TEAM DOC

(Detection of Cancer)

Enhancement of Detection and Diagnosis of Non-Small Cell Lung Cancer Through The Improvement of Machine Learning and AI Models

March 29th, 2024

Team Members: Yael Beshaw, George Cancro, Darren Chang, Jayda Fomengia, Vanshika Mehta, Arjun Vedantham, Ritvik Yaragudipati

Mentor: Dr. Soheil Feizi, Department of Computer Science at the University of Maryland

TABLE OF CONTENTS

Abstract

Due to low survival rates and an unparalleled burden of non-small cell lung cancer on underserved communities, there is great urgency for innovative and accessible methods that will improve healthcare access for lung cancer patients. To combat this inequity, Team DOC aims to develop an AI model that is able to not only improve lung cancer diagnoses but also predict the progression of non-small cell lung cancer. We intend to evaluate the performance of a convolutional neural network on the LIDC-IDRI dataset and retrain the final layers of the model to improve its performance on the same dataset. Repeating this process on different model architectures allows us to determine which model performs optimally, providing a foundation to develop an end-to-end explainable AI workflow that can extract clinically relevant predictions of cancer progression for further analysis. Throughout our training process, we resolve to address the accuracy and potential for bias. Additionally, we are carrying out a survey among underserved populations and communities to discern the need for our improved cancer detection model. We hope that our model will be able to be implemented in communities with lack of access to healthcare systems to bridge the gap between underprivileged communities and unbiased care.

Key Words: *Lung Cancer, Non Small Cell Lung Cancer, AI, Datasets, Models, Algorithm, Bias, Survey, Detection, Diagnosis, Diagnostics*

Introduction

With more than 1.6 million fatalities a year and low five-year survival rates, there is a great need for early diagnosis of lung cancer (Ardila et al., 2019). Currently, diagnostic techniques rely on Computed Tomography (CT) imaging which provides in-vivo capabilities to measure extent and location of lung lesions and morphological manifestations (Purandare & Rangarajan, 2015). After investigation of lung, brain, and breast cancers, the determination to focus our research on lung cancer was on the basis of available datasets, the lackluster progression in the field for this specific cancer relative to the other two, and society's heightened need for more efficient detection.

While technologies and screening methods such as CT imaging are key to surviving cancer, cancer is most treatable when it is detected during its early stages. However, some cancers, particularly lung cancer, go along undetected until they are at an advanced stage, drastically decreasing survival rate (Midthun, 2016). This issue is what perpetuates lung cancer as the leading cause of cancer-related deaths in men and women internationally, with about 1.6 million deaths per year (Global Burden of Disease Cancer Collaboration, 2015). These deaths are even more prominent among low-income communities and people of color (National Institutes of Health, 2015). While cancer itself is the uncontrollable division of cells in a specific location of the body, many have linked the causes of lung cancer to patterns of cigarette smoking and regions of industrialization (Barta et al., 2019). Including factors such as disparities in healthcare access, environmental conditions, and cultural barriers it has become increasingly imperative to explain why the detection and diagnosis of lung cancer is often delayed among those in the developing world or regions that lack advancements in healthcare (Barta et al., 2019). As a

result, there are varying degrees in the prevalence of lung cancer in regions around the world, with cigarette smoking being a key factor.

The delayed diagnosis of lung cancer is a great cause for concern and has a large impact on the treatment and survival rate. For nearly all types of cancer, survival rates drastically increase when the disease is caught at an early stage (American Cancer Society, 2021). Cervical cancer, for example, has a 15% survival rate when diagnosed at an advanced stage, compared to a 93% survival rate if diagnosed prior to when the cancer spreads (American Cancer Society, 2021). There is an established socioeconomic barrier to accessing early cancer detection, as those from underserved communities may not be able to afford routine checkups by a primary care physician, or may have poor health awareness. “Delays in the Diagnosis of Lung Cancer” , suggests that the delay in lung cancer detection is due to a variety of reasons, including ignorance of what symptoms indicate the presence of lung cancer and delays in doctor’s visits for said symptoms (Ellis, 2011). Many times patients get treated for infections in the lung rather than getting screened for lung cancer, which prolongs the time to receive a completed diagnosis in addition to a relatively complicated diagnostic process.

The need for early detection is of the utmost importance. Currently, only 16% of lung cancer cases are detected early, which is low relative to other cancers (American Lung Association, 2020). This is due to the fact that symptoms of lung cancer do not manifest until later on in the lung cancer process (National Cancer Institute, 2020). It may also be because of the stereotypes that surround lung cancer, such as only those who smoke cigarettes or vape can get lung cancer. In fact, in the United States, 10-20% of lung cancers occur in non-smokers every year (CDC, 2020). This stigma is untrue and discourages people from getting screened for lung cancer. Our team hopes to create a model that is accessible for physicians to use in their primary

care facilities, so that even those who believe they may not be at-risk can receive a screening at their yearly check-up.

