Data Science & Analytics

Lung Cancer Detection

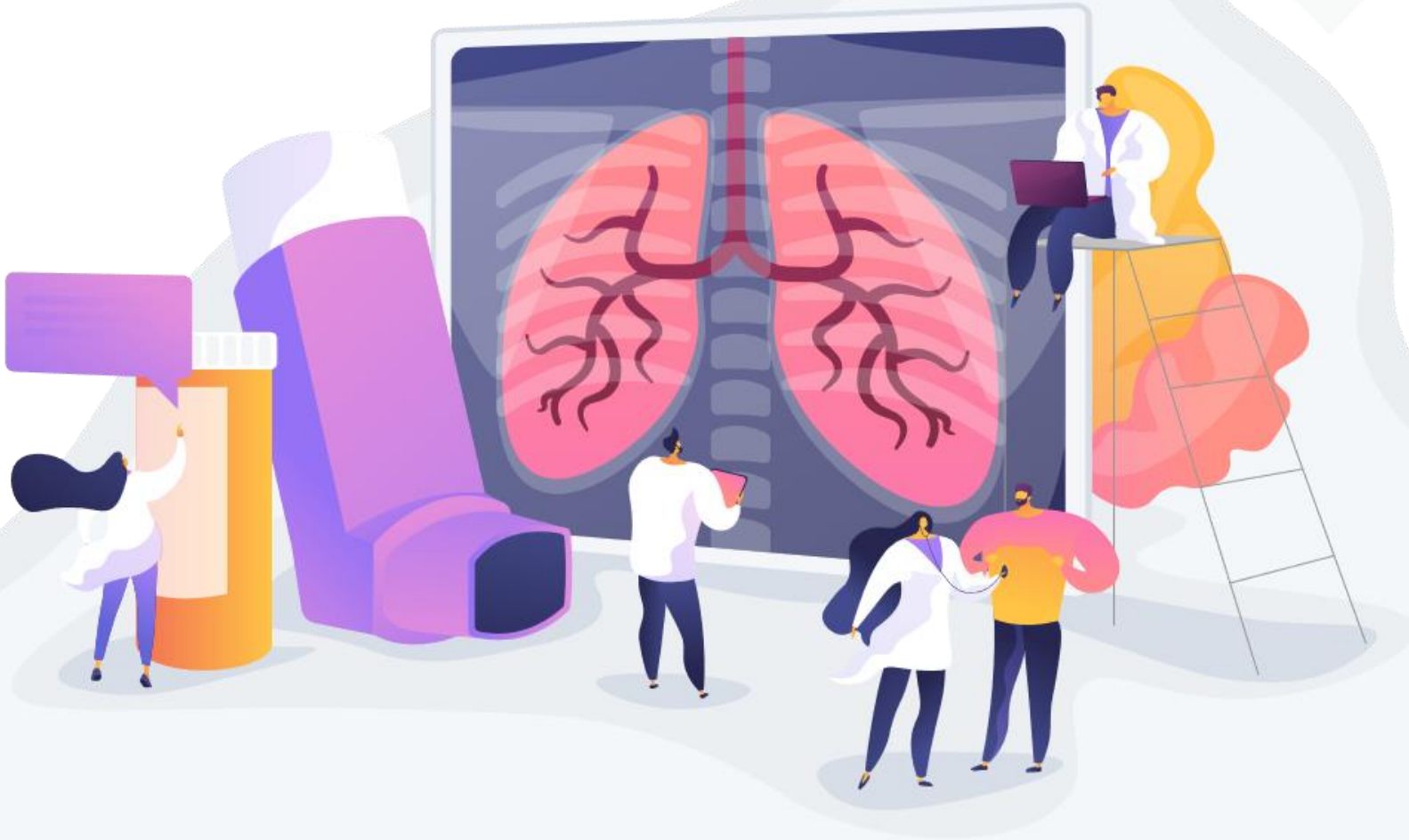
www.geninvo.com

North America | UK | Europe | India

Lung Cancer Detection

Machine Learning methodologies for

Automatic Segmentation and Diagnosis



Business Need

- Lung cancer is the leading cause of cancer related deaths in United States, Europe and many Asian countries. Owing to the low survival rates among lung cancer patients, early-stage detection and treatments are necessary.
- Recently other than conventional radiography, Positron emission tomography and Computed tomography (PET/CT) examination is being used as a cancer screening tool in some countries and is useful for the early detection of lung cancer. However, it's primary disadvantage is that it generates about 1000 slice images per scan, so the diagnosis and interpretation of a large amount of PET/CT images becomes cumbersome and a technique for obtaining high diagnostic accuracy with little effort is required.
- Computer-aided detection (CAD) is set of computational techniques developed primarily to serve as a second reader to support the radiologist's diagnosis and assist in evaluating a large number of images to identify pulmonary nodules and arrive at the diagnosis. A fully integrated CAD system based on Deep learning methodologies of classification and segmentation could be beneficial for practical applications of future medical imaging systems.

Challenges

Lung Even with many promising results from deep learning based image processing studies there are certain challenges to be considered when developing a practical application:

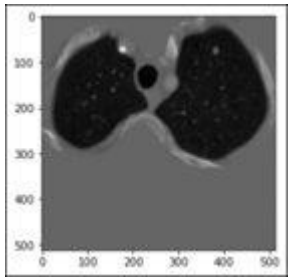
- Extensive cooperation between SMEs / radiologists and data scientists is required. This will help programmers to develop the deep learning based CAD system more in line with clinical needs and thereby improving diagnosis efficiency and accuracy.
