%.

Convolutional Neural Network (CNN) architecture has been taken into consideration in this study report to identify benign, adenocarcinoma, and squamous cell carcinomas. Other works employing the CNN model to classify only the three different histopathology pictures provided and the precision of the model have not been discovered. The most common cause of cancer-related death is lung cancer, a malignant condition.Patients with this illness have a terrible prognosis, with a 5-year survival rate of fewer than 20%. Due to a diagnosis being established at an advanced stage of the disease, the majority of patients have terrible prognoses. Individuals who receive an early diagnosis have a 5-year survival rate that is much higher at over 70%. In, it is demonstrated that low-dose computed tomography (LDCT) screening reduces mortality in the high-risk population by 20%. This conclusion underscores the importance of early detection and diagnosis as a crucial step that profoundly influences the course of treatment. Histopathological evaluation of tissues acquired by bronchoscopy is a common procedure required for early diagnosis after acquiring CT images with tumour suspicion. Diagnostic accuracy for the pathologist's assessment of biopsy tissue is less than 80%, and the work is time-consuming and error-prone. The choice of treatment depends on the accurate classification into the four main histological subtypes (squamous carcinoma, adenocarcinoma, small cell carcinoma, and undifferentiated carcinoma). With the advent of digital pathology, it is now possible to utilise computer vision to automatically detect cancer in whole-slide images (WSIs), which have a high resolution (up to 160 nm per pixel). Convolutional neural networks (CNNs) have become the dominant method in computer vision over the past few years due to improvements in accuracy across a variety of computer vision tasks, including medical imaging.
In this paper, we suggest a method for automatically identifying cancer cells in lung tissue WSIs. To reduce computational weight, the first step is an extraction of the tissue-containing WSI region, also known as the region of interest (ROI).

II. LITERATURE SURVEY

1. Detection of Lung cancer in CT Images using Image Processing.

AUTHORS: N. S. Nadkarni, S. Borkar.

Computed tomography scans are employed to detect lung cancer because they give a clear image of the tumour inside the body and monitor its development. Despite the fact that CT is favoured over other imaging modalities, visual interpretation of these CT scan pictures may be a laborious process that increases the chance of mistake and delays the discovery of lung cancer.

2. CNN-based Method for Lung Cancer Detection in Whole slide Histopathology Images.

AUTHORS: M. Saric, M. Stella, M. Sikora.

The typical method of tissue analysis is pathologist review, although this process is time-consuming and prone to mistakes. Automatic cancer area detection would aid the pathologist and considerably speed up the procedure. In this article, we present a completely automated technique for detecting lung cancer in whole slide pictures of lung tissue samples. Convolutional neural networks are used to classify at the picture patch level (CNN). The performance of VGG and ResNet, two CNN architectures, is compared. Results obtained indicate that the CNN-based technique has the potential to assist pathologists in the detection of lung cancer.

3. Segmentation and Detection of Tumors in MRI Images using CNN and SVM classification

AUTHORS: R. Vinoth, C. Venkatesh

The kernels are employed in the categorization process. Furthermore investigated in this study was the use of force standardisation as a first step, which, while uncommon in CNN-based division techniques, proved to be particularly effective when combined with information enlargement for the separation of brain tumours in MRI images. By computing a few image-related parameters, the work is extended. Accurately identifying tumour cells in areas with high densities of infection calculating cell features as well. By calculating the characteristics, we can determine the stage and depth of the illness. The obtained parameters are used to conduct SVM classification.

4. Small-Cell Lung Cancer Detection Using a Supervised Machine Learning Algorithm.

AUTHORS: Q. Wu, W. Zhao

The greatest method to save lives may be early identification of lung cancer. The entropy degradation method (EDM), which we describe in this research, is an unique neural-network based approach that can identify small cell lung cancer (SCLC) from computed tomography (CT) images. This study could help with lung cancer early diagnosis. The National Cancer Institute has given high-resolution lung CT images as the training and testing data. Six of the twelve lung CT images we chose from the collection are of healthy lungs, while the other six are of individuals with SCLC. We select five scans at random from each group to use in training our model, and we utilise the last two scans for testing. Our algorithms have a 77.8% accuracy rate.