With a net five-year survival rate of 18.6% in 2020, the high mortality and morbidity of lung cancer accentuates the need for efficient, accurate lung cancer screening techniques (American Lung Association, 2020). With significant innovations in immunotherapy and targeted oncology treatments, highly accurate morphological classification, localization, and temporal modeling are in urgent need for early diagnosis (Mayo Clinic). Currently, 75% of cases are detected only in its advanced metastatic stages, associated with significant pulmonary nodal spread (Wang et al., 2020). Over the past 5 years, significant progress has been made in lung cancer prevention and treatment. Results from the 2018 USA National Lung Screening Trial (NLST) have demonstrated a 20% overall reduction in lung cancer mortality, widely attributed to the nationwide lung cancer screening program in the USA and the increasing utilization of low-dose computed tomography (LDCT). LDCT is a screening test which uses a low dose of X-ray radiation to make detailed images of the lungs. However, one of the largely known risks of LDCT is its rate of high false positive results (tests which suggest that a person has lung cancer when no cancer is actually present), and overdiagnosis (the identification of non-problematic cancer instances) leading to unnecessary treatment. For instance, the rate of positive diagnoses in the NLST was ~27% and ~17% in the first 2 years of screening. For the purposes of this study, an CT instance was considered positive if it contained visible abnormalities, including a non-calcified nodule of at least 4 mm on the long axis. But a retrospective analysis revealed that over 96% of such positive diagnostic results were actually false positives, and about 72% of these necessitated some form of secondary testing (Ardila et al., 2019). Thus, it is evident that the fine differentiation of benign and malignant 5-15 mm sized pulmonary nodules, which are

significant biomarkers of cancer progression, using a minimal number of CT scans, while reducing the rate of false positives, is a challenging task for radiologists.

More recently, results from a number of studies have pushed for a shift from the traditional optical microscopic examination to more computational approaches using artificial intelligence and computer vision techniques to characterize pulmonary nodules with a high accuracy, all the while requiring a minimal number of CT scans. AI is a general class of self-learning algorithms that enable computers to simulate intelligent behavior and critical thinking, increasingly on-par with those of humans. Machine learning, a large branch of AI, focuses on the paired usage of data and algorithms to imitate the way humans learn, gradually improving its accuracy over time. It is increasingly based on the idea that systems can directly learn from data, identify intrinsic patterns, and consequently make ever-improving decisions with minimal human intervention.

Deep learning, also known as hierarchical learning or deep structured learning, is a specific class of AI techniques that uses a layered algorithmic architecture to analyze data. As denoted by their hierarchical or layered paradigm, it is loosely based on the way in which biological neurons interconnect, where data is filtered through a cascade of multiple layers, with successive layers extracting low to high-level features with increasing specificity. Deep learning models are self-learning, and they become increasingly better at prediction as they process more data and refine their ability to make correlations and connections. These classes of algorithms have gained significant traction in the field of computer vision (which studies applications of image processing techniques to analyze visual data), and these techniques have specifically shown unmatched efficacy in tasks like image classification. Our core aim in this project is to develop a robust, automated, and intelligent AI-driven system that (1) make improved lung

cancer diagnostic and long-term survival rates predictions by mining imaging features from 3D volumetric data, (2) enhance the explainability of these probabilistic predictive models to support robust, human guided decision-making, (3) are portable and easy to use, especially in regions with under-resourced and underserved populations.

Literature Review

In this section, we provide a foundational overview of the chief focus of our study, lung cancer, along with AI algorithms such as deep neural networks which are driving state-of-the-art advancements in lung cancer early diagnostics over the past decade.

1. Cancer

Although many have heard of the term "cancer" and associate it with illness and fatality, there remains uncertainty about what cancer actually is, how it spreads, and how deadly it can be. The National Cancer Institute (NCI) defines cancer as the uncontrollable and abnormal division of cells that is able to spread to and invade tissue (National Cancer Institute at NIH, 2021). A common term regarding cancer is "tumor"- tumors are lumps of tissue, and they are often caused by the abnormal division of cells in the body. The location of these tumors often give medical professionals a clue about where the cancer has started and the severity of it. There are two types of tumors; benign (meaning there is no cancer present) or malignant (indicates presence of cancer) (Patel, 2020). There are key differences between the two that allow medical professionals to determine the next steps for their patients. Being non-cancerous, benign tumors do not spread to other parts of the body or invade tissue otherwise known as metastasis. As a result, they are not usually life-threatening and do not grow at a fast rate while malignant tumors do the

opposite. The table below provides an overview of the differences between benign and malignant tumors that both professionals and patients are able to refer to when diagnosing breast cancer.