- Medical images are different from natural images and are based on different physical principles. The quantification of the images is also complicated by limited resolution, lack of contrast, noise and presence of artifacts which creates a major problem in data preprocessing. Computational strategies need to take this variability into account so the application is sufficiently robust to perform well under a variety of conditions.
- Medical images are typically 3D tomographic images and deep learning algorithms need to be designed to process spatial relationships in all three dimensions of the 3D image. The training of such algorithms requires a large amount of dataset support which may also lead to tendency of overfitting. Robust evaluation strategies need to be implemented to mitigate overfitting in models.

Solution Methodology

- The solution developed for automating lung cancer detection in CT scans is a two-stage framework.
- The first stage involves semantic segmentation of the lung region from CT images to isolate the lungs and localize the nodules in the input images. The detection of nodules within the processed CT scan image indicates the possibility of cancer. The second stage involves generating specific features from the localized nodules in order to train a classifier for detecting cancer.
- The data preprocessing step in solution pipeline includes image transformation techniques for generating region of interest (ROI) maps. Thresholding is one such technique that removes non-lung region from the image which do not apply equally to all datasets thus reducing the possibility of false positives outside region of interest.
- The automatic segmentation of pulmonary nodules is done by U-Net model which is a fully convolutional network (FCN) composed of encoder-decoder architecture. The network ingests CT scan image and outputs a binary mask that highlights nodule locations.
- The final step of solution pipeline involves the Lung Cancer Detection model which is a classifier trained on the features extracted from CT scan image data and nodule masks.

