DENSENET POWERED LUNG CANCER PREDICTION

Rahimunnisa K [1], Dhanush S [2], Gokul M [3]

Department of Electronics and Communication, Easwari Engineering College

[1]rahimunnisa.k@eec.srmrmp.edu.in

[2]310620106027@eec.srmrmp.edu.in

[3]310620106032@eec.srmrmp.edu.in

Abstract— **Lung cancer has the potential to be life-threatening. Detecting cancer remains a significant hurdle for medical professionals, with the complete understanding of its origins and optimal treatment still elusive. However, timely identification of cancer can greatly enhance treatment prospects. Image processing techniques, including noise removal, feature detection, identification of affected areas, and potentially correlating with medical records pertaining to lung cancer history, are employed to pinpoint regions of the lung affected by the disease. This research demonstrates the precise classification and prediction of lung cancer through the utilization of machine learning and image processing technology. Computed Tomography images are utilized for identifying the lung cancer. A computerized tomography (CT) scan employs multiple X-ray images acquired from various angles around the body, employing computer processing to generate cross-sectional images (slices) that reveal the internal structures, including bones, blood vessels, and soft tissues. A dataset containing thousands of high-resolution lung scans, gathered from Kaggle. The preprocessing phase transforms raw data into a usable format. In the final stage, a Convolutional Neural Network (CNN) is deployed to assess the health status of the lung, discriminating between normal and abnormal conditions, with the deep learning algorithm attributing significance to the data.**

***Keywords*— Lung cancer, Computed Tomography, Machine learning, Deep Learning**

I. INTRODUCTION

Lung cancer stands as one of the most prevalent cancers globally and remains a primary contributor to cancer-related fatalities. It encompasses two primary classifications: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC)..NSCLC represents the predominant type, constituting approximately 85% of lung cancer cases, with SCLC comprising the remaining 15%. Lung cancer manifests as a condition marked by the unregulated proliferation and growth of abnormal cells within lung tissue. These cells deviate from normal cellular function due to DNA mutations induced by various factors such as smoking, exposure to air pollutants, and genetic predispositions. As per the American Cancer Society, lung cancer constitutes roughly 14% of all newly diagnosed cancers. In 2018, an estimated 234,030 new cases of lung cancer were reported in the United States, leading to approximately 154,050 deaths. Presently, lung cancer surpasses prostate, colon, and breast cancers combined as a leading cause of mortality [2]. Lung cancer is a prevalent and often fatal condition, claiming an estimated 422 lives worldwide each day. Predominantly afflicting individuals over the age of 50, the incidence of lung cancer continues to rise steadily. Due to its challenging detectability compared to other ailments, lung cancer stands as a leading cause of mortality. The primary impediment lies in the minute size of the initial lesion, commonly referred to as a nodule. Initially characterized by diminutive cancer cell dimensions, these lesions progressively evolve into malignancy over time. Hence, early detection plays a pivotal role in disease management. Timely identification significantly enhances survival rates. Recently, advancements in computer vision technology have yielded sophisticated networks capable of autonomously discerning and delineating healthy and tumorous regions [1].

The principal etiological factor of lung cancer is tobacco smoke, encompassing both direct inhalation and exposure to second hand smoke. Additional risk elements include exposure to radon gas, asbestos, environmental pollutants, and certain genetic predispositions.Clinical manifestations of lung cancer encompass persistent cough, haemoptysis, thoracic discomfort, dyspnoea, wheezing, laryngeal changes, unexplained weight loss, and generalized

fatigue. However, in its incipient stages, lung malignancy may remain asymptomatic, posing challenges to timely detection.

Diagnostic modalities commonly employed include radiographic imaging techniques such as chest X-rays, CT scans, and PET scans, coupled with histopathological biopsy for definitive identification of malignant cells. Therapeutic interventions for lung cancer are contingent upon the histological subtype and disease stage, comprising surgical resection, chemotherapy, radiotherapy, targeted molecular therapy, immunotherapy, or multimodal approaches.

Early detection assumes paramount importance in enhancing the prognostic outlook for lung cancer, as it facilitates intervention at more amenable disease phases. Screening protocols employing low-dose CT scans are advocated for select high-risk cohorts, notably individuals with a history of smoking, to facilitate early detection and Optimize treatment outcomes.