Features	Benign	Malignant
Growing	Slow growing	Fast growing
Capsulated	Yes	Non- capsulated
Invasive	Non- Invasive	Invasive & infiltrate
Metastasis	Non	Yes
Shape	Smooth/ oval / lobulated / regular	Nodular /Stellate / Irregular
The pains	Painful	Painless
Skin	No skin dimpling	Skin dimpling
Nipple	No nipple retraction	Nipple retraction
Prognosis	Good	Bad
Damage to human body	Relatively smaller	Relatively bigger

Figure 1.1: Hamouda, et al., (2017). Table of differences between benign and malignant breast cancer tumors.

Lung cancer tumors can be differentiated in a similar fashion. Providers may chart the nodule size and its rate of growth, with smaller, slowly growing nodules considered more likely to be benign. Malignant tumors grow rapidly, doubling in size less than four months (Cleveland Clinic, 2024). However, unlike breast cancer, lung nodules are also examined for their content and shape. Besides differences in “Shape”, “The pains”, “Skin”, and “Nipple”, this table can be used to differentiate tumors across various cancers. In order to actually take a look at these tumors, medical professionals are able to utilize an array of methods and technologies. According to the Mayo Clinic, physicians are able to perform physical exams, biopsies, laboratory tests, and imaging tests to diagnose cancer (Mayo Foundation, 2021).

Physical exams consist of the physician examining the physical manifestations of the tumor or cancer by indicating changes in skin such as the color, texture, or by feeling the tissue

mass if possible (Mayo Foundation 2021). Indications of breast cancer are commonly examined this way upon a patient's first visit and cancer institutes recommend self examination using physical exam methods (American Cancer Society, 2021). However, this may usually require a confirmation of the physician's diagnosis by using one of the next three methods as physical examinations can only take a look at how the tumor or cancer is impacting the outside of the body. Thus making it difficult to determine whether the tumor is cancerous or not. The next three methods are used depending on the type of cancer that is cause for concern for the patient, meaning not all technologies and methods can be or should be used for all cancers. Biopsies are considered to be the most invasive way to detect and diagnose cancer but are the most common (National Cancer Institute, 2021). This method requires the physical removal of tissue or cells from the region of concern, while there are many removal techniques, they are still invasive and may be considered unnecessary for patients who undergo them after only a physical examination. Laboratory tests are most commonly used for cancers that indicate a change in bodily fluids such as urine or blood. For example, to diagnose leukemia (a cancer of blood forming tissues), blood samples are taken to a lab to detect changes in blood cell count (white and red blood cells) using a complete blood count (CBC) (Mayo Foundation, 2021).

Imaging tests are the most commonly known and include using technologies such as ultrasounds, X-rays, Magnetic Resonance Imaging (MRI) scans, Computed Tomography (CT) scans, etc. This method uses non-invasive technologies that are able to provide physicians with images that clearly outline the absence or presence of the tumor in question, the type of tumor it is, and the scope of it as well (National Cancer Institute, 2021). In the MRI scans of two brains provided, there is clear evidence for the presence of a tumor and differentiation between a benign and malignant tumor (Nasim et. al, 2019). The benign tumor is localized (the white region), with

no indication of any spreading or growth beyond the centralized location. The malignant tumor, however, is decentralized and spreads throughout the brain. Since this method is the most non-invasive and is able to provide physicians, researchers, and patients alike with images that allow for indications of tumors and stark differences between what type of tumors are present, it is the method of interest for many technological advances (Fass, 2008). For cancers such as lung cancer, there is a wide range of imaging techniques that are able to aid medical professionals in the detection and diagnosis of lung cancer including MRI, CT, Positron Emission Tomography (PET), and chest radiography. According to a study comparing these techniques CT and PET scans are concluded to be the best technique to use (Wever et al., 2011). Team DOC aims to utilize the technological potential of imaging scans to enhance the detection of lung cancer. It is our hope that these efforts can aid in avoiding invasive or inaccurate techniques of detection and further diagnosis.