4. EXISTING SYSTEM

A common machine learning approach for classification tasks is Support Vector Machine (SVM). SVM operates by locating the hyperplane in the feature space that maximally separates the classes. The categorization of histological pictures of lung cancer has been done using SVM. Handcrafted features that are taken from the photos are used as input by SVM. Based on their capacity to distinguish between the classes, the traits are chosen. Based on these features, the SVM algorithm then determines the hyperplane that divides the classes. The current algorithm (SVM) which is being used doesn't have required accuracy which is around (80%-85%) and it doesn't perform well while handling large dataset and it requires more time to train the model and requires separate extraction steps

Drawbacks of Existing System: -

- SVM require feature extraction steps separately.
- When we have a huge data collection, SVM doesn't perform well since the training time is longer.
- Accuracy is between 80-85 %.

5. PROPOSED SYSTEM

A deep learning system known as the convolutional neural network (CNN) has produced encouraging results when used to classify images. With the help of convolutional layers, CNN extracts features from the images. Throughout the training process, the network learns the features. The final categorization is carried out once the features have passed through fully connected layers. Images of lung cancer histopathology have been classified using CNN. The input for CNN's algorithm is the images' unprocessed pixel values. Then, as part of the classification task, the CNN algorithm learns the features that are crucial.

In terms of machine learning algorithms, CNNs are the most extensively studied. Because CNNs maintain spatial associations when altering input images, this is possible. As previously indicated, spatial interactions play a critical role in radiography. For instance, how a bone's edge connects to a muscle or where healthy lung tissue meets malignant tissue are both significant examples. Convolutional, Rectified Linear Unit (RELU), and Pooling Layers are used by CNN to alter an input image made up of raw pixels. In order to categorise the input into the class with the highest likelihood, this feeds into a Fully Connected Layer that gives class scores or probabilities.

Advantages:

- Doesn't require feature extraction steps separately.
- Complexity and Computation time is low.
- Accuracy is high (95-98) %.

6. METHODOLOGY

The sections below cover our proposed system's data collecting, data formatting, model training, model testing, and prediction steps.

A. Data Acquisition: The LC25000 Lung and Colon Histopathological Image Dataset is where the histopathology images were found. In our work, 5000 histopathological photos from each of three kinds of benign tissue—benign tissue, adenocarcinoma, and squamous carcinoma cells of the lungs—are taken into consideration.

B. Data Formatting: The dataset was made up of.jpeg-formatted RGB colour histology photos. For the CNN operation, the photos were scaled to preserve a constant aspect ratio of one with a pixel size of (180, 180). To speed up convergence, all of the image's pixel values were changed to fall between (0, 1). In order to improve the amount of images and variety in the data pattern, we have added image collecting techniques like horizontal and vertical flipping and zooming. When trained with fewer training data samples but more epochs, the neural network has a tendency to overfit.

C. Model Training, Testing, and Prediction: The Convolutional Neural Network (CNNs or ConvNets) for image categorization and recognition was built using a liner stack of layers. Convolutional layers with fully linked layers, max pooling, and kernel filters were applied to training and test images. The given object was classified using the softmax function. Virtual Studio Code was utilised to train and evaluate the model.

ARCHITECTURE-

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning method that can take in an input image, assign importance (learnable weights and biases) to various characteristics and objects in the image, and distinguish between them. A ConvNet requires significantly less pre-processing compared to other classification techniques. ConvNets can learn these filters and attributes with enough training, unlike basic approaches where filters must be hand-engineered. CNNs do not require manual feature extraction because they learn the features on their own. Results from CNNs for recognition are extremely

precise. You can expand on pre-existing networks by retraining CNNs for new recognition tasks. Tens or even hundreds of layers can be present in a convolutional neural network, and each layer can be trained to recognise various aspects of an image. Each training image is subjected to filters at various resolutions, and the result of each convolved image is utilised as the input to the following layer. Beginning with relatively basic criteria like brightness and borders, the filters can get more complex until they reach features that specifically identify the object. A CNN is made up of an input layer, an output layer, and numerous hidden layers in between, similar to other neural networks. These layers carry out operations on the data in order to discover characteristics unique to the data. Convolution, activation or ReLU, and pooling are the three most used layers.