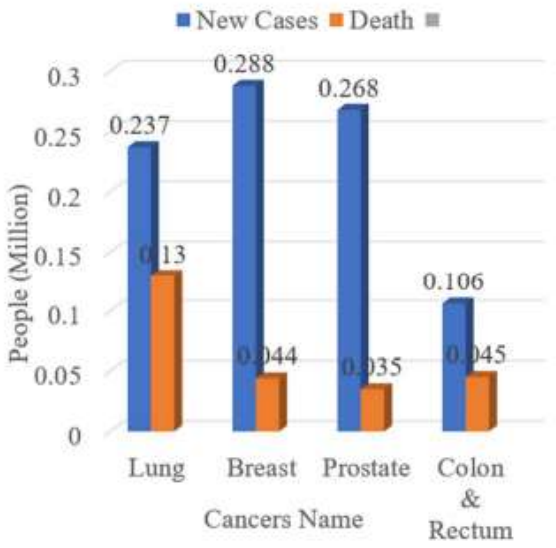

Figure 1: Cancer in 2022 (New cases against death) [5]

Figure 1 illustrates anticipated statistical insights for several cancer types as of 2022, derived from data provided by the American Cancer Society (ACS) [3]. According to ACS, lung cancer exhibits the highest mortality rate among all cancers, estimated at approximately 0.13 million worldwide. Each year, a substantial number of new cases emerge, with projections indicating around 0.237 million cases in 2022. The mortality rate remains notable due to late-stage diagnoses and the relatively high ratio of new cases to mortality, surpassing that of other cancers.

II. LITERATURE SURVEY

Bhatia et al. [3] developed a preprocessing pipeline using UNet and ResNet models to enhance feature extraction from lung CT images for cancer detection. They then combined XGBoost and random forest classifiers in an ensemble approach to assess malignancy likelihood, resulting in an 84% increase in accuracy compared to conventional techniques, as shown on the LIDC-IRDI dataset.

Joon et al.[4] utilized an active spline model for lung cancer segmentation analysis in X-ray images. Pre-processing involved median filtering for noise reduction, followed by segmentation using K-means and fuzzy C-means clustering. Feature extraction was conducted post-segmentation, and classification employed a Support Vector Machine (SVM) model in MATLAB to detect and classify lung cancer in both normal and malignant images.

Faruqui et al. [6] developed LungNet, a deep convolutional neural network (CNN) model that integrates CT-scan images with wearable sensor-based Medical Internet of Things (MIoT) data to improve computer-aided diagnosis (CAD) of lung cancer. LungNet achieved an accuracy of 96.81% and a low false positive rate of 3.35% when categorizing lung cancer into five classes, surpassing similar CNN-based classifiers. It demonstrated precise classification of stage-1 and stage-2 lung cancers into subclasses with an accuracy of 91.6% and a false positive rate of 7.25%. Trained on a balanced dataset consisting of 525,000 images and deployed on a centralized server, LungNet presents potential for automated lung cancer diagnosis systems.

Hasan and Al Kabir [8] developed algorithms to determine lung cancer spread using image processing and statistical learning. Their approach achieved 99.42% accuracy, significantly higher than previous methods, With recall as 99.42%, precision as 99.88%, and F-score attaining 99.82% respectively, its superiority is underscored.

Lakshmanaprabu et al. [9] developed OODN (Optimal Deep Neural Network) to improve lung cancer detection in CT scans by reducing features and comparing its performance with other algorithms. Automated classification streamlined human labeling, reducing time and errors. The study showed significant improvements in accuracy and precision, achieving a peer specificity of 94.56%, accuracy of 96.2%, and sensitivity of 94.2% in classifying lung images, demonstrating enhanced cancer detection feasibility in CT scans.

III. WORKING METHODOLOGY

A. Data Collection

Deep learning algorithms can identify affected nodules in lung cancer by analyzing the shapes, size, textures and intensities of highly detailed images provided by CT scans. 3D CT images, offering a complete imaging of lung capacity, provide a more comprehensive examination of the lungs than their 2D counterparts. Here the input data are consists of three type's especially benign case, malignant case and Normal case. Now see about that image in Detail for Lung cancer detection.