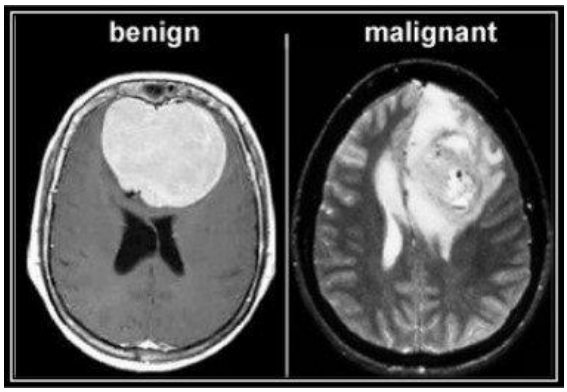

Figure 1.2: Nasim, et al., (2019), MRI scans of two brains (left; benign/non-cancerous, right; malignant/cancerous)

2. Lung Cancer

Lung cancer is one of the most common fatal malignancies in adults worldwide (Detterbeck et al., 2013). Cancerous tumors of the lung can be classified into three sub-groups: non-small cell lung cancer (NSCLC), small-cell lung cancer (SCLS), or lung-carcinoid tumors (American Cancer Society, 2021). NSCLC is the most common of the three, affecting 80-85% of all lung cancer patients. It means that the cancer has not spread outside the lung and is localized. SCLC is less common (10-15%), but is more aggressive than NSCLC. About 70% of people who

have SCLC are only diagnosed once the cancer has already spread to the rest of the body. Lung-carcinoid tumors make up less than 5% of lung cancers and generally grow more slowly than NSCLC and SCLC (American Cancer Society, 2021).

Lung cancer is highly linked with cigarette smoking (Alberg et al., 2013). Cigarette smoke, which consists of over 20 different cancer-causing chemicals called carcinogens, when inhaled, can damage the cells that line the lungs (Minna et. al, 2002). By increasing the frequency of smoking, it also increases the exposure of the lung cells to the carcinogens, increasing the damage. Over time, the cells will start to not function properly and develop into cancer (Mayo Foundation, 2021). People with lower levels of education or living below the poverty level have higher rates of cigarette smoking, and therefore higher rates of lung cancer (Centers for Disease Control and Prevention (CDC), 2021). According to the CDC, lung cancer rates are 18-20% higher in populations living in rural, deprived areas compared to those living in suburban, wealthy areas. The higher lung cancer rates can also be attributed to the fact that lower-income populations have less access to healthcare. This makes it more likely that their cancer will not be detected and diagnosed until it is untreatable. Additionally, research has concluded that perceptions of lung cancer differ from others, as it is comparable to a "punishment" or "loss of autonomy" (Mazières et al., 2015). This societal bias contributes to the difficulty of diagnosing patients early and creating a sense of urgency to address existing disparities.

a. Characteristics & Types of Lung Cancer

As mentioned previously, lung cancer can be broken up into two major categories: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). Small cell lung cancer is distinguished by its aggressive path from the bronchi, or breathing tubes, in the chest. Though these cancer cells are small, they grow very rapidly and form large tumors (American Cancer Society, 2021). The 5-year survival rate of small cell lung cancer can vary from 7-27% depending on how much the cancer has spread at time of diagnosis (American Cancer Society, 2021). It is typically treated with chemotherapy and radiation therapy.

Non-small cell lung cancer is any lung cancer not classified as small cell lung cancer. It can be broken up into three main types: squamous cell carcinoma, large cell carcinoma, and adenocarcinoma (National Cancer Institute at NIH, 2021). While each of these subtypes differ in their lung cells of origin, they are grouped together as NSCLC because of their similar treatment and outlook. NSCLC is less aggressive than SCLC, with a 5-year survival rate of 25-63% (American Cancer Society, 2021). About 80-85% of lung cancers fall under NSCLC.

3. Current Diagnostic Methods of Lung Cancer

The most common method of lung cancer imaging throughout the world is through chest radiography, due to its low cost, low radiation, and widespread availability. In combination with temporal subtraction, which deletes the gross anatomical structures from the radiography image, the progression of the cancer in the tissue can be determined without distraction from the overlying structures. While chest radiography is the first method used for lung screening and remains to be the most popular, studies in the past have shown that the technique does not significantly reduce lung cancer mortality rates (National Cancer Institute at NIH, 2021). Since then, doctors in first-world countries such as the United States and the United Kingdom have

moved to abandon chest radiography as a way to screen lung cancer. Chest radiography does prove to be useful in characterizing lung lesions as benign or malignant as well as lesion size; however, it cannot demonstrate the full metastasis of the cancer, such as whether it has spread into the walls of the chest or the lymph nodes.

While chest tomography and sputum cytology have both been used to screen lung cancer, there is no support that the usage of these methods decrease the rates of lung cancer or decrease lung cancer mortality (National Cancer Institute at NIH, 2021). However, complete absence of screening can prove to be disadvantageous and increase incidence and mortality rates of lung cancer.