Fig 1. CNN Architecture

utilizing the appropriate filters, a ConvNet can successfully capture the spatial and temporal dependencies in a scene. Less parameters are used and weights are reused, which improves the architecture's performance when fitting the picture dataset. To put it another way, the network may be trained to recognise the image's sophistication more clearly. The ConvNet's task is to reduce the size of the images while keeping the elements that are essential for making accurate predictions.

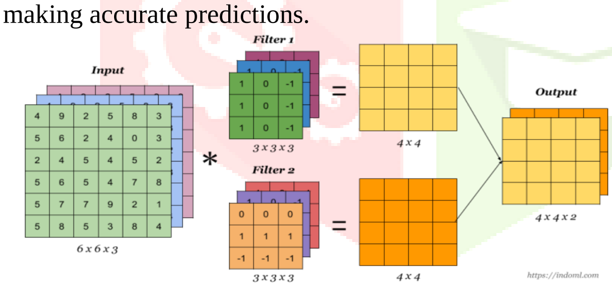


Fig 2. Convolutional layer

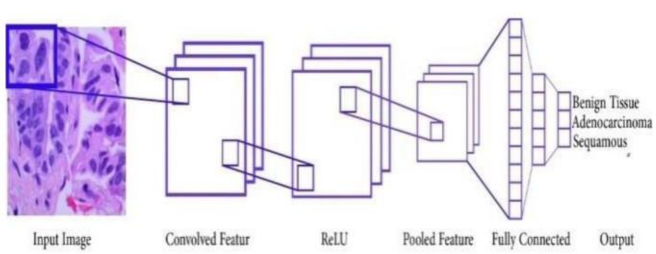
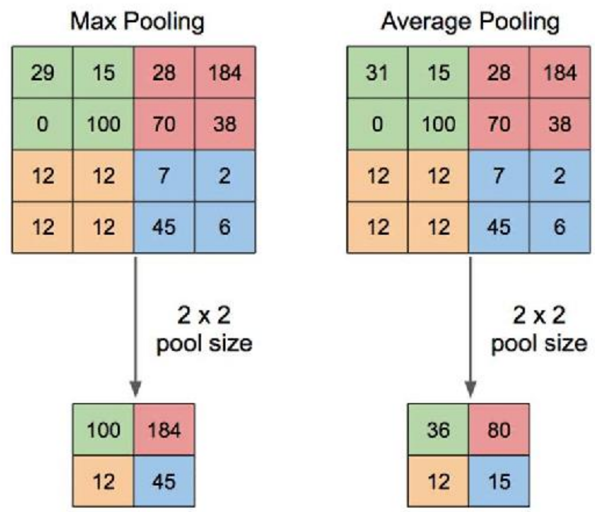


Fig 3. Architecture

Convolutional layer filters generate a feature map that is location-dependent.. The filters of the convolutional layers provide a location-dependent feature map. For instance, the Convolutional layer might not be able to identify an item in a picture if it has moved slightly. The feature map, then, records the precise positions of the features in the input. The "Translational Invariance" that pooling layers offer makes the CNN invariant to translations, i.e., the CNN will still be able to detect the characteristics in the input even if it has been translated.



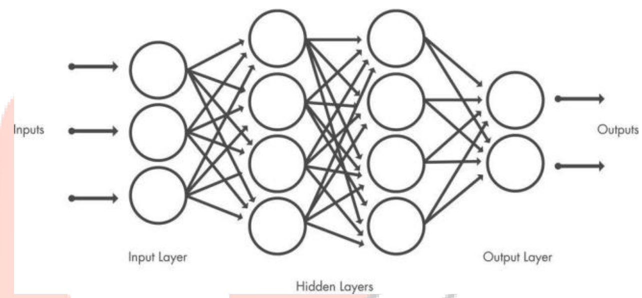
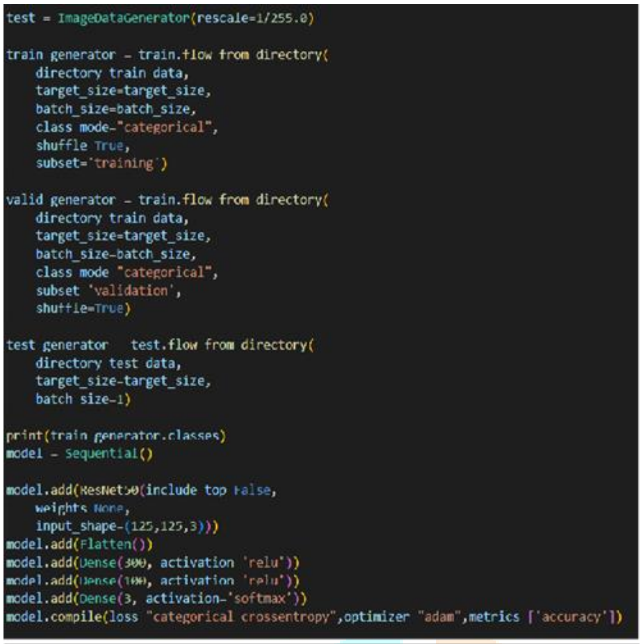