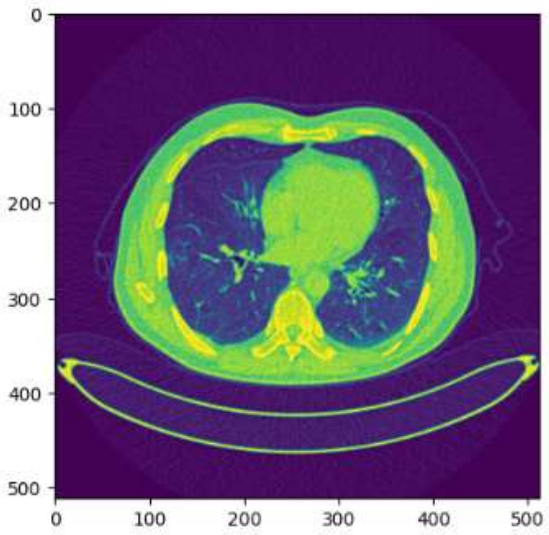

Figure 2: Benign Case Lung Cancer image

Benign lung tumors are growths that do not spread (metastasize) to other parts of the body and are typically not life-threatening. These tumors are often discovered incidentally during imaging tests conducted for other purposes,

such as an x-ray. The causes of most types of benign lung tumours are unknown. Some risk factors include: Genetics, infection, smoking. Benign lung tumors typically exhibit small sizes (less than 3 centimetres or roughly 1.5 inches), smooth and regular shapes with borders, and may have variable doubling times, which can be either fast or slow.

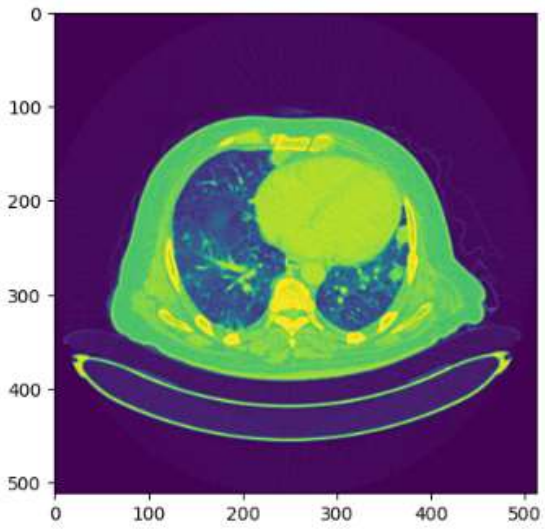

Figure 3: Malignant case lung cancer image

Malignant lung cancer CT images typically show tumors that are small, with sizes less than 3 centimeters (roughly 1.5 inches), and exhibit irregular shapes and borders. The doubling time of these tumors can vary, ranging from fast to slow growth rates.

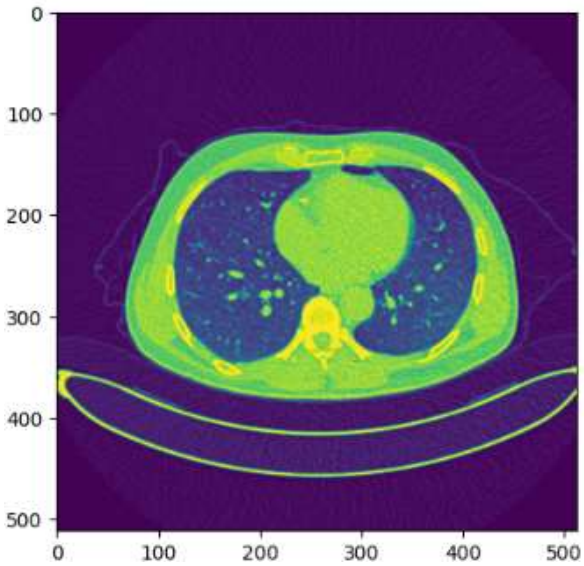

Figure 4: Normal Case Lung cancer image

B. Data Pre-processing

Preprocessing in lung cancer detection images involves several steps to enhance the quality and usability of the data. This typically includes tasks such as noise reduction, normalization of pixel values, resizing or rescaling of images to a standardized resolution, and sometimes augmentation techniques to increase the diversity of the training dataset. Additionally, contrast enhancement may be applied to improve the visibility of subtle features within the images. Overall, preprocessing aims to optimize the input data for more accurate and efficient analysis by machine learning algorithms in detecting lung cancer.