Solution Pipeline

CT Scan Image

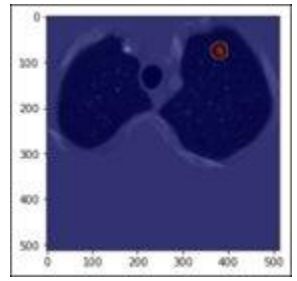


Preprocessing and ROI Isolation

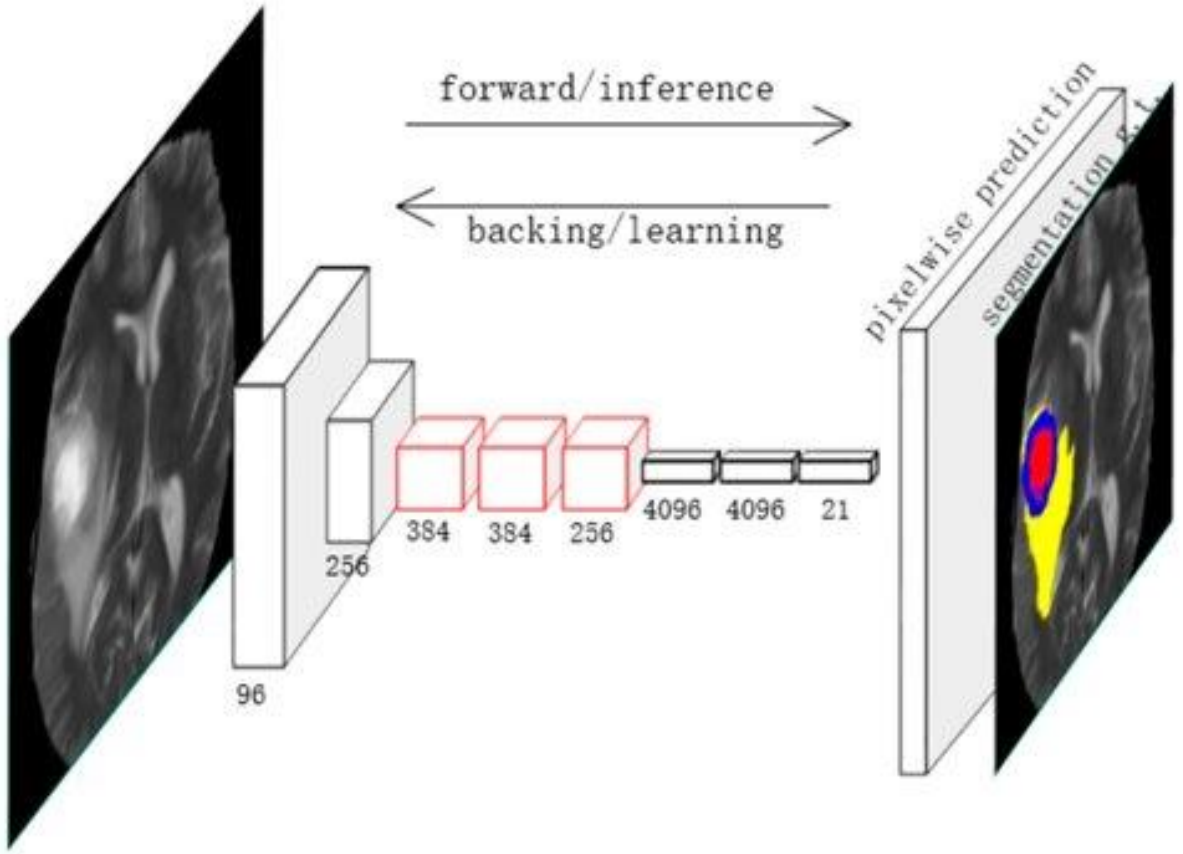
Segmentation and Nodule Detection

Inference

Cancer Detection

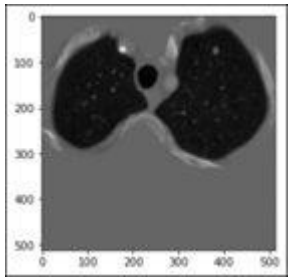


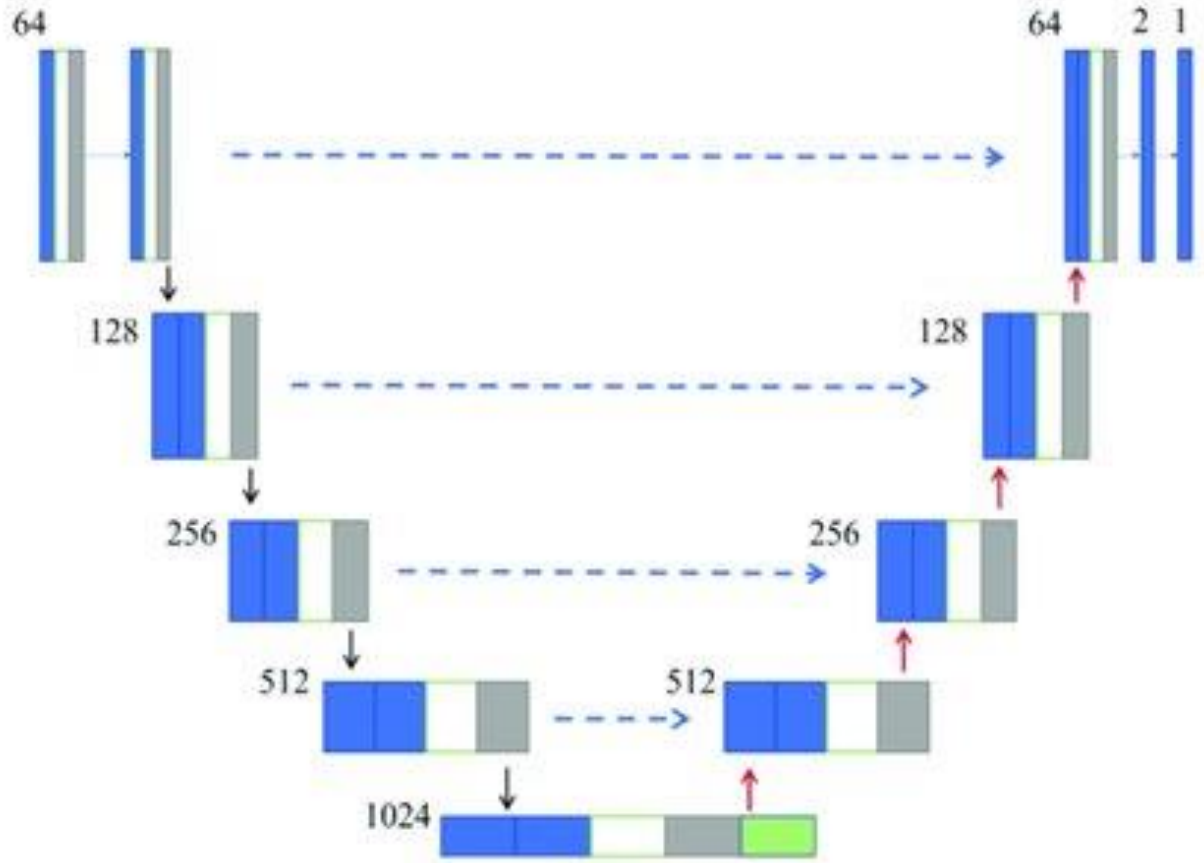
Example of FCN architecture for Image Segmentation



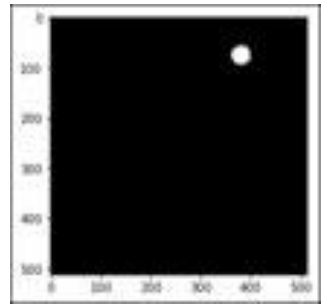
CT Scan Image Segmentation with U-Net

CT Scan Image
(512 x 512 x 1)





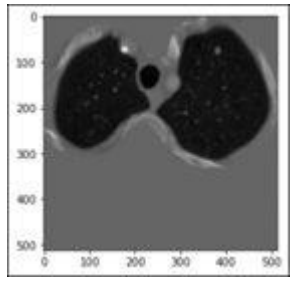
Nodule Mask
(512 x 512 x 1)



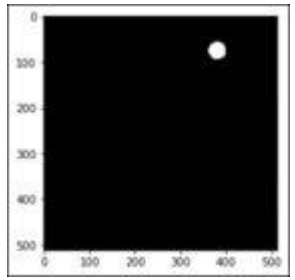
Con2D
Batch normalization
Leaky ReLU
Dropout
Pooling
Up-sampling
Copy and crop

Lung Cancer Prediction

Image



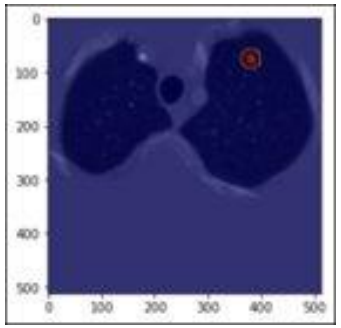
Nodule Mask



Feature Extractor

XGBoost Classifier

Cancer Detection



Business Impact

- The recent transformation of Healthcare industry towards personalized medicine requires analysis of large amount of medical imaging data. Robust automated technologies are essential for efficient and accurate analysis of multi-modal medical images in the context of early diagnosis, optimal treatment planning and treatment follow-up.
- The computer vision community is fully dedicated to developing new machine learning algorithms, neural network architectures, curated datasets and open source software.
- Computer-aided detection benefits from advances in deep learning to develop intelligent image analysis systems that are less biased by heuristic assumptions and thus minimize human error in clinical diagnosis.
- Semantic segmentation methods have achieved good results in recent years. Automatic image segmentation acting as second reader can effectively help radiologists to confirm their diagnosis thus increasing accuracy of diagnosis and greatly reducing the workload.

Ideas
Delivered
Intuitively

Thank You!

www.geninvo.com | Reach out to us info@geninvo.com