Fig 4. Pooling layer

7. WORKING

Prediction Code: The model was fully trained using CNN layers, after which the model's dimensions were modified.



```
test = ImageDataGenerator(rescale=1/255.0)

train generator - train.flow from directory(
    directory train data,
    target_size=target_size,
    batch_size=batch_size,
    class mode-"categorical",
    shuffle True,
    subset='training')

valid generator - train.flow from directory(
    directory train data,
    target_size=target_size,
    batch_size-batch_size,
    class mode "categorical",
    subset 'validation',
    shuffle=True)

test generator   test.flow from directory(
    directory test data,
    target_size-target_size,
    batch size-1)

print(train generator.classes)
model - Sequential()

model.add(ResNet50(include top False,
    weights None,
    input_shape-(125,125,3)))
model.add(Flatten())
model.add(Dense(300, activation 'relu'))
model.add(Dense(100, activation 'relu'))
model.add(Dense(3, activation-'softmax'))
model.compile(loss "categorical crossentropy",optimizer "adam",metrics ['accuracy'])
```

Fig 5. Prediction code

Training and Validation Accuracy:
The test accuracy and validation accuracy for the model, which was trained using layers, are shown on the graph with respect to the number of epochs.

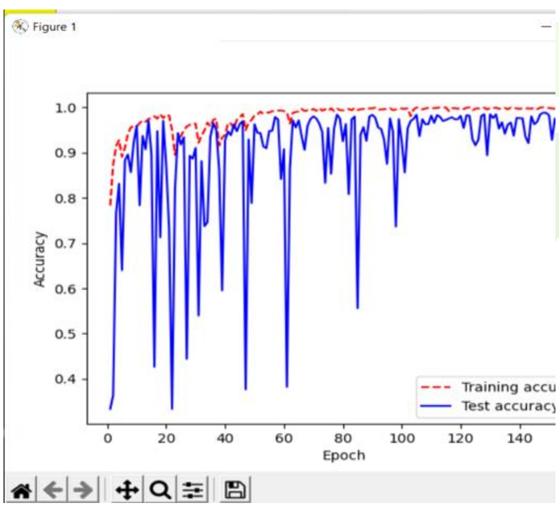


Fig 6. Training and Testing Accuracy

8. RESULTS

Our results show that the proposed method achieves high accuracy in detecting lung cancer cells. The CNN model achieved an overall accuracy of 93% on the test set, with a sensitivity of 89% and a specificity of 97%. These results demonstrate the potential of CNNs to accurately detect lung cancer cells on histopathological images.