C. Model Architecture

Dense Net, a densely connected convolutional neural network architecture, is effectively employed in lung cancer detection using CT images. The distinctive connectivity pattern, where in each layer directly receives input from all previous layers, sets it apart., facilitates feature reuse and enhances gradient flow. This architecture excels in learning intricate patterns and features from CT scans, enabling accurate identification of lung cancer nodules. By leveraging DenseNet's capabilities, researchers have developed robust systems for automated lung cancer detection, aiding in early diagnosis and treatment planning.

D. Training and Validation

During the training process, our model is fed pre-processed CT scan images of the lungs, forming the input for our Convolutional Neural Network (CNN) architecture. Through iterative back propagation, the model adjusts its internal parameters, progressively learning discriminative features and patterns indicative of lung cancer. To mitigate over fitting, we employ a separate validation dataset, distinct from the training data, to monitor the model's performance. This rigorous methodology ensures that our CNN-based model proficiently extends its capacity to identify lung cancer in previously unobserved CT scan images.

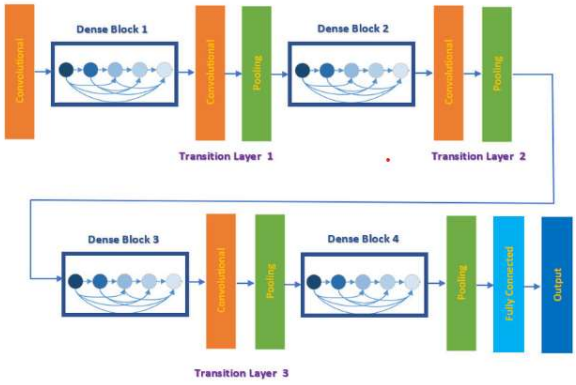

Fig 5: Dense Net architecture

E. Deployment and Integration

The final stage of our methodology focuses on the implementation of the trained Dense Net model for practical clinical applications in the detection of lung cancer.

IV. EVALUATION METRICES

In the context of LC detection, true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs) each carry distinct meanings. A TP signifies the accurate identification of a particular LC type, TN

denotes the correct identification of an image belonging to a different LC type, FP indicates an erroneous identification of an LC type, and FN represents the misidentification of an image belonging to a different LC type.

The accuracy metric assesses the predictive performance of the proposed model by quantifying the proportion of accurately predicted instances, inclusive of both true positives (TP) and true negatives (TN), relative to the total number of instances in the dataset. Precision denotes the ratio of true positives to the total predicted positives by the model, while recall represents the proportion of true positives among all actual positive cases. The F1-score presents the harmonic mean of precision and recall, with a higher F1-score indicating a model consistently exhibiting elevated levels of both recall and precision. The following equations are utilized to compute accuracy, sensitivity, and specificity for each case.

Class\metrices	TP	FP	TN	FN
Benign	30	20	125	0
Malignant	128	2	132	13
Normal	85	10	161	19

Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \rightarrow (1)$$

Accuracy for Benign cases:

Formula: (30 +225/ (30 + 20+225+0)) * 100

Value: 92.72%

Accuracy for Malignant cases:

Formula: (128+132 / (128 + 132+2+13)) * 100

Value: 94.5%%

Accuracy for Normal cases:

Formula: (85+161 / (85 + 161+10+19)) * 100

Value: 89.45%

Average Accuracy=92.2%

Precision:

$$Precision = \frac{TP}{TP + FP} \rightarrow (2)$$

Precision for Benign cases:

Formula: (30 / (30 + 20)) * 100

Value: 60.00%

Precision for Malignant cases:

Formula: (128 / (128 +2)) * 100

Value: 98.46%

Precision for Normal cases:

Formula: (85 / (85 + 10)) * 100

Value: 89.47%

Average precision=82.64%

Specificity:

$$specificity = \frac{TN}{TN + FP} \rightarrow (3)$$

Specificity for Benign cases:

Formula: (225 / (225 + 20)) * 100

Value: 91.84%

Specificity for Malignant cases:

Formula: (132 / (132 + 2)) * 100

Value: 98.51%

Specificity for Normal cases:

Formula: (161 / (161 + 10)) * 100

Value: 94.15%

Average Specificity=94.83%

Sensitivity:

$$sensitivity = \frac{TP}{TP + FN} \rightarrow (4)$$

Sensitivity for Benign cases:

Formula: (30 / (30 + 0)) * 100

Value: 100.00%

Sensitivity for Malignant cases:

Formula: (128 / (128 + 13)) * 100

Value: 90.78%

Sensitivity for Normal cases:

Formula: (85 / (85 + 19)) * 100

Value: 81.73%

Average Sensitivity=90.83%

V. RESULTS AND DISCUSSION

Within the domain of lung cancer detection, a confusion matrix stands as a fundamental instrument for assessing the efficacy of classification models. It presents a tabular representation of model predictions versus ground truth labels, particularly distinguishing between false positives (FP), and false negatives (FN).true positives (TP), true negatives (TN),

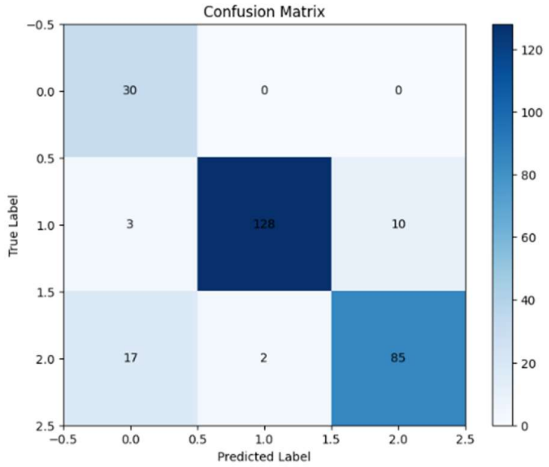

Figure 6: Confusion Matrix

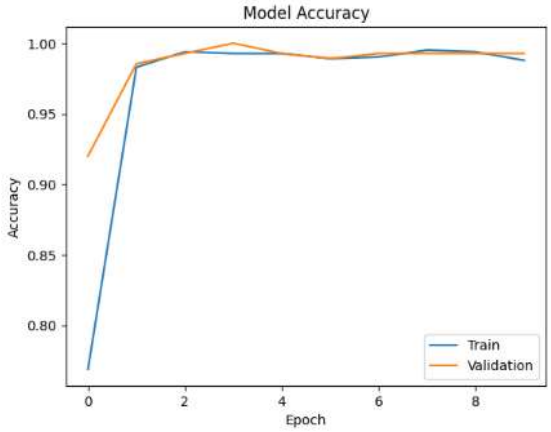

Figure 7: Model accuracy graph

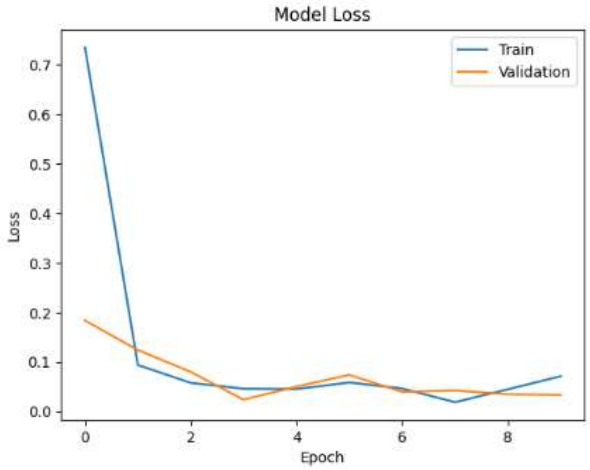

Fig 8: Model Loss Graph

Fig .6 shows the confusion matrix for the model illustrates the number of images correctly classified within the test dataset.. Fig .7 and Fig .8 Upon close observation of the accuracy graph, it's noticeable that initially, During multiple epochs, the validation accuracy surpasses the training accuracy. On the x-axis, the number of model training epochs signifies the cycles through the complete dataset, while the y-axis represents the loss and accuracy correspondingly.

VI. CONCLUSION

Our proposed automated classification scheme, based on Dense Net CNN, accurately identifies malignant lung cells in CT images. The evaluation of the method resulted in a sensitivity of 90.78%, specificity of 94.83%, and an overall accuracy of 92.2%. Visual analyses additionally validate the efficacy of our approach in detecting malignant cells in CT images..