The most successful method of screening for lung cancer is through low-dose helical CT scanning, which unlike chest radiography, has reduced lung cancer mortality rates 20-24% (National Cancer Institute at NIH, 2021). The success of CT imaging in diagnosing cancer is due to its design and ability to give a detailed picture of not only the localization and spread of the tumor, but also how the cancer has affected the lymph nodes and whether it can be classified as either benign or malignant (Wever et.al, 2011). When integrated with positron emission tomography (PET) imaging, the scans can reveal both anatomical information about the cancer and the metabolic activity of the cancer cells. It also allows for 3D reconstruction of the patient's lung cancer, which is beneficial for physicians to determine the best course of treatment. In addition to improved spatial resolution, CT imaging brings the benefit of increased temporal resolution, as data is acquired within a single breath of the patient (Wever et.al, 2011).

This high-resolution, high-dimensionality data fits well with image classifier architectures like 3D convolutional neural networks (CNNs). Like traditional neural networks, CNNs are graph-based data structures consisting of an input layer in which each pixel of the

supplied image is mapped to a specific graph node. Each intermediate layer consists of nodes that transform the input pixel values by applying a multiplicative weight or additive bias value, before sending those values forward in the graph. In contrast to traditional neural networks, CNNs also have convolutional layers, which pool together values from a part of the intermediate graph to reduce the dimensionality and can implicitly extract key features of an image - a type of implicit, automatic feature extraction that avoids manual feature engineering. Finally, these are mapped to the output layer, consisting of nodes that correspond to a set of labels. In this last layer, the node with the highest value generates the network's overall prediction (O'Shea & Nash, 2015).

We began by evaluating existing approaches in the literature to identifying lung cancer nodules. One of the projects that we found was DeepLung, a project from researchers at UC Irvine and Tencent that aimed to improve segmentation accuracy (Zhu et al, 2018). Although not identical to detection, segmenting a nodule candidate from a large image is an important step in the diagnosis pipeline for future malignancy classification. However, while this was a good baseline to understand the current state of the art, we identified key issues with the paper and its attached code. First, there were no attempts at explaining how the model had identified a nodule candidate site during the segmentation step, or explaining what factors in a nodule image led to a benign or malignant classification. This represents a large gap that needs to be addressed when autonomous systems like this are deployed to clinical settings, where models are expected to work in tandem with human physicians.

In addition, while DeepLung did provide a repository with the code used for their project, much of the code was written for much older software libraries, and the results presented in this paper could not be replicated on modern systems. Based on these factors, we identified model

interpretability as a prime research area in our project. Other existing approaches relied on human-defined image segmentations and would classify 3D volumes that corresponded to potential nodules sites. While models in this paradigm may produce more accurate results, they still require manual intervention to pick out potential sites, and cannot handle scans as they are generated by medical imaging equipment. Ideally, our model would give accurate results, while minimizing the amount of manual preprocessing work (Alakwaa, 2017).

a. GradCam and Image Saliency

When using computer vision in convolutional neural networks, it is important to consider the interpretability of their outputs. When evaluating results of these models, we should prioritize an understanding of our models' learning and decision-making. Techniques such as image saliency maps provide insight into the regions of interest of the input image with respect to the output label. Image saliency is designed to mimic the visual perception of a human to provide structure to visual data.

GradCAM (Gradient-weighted Class Activation Mapping) is a tool for improved convolutional neural network interpretation and explanation. GradCAM produces a saliency heat map that highlights the regions of an image that contribute most significantly to the prediction of the model. GradCAM is an improved version of CAM (Class Activation Mapping) which is based on activation maps in the final convolutional layer of the network that indicates the learned abstract feature of an input image [2]. By identifying these activation maps, CAM is able to extract the regions that contributed most significantly to the target classification. GradCAM builds on this model by adding the additional steps of gradient computation on the target class with respect to the feature maps, then a weighted combination with gradients using global average pooling (GAP) and ReLU activation. The resulting activation map provides more

accurate and localized visual explanations for the classification, with more accurate localized salient regions than CAM with the inclusion of gradients (Zhou et al, 2015).

GradCAM can be applied to medical image analysis, where the regions of interest can be highlighted in radiological images. The CT scan image datasets that were explored and used to build our CNN model demonstrate an excellent use case of GradCAM to provide better understanding and explainability of our model's predictions. In our CNN model designed to classify CT scan images of patients' lungs as benign or malignant cancer, GradCAM was implemented to construct heat maps based on the gradients in the final convolutional layer to provide visualization of this model's decision-making process. GradCAM can be applied to an input CT scan by loading our pre-trained CNN model, preprocessing the CT scan into a suitable input vector, then extracting feature maps and gradients from the loaded model to obtain weighted feature maps that can be visualized as a heatmap and overlaid on the input image. The initial use case when applying GradCAM to our model is visual confirmation that the regions of interest of the model aligned with specific than localized areas of the lung that contained a nodule or tumor, instead of heavily weighting the whole image or other areas of the lung with no meaningful features.

b. MedMNIST v2

MedMNIST v2 is a large-scale dataset collection based off of the MNIST (Modified National Institute of Standards and Technology) database (Yang et al, 2023). MedMNIST v2 consists of 12 datasets for 2D images and 6 datasets for 3D images. MedMNIST v2 was designed to perform classification on 2D and 3D images.

We focused on one of the 3D classification models called NoduleMNIST3D. This model is based on the LIDC-IDRI (Lung Image Database Consortium Collection) dataset which contains images from thoracic CT scans. NoduleMNIST3D was designed for both lung nodule segmentation and to perform binary, malignancy classification tasks on thoracic CT scans. This model used 8 methods to classify scans: ResNet-18+2.5D, ResNet-18+3D, ResNet-18+ACS, ResNet-50+2.5D, ResNet-50+3D, ResNet-50+ACS, auto-sklearn, and AutoKeras. This model was trained using 1,158 thoracic CT scans, validated using 165 thoracic CT scans, and tested using thoracic CT 310 scans.

We used this model because the LIDC-IDRI dataset used by NoduleMNIST3D and binary classification were aligned with the thoracic CT scan datasets we were using and our goals. We used this model as a benchmark for our own model and to compare accuracies.

4. Data

For the team's specific research, medical privacy rules presented a key barrier towards obtaining medical datasets of non-small cell lung cancer images. However, high-quality, anonymous public data sets are available, and we analyzed several popular and freely available datasets to determine which ones would be best suited for our model. Ideally, each data set needed to have a variety of patients being scanned based on race, gender, age, etc. When we were initially considering our model, we decided on the following five datasets as the best options to consider:

a. NSCLC-Radiomics (Lung1 Dataset)

The first dataset we considered is the NSCLC-Radiomics, Lung1 Dataset. This dataset is provided by The Cancer Imagine Archive Public Access, and is available for free installation in

all machines. The Lung1 dataset contains images from 422 lung cancer patients, including pretreatment scans, tumor specific information and clinical outcome data. The dataset includes 3D image scans of the tumors with a radiation oncologist's official diagnosis, along with other important structural information that contribute to identifying the tumor.(Kirby, 2021). This in-depth analysis of the tumor is helpful when looking for trends amongst scans, as well as choosing factors that impact the growth/span of the tumor itself. Some applications of this dataset in current research include helping build the "prognostic radiomic signature."(Kirby, 2021)

b. NSCLC-Radiomics (Lung3 Dataset)

Another dataset option to consider is the NSCLC-Radiomics, Lung3 Dataset. Similar to the Lung1 Dataset, the Lung3 Dataset is provided by The Cancer Imaging Archive Public Access. The Lung3 Dataset contains images of 89 Lung Cancer patients that have had surgery as a treatment form. The Dataset provides information including an initial CT scan, gene models and general clinical data available. An analysis of the dataset revealed "We found that a large number of radiomic features have prognostic power in independent data sets, many of which were not identified as significant before. Radiogenomics analysis revealed that a prognostic radiomic signature, capturing intra-tumour heterogeneity, was associated with underlying gene-expression patterns." The analysis of the dataset revealed the importance and impact this dataset's research held in advancing the overall lung cancer research. An example of applications of the dataset were to analyze gene expression models. (Kirby, 2021)

c. RIDER Lung CT - Cancer Imaging Archive

The RIDER Lung CT data set is created in order to study the "variability of tumor unidimensional, bidimensional, and volumetric measurements on same-day repeat computed

tomographic (CT) scans in patients with non–small cell lung cancer." (Commean, 2021). This data set, consisting of 32 patients, has a total of 15,419 images. 3 radiologists performed two CT scans per patient, both in a 15-minute time span. They measured the two greatest diameters of each lesion on both scans, as well as the same tumors on the first scan. With the help of computer software, " Concordance correlation coefficients (CCCs) and Bland-Altman plots were used to assess the agreements between the measurements of the two repeat scans (reproducibility) and between the two repeat readings of the same scan (repeatability)" (Commean, 2021). The concordance correlation coefficients are used as benchmarking to compare the team's AI model with the current "gold standard". Bland-Altman plots help identify any systematic difference between the measurements (i.e., fixed bias) or possible outliers. This data set is a strong candidate for the model to be trained on. Even if the size of the data is on the smaller side, the reproducibility and repeatability of both the radiologists and computer measurements were both above the 95% mark.

d. LIDC-IDRI (Lung Image Database Consortium Collection)

The Lung Image Database Consortium image collection (LIDC-IDRI) consists of over 244,000 images of CT lung cancer scans. This data set was created by seven academic centers and eight medical imaging companies, and initiated by the NCI. 1018 cases are contained in this specific data set, and "each subject includes images from a clinical thoracic CT scan and an associated XML file that records the results of a two-phase image annotation process performed by four experienced thoracic radiologists" (Vendt). In the initial phase, each radiologist reviewed the CT scans independently and categorized each lesion into 3 sections: nodule >= 3 mm, nodule < 3 mm, and non-nodule >= 3 mm. During the second phase, "each radiologist independently reviewed their own marks along with the anonymized marks of the three other radiologists to

render a final opinion." (Vendt, 2020). The reason this data set is a strong candidate for training the model is because of the amount of review that went into each image. Because of the anonymous analysis, there is no "forced consensus" of any lung nodules in each CT scan. Not only that, but, inevitably, we chose to use this dataset since it encodes image data using the Digital Imaging and Communications in Medicine (DICOM) standard, which is the file format used by most medical scanners. For the purposes of training the classifier, we will be using the subsection of the dataset containing the thoracic CT scans as this constitutes a majority of the dataset and is the type of scan most frequently used in existing literature about computer assisted lung cancer detection.

e. HRCTCov19-a high-resolution chest CT scan image dataset

The HRCTCov19 dataset is a comprehensive collection of 181,106 chest high-resolution computed tomography (HRCT) images from 395 patients. These images are recorded in the Digital Imaging and Communications in Medicine (DICOM) standard as 16-bit grayscale images with a resolution of 512 × 512 pixels. The dataset includes images labeled with various conditions, such as Ground Glass Opacity (GGO), Crazy Paving pattern, Airspace Consolidation, etc. The team is focused on the subsection of the dataset which provides clean DICOM images of lung nodules as a baseline for comparison for the model. This dataset is a strong contender for training, as it is one of the few data sets with a large selection of baseline images for training. This is essential in order to establish a correct detection of non-cancerous images, hence why the team will be going forward with this image dataset for training purposes.

Methodology

- **AI**

As a baseline, we started training our classifier by using a 2D image classifier based on the documentation provided by the Keras and Tensorflow machine learning libraries (Tensorflow Developers, 2023). However, 2D classifiers discard the contextual information present within a scan, and instead treat images in a particular class as independent samples drawn from a common distribution for a particular image class. However, our datasets specifically contained higher dimensional image data for each scan, and downsampling this data into a sequence of 2D slices meant that we lost significant contextual information and interpretability. Following this, we explored model architectures that were capable of accommodating not just 2D image data, but 3D image data with an arbitrarily high number of 2D slices.

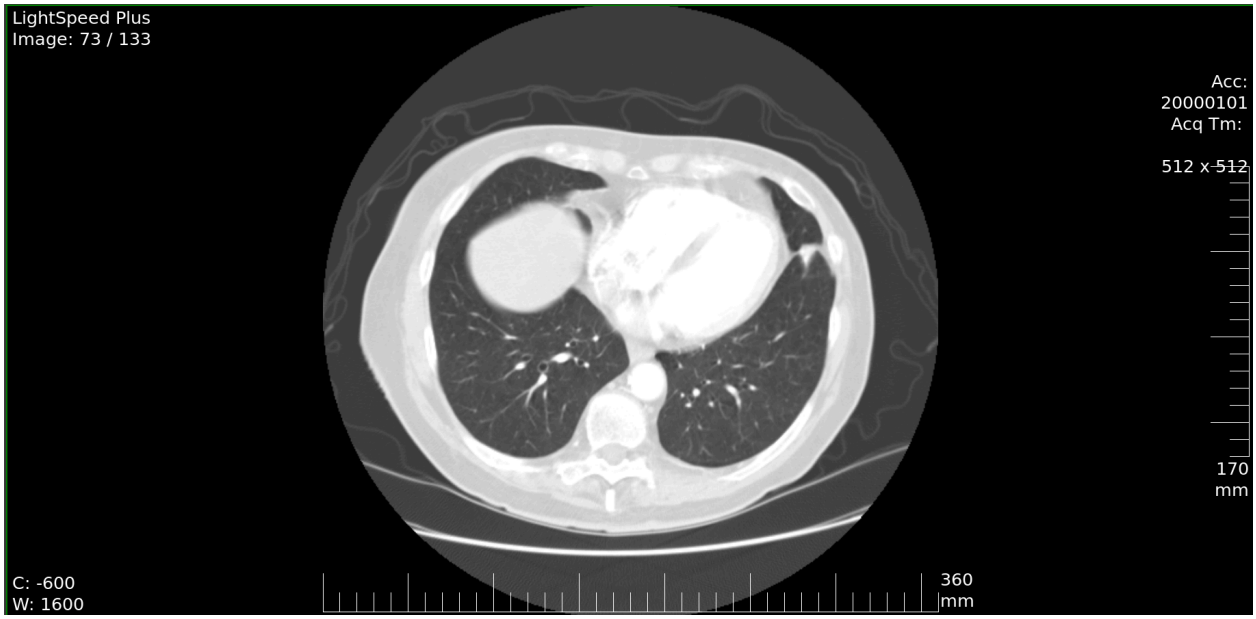

Figure 2: An example of a 2D scan slice, about halfway through the scan depth. Each scan in the LIDC-IDRI set consists of approximately 100 of these slices, depending on the exact resolution of the scanner used.

To this end, we used a 3D convolutional neural network, based on a model architecture explored in the documentation for Keras (Zunair, 2020). We also evaluated using a temporal convolutional network (TCN) as a comparison point, due to its ability to handle higher dimensional image data. While TCNs are generally used for problems like video classification and are well-suited for time-series tasks, we also believed that our high depth medical scans could also be well-suited for training a TCN based model. However, we found that using a 3D convolutional network was easier to use, and that our experimental model accuracy was unaffected by scaling the volume of the input matrices, which is why we had considered using a TCN-based model in the first place. As such, we continued to build off of 3D convolutional networks due to their robust classification ability, and widespread usage in existing lung cancer detection literature.

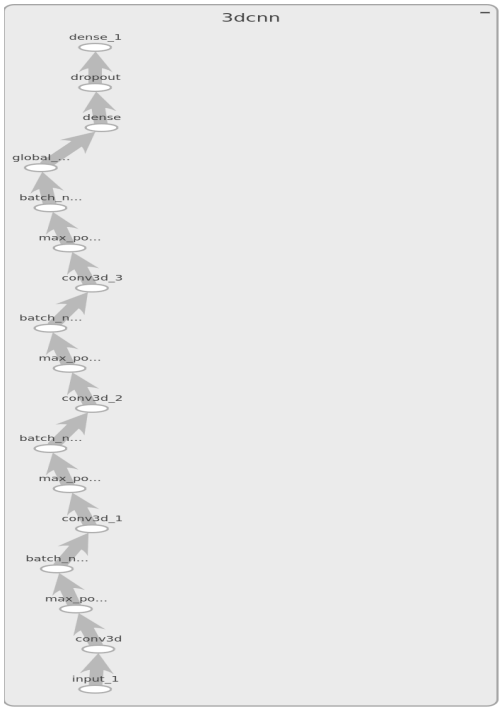

Figure 3: Our 3D convolutional neural network graph. This graph is inverted, such that data enters at the bottom of the visualization (“input_1”) and passes through several convolution and pooling layers before getting to the final (“dense_1/dropout”) layers

DICOM images were then loaded in from the LIDC-IDRI dataset. These images contain 512 x 512 x n dimensional pixel arrays, which can be loaded into existing machine learning toolchains by extracting the pixel data across a set of images from a specific scan, and serializing them into a single, high dimensional array.

Critically, image data stored in the DICOM format needs to be converted to normalized pixel values using a Houndsfield transformation (Hounsfield Unit/CT Unit), which is a linear transformation utilized by radiologists to create a grayscale image where the denser tissue (which is more likely to have a cancerous nodule) appears brighter than lower density tissue (DenOtter & Schubert, 2023). Following this transformation, the scans are scaled in X and Y to match the size of the input layer of the network, usually 256 x 256, or a ½ scaling. This scaling was applied to cut down on the resources required to train the model, as it may not be possible to keep the full-sized image in memory during the training process due to memory usage constraints. This fully preprocessed array is then written back to disk in order to provide a cleaner way to repeatedly train off of the data while minimizing preprocessing overhead for each training iteration. Prior to preprocessing, we used the attached metadata file in the LIDC-IDRI dataset to focus on just thoracic CT scans, totalling 232 different scan samples. We applied a similar filter to images from the two other datasets that we used to obtain normal scan data, and used 25 normal scans to train our classifier. The model performance (including sensitivity and specificity results) are given in the next section.

When we begin training the model, these serialized arrays are then loaded off of the disk 256 x 256 x 64 sized array that can be fed into the first layer of the neural network. Following the input layer, we