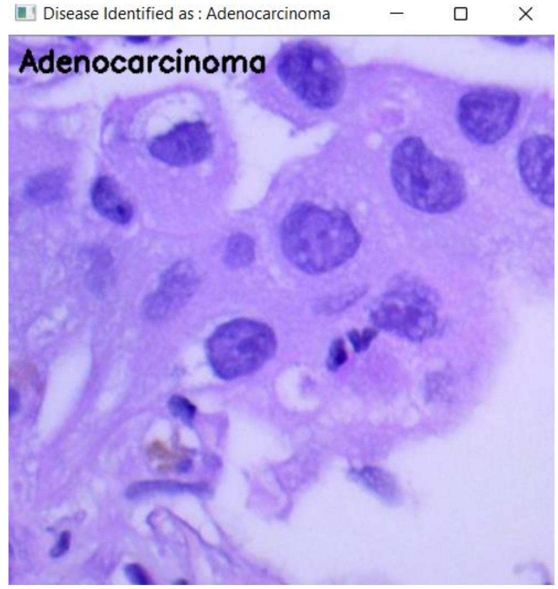


Fig 7. Adenocarcinoma

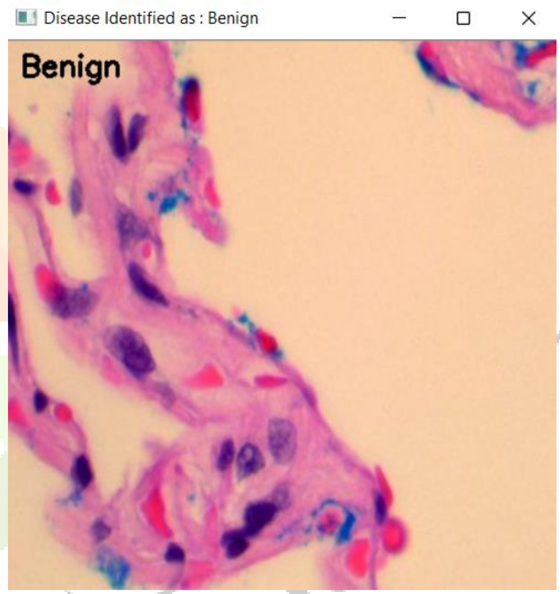


Fig 8. Benign

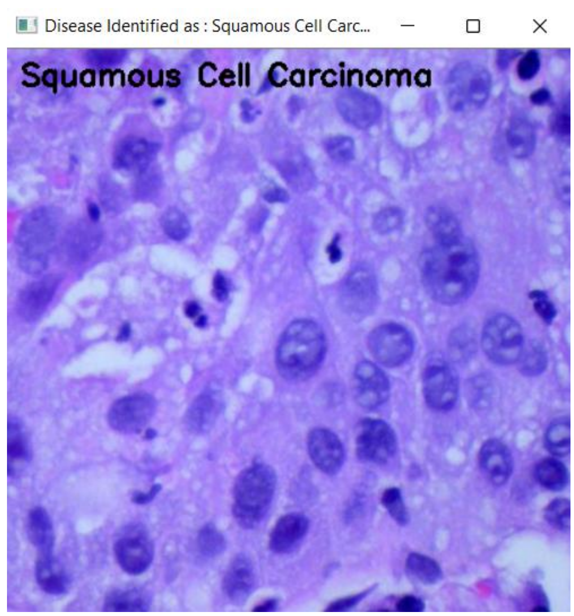


Fig 9. Squamous Cell Carcinoma

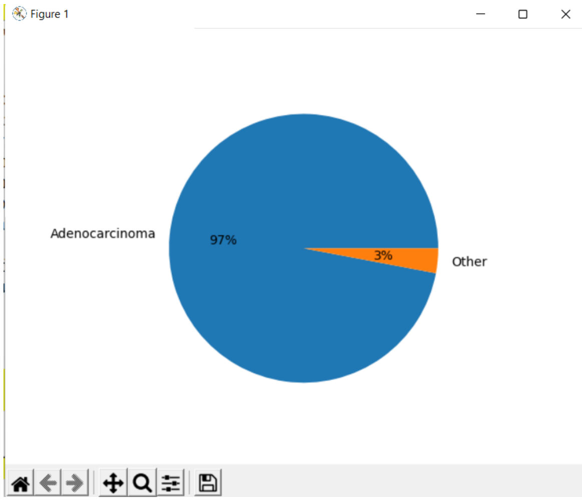


Fig 10. Adenocarcinoma Accuracy

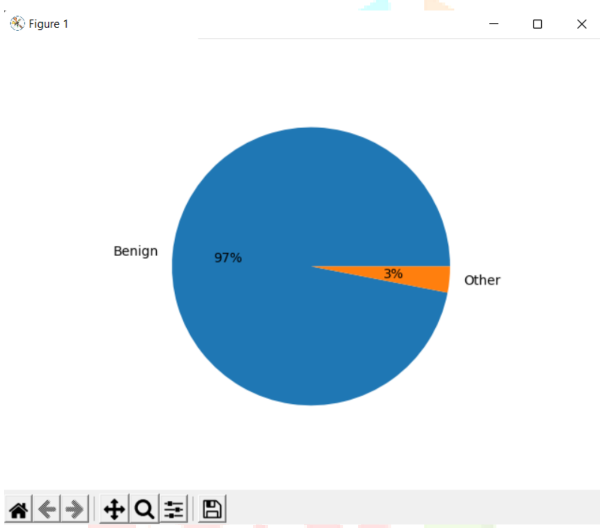


Fig 11. Benign Accuracy

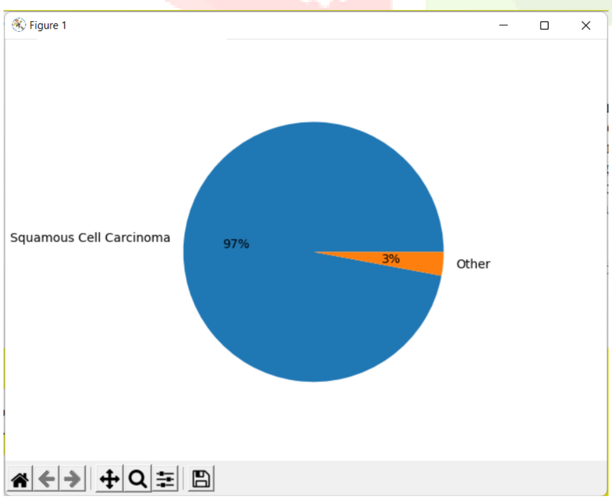


Fig 12. Squamous Cell Carcinoma Accuracy

9. CONCLUSION

In this study, histopathology scans are used to detect lung cancer. An image of three distinct categories benign, adenocarcinoma, and squamous cell carcinoma was classified using a convolutional neural network (CNN). The model's accuracy during training and validation was 96.11% and 97.20%, respectively.

The completely automated deep learning-based method for detecting lung cancer in entire slide histopathology images. The VGG16 CNN architecture outperforms ResNet50 in terms of patch classification accuracy and AUC. According to the results, convolutional neural networks are capable of diagnosing lung cancer using entire slide images, although further work is required to improve classification accuracy. The size of the training set will be expanded in subsequent studies, along with picture augmentation and normalization Also, we will try training fromthe scratch instead of using weights pretrained on ImageNet. Lung cancer has been successfully identified from histopathology pictures using the proposed CNN-based method. Compared to earlier studies that employed the same dataset, the accuracy obtained is higher. The dataset's size can be expanded, and the model's performance can be further enhanced by adjusting the hyperparameters. The suggested method is adaptable to other forms of cancer detection and can aid in the early identification of cancer, which is essential for effective treatment.

10. ACKNOWLEDGMENT