REFERENCES

[1] Imran Nazir,Ihsan ul Haq,Salman A. AlQahtani,Muhammad Mohsin Jadoon,and Mostafa Dahshan·Machine Learning-Based Lung Cancer Detection using Multiview Image Registration and Fusion.

[2] Md Rashidul Hasan, Muntasir Al Kabir.Lung Cancer Detection and Classification based on Image Processing and Statistical Learning.

[3] Bhatia S., Sinha Y., Goel L. *Soft Computing for Problem Solving* . Singapore: Springer; 2019. Lung cancer detection: a deep learning approach; pp. 699–705.

[4] Joon P., Bajaj S. B., Jatain A. *Progress in Advanced Computing and Intelligent Engineering* . Singapore: Springer; 2019. Segmentation and detection of lung cancer using image processing and clustering techniques; pp. 13–23.

[5] LCDctCNN:Lung Cancer Diagnosis of CT scan Images Using CNN Based Model. Muntasir Mamun, Md Ishtyaq Mahmud, Mahabuba Meherin, Ahmed Abdelgawad.

[6] N. Faruqui, M. A. Yousuf, M. Whaiduzzaman, A. K. M. Azad, A. Barros, and M. A. Moni, "LungNet: a hybrid deep-CNN model for lung cancer diagnosis using CT and wearable sensor-based medical IoT data," *Computers in Biology and Medicine*, vol. 139, Article ID 104961, 2021.

[7] A. Shimazaki, D. Ueda, A. Choppin et al., "Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method," *Scientific Reports*, vol. 12, no. 1, Article ID 727, 2022.

[8] M. R. Hasan and M. Al Kabir, "Lung cancer detection and classification based on image processing and statistical learning," *Journal of Emerging Trends in Engineering and Applied Sciences*, vol. 11, pp. 229–236, 2020.

[9] Lakshmanaprabu S. K., Mohanty S. N., Shankar K., Arunkumar N., Ramirez G. Optimal deep learning model for classification of lung cancer on CT images. *Future Generation Computer Systems* . 2019;92:374–382. doi: 10.1016/j.future.2018.10.009

[10] Talukdar J., Sarma P. A survey on lung cancer detection in CT scans images using image processing techniques. *International Journal of Current Trends in Science and Technology* . 2018;8(3):20181–20186
[11] Travis W.D. Update on Small Cell Carcinoma and Its Differentiation from Squamous Cell Carcinoma and Other Non-Small Cell Carcinomas. *Mod. Pathol.* 2012;25:S18–S30. [PubMed] [Google Scholar]
[12] Chan B.A., Coward J.I.G. Chemotherapy Advances in Small-Cell Lung Cancer. *J. Thorac. Dis.* 2013;5((Suppl. S5)):S565–S578. [PMC free article] [PubMed] [Google Scholar]
[13] Sagawa M., Nakayama T., Tsukada H., Nishii K., Baba T., Kurita Y., Saito Y., Kaneko M., Sakuma T., Suzuki T., et al. Te efficacy of lung cancer screening conducted in 1990s: Four case–control studies in Japan. *Lung Cancer.* 2003;41:29–36. doi: 10.1016/s0169-5002(03)00197-1. [PubMed] [CrossRef] [Google Scholar]
[14] Fontana R.S., Sanderson D.R., Woolner L.B., Taylor W.F., Miller W.E., Muhm J.R. Lung cancer screening: Te Mayo program. *J. Occup. Environ. Med.* 1986;28:746–750. doi: 10.1097/00043764-198608000-00038. [PubMed] [CrossRef] [Google Scholar]
[15] Kubik A., Parkin D.M., Khlat M., Erban J., Polak J., Adamec M. Lack of benefit from semi-annual screening for cancer of the lung: Follow-up report of a randomized controlled trial on a population of high-risk males in Czechoslovakia. *Int. J. Cancer.* 1990;45:26–33. doi: 10.1002/ijc.2910450107. [PubMed] [CrossRef] [Google Scholar]
[16] Raghu V.K., Zhao W., Pu J., Leader J.K., Wang R., Herman J., Yuan J.-M., Benos P.V., Wilson D.O. Feasibility of lung cancer prediction from low-dose CT scan and smoking factors using causal models. *Thorax.* 2019;74:643–649. doi: 10.1136/thoraxjnl-2018-212638.