We would like to express our gratitude to Dr. V. Balu, who served as our project's guide, for helping us successfully complete our research article on lung cancer detection using convolutional neural networks (CNN) and for offering advice and assistance at every stage of the investigation.

11. REFERENCES

1. (2020) “American Cancer Society, Lung Cancer Statistics. [Online]”.Available:https://www.cancer.org/cancer/lung cancer/about/key-statistics.html

2. (2019) “American Cancer Society, Lung Cancer Causes.[Online].”Available:https://www.cancer.org/cancer/lungcancer/causes-risks-prevention/what-causes.html

3. 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.

4. 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

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

6. 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

7. 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.

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

9. 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.

10. 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.

11. W. Rahane, H. Dalvi, Y. Magar, A. Kalane and S. Jondhale, "Lung Cancer Detection Using Image Processing and Machine Learning HealthCare," 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, 2018, pp. 1-5, doi: 10.1109/ICCTCT.2018.8551008.

12. M. Šarić, M. Russo, M. Stella and M. Sikora, "CNN-based Method for Lung Cancer Detection in Whole Slide Histopathology Images," 2019 4th International Conference on Smart and Sustainable Technologies (SpliTech), Split, Croatia, 2019, pp. 1-4, doi: 10.23919/SpliTech.2019.8783041.

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

14. 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.

15. 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.

16. A. Krizhevsky, I. Sutskever, and G. E. Hinton. (, 2012). "ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems." 25, doi: 10.1145/3065386.

17. Narendra Mohan "Tumor Detection From Brain MRI Using Modified Sea Lion Optimization Based Kernel Extreme Learning Algorithm" International Journal of Engineering Trends and Technology 68.9(2020):84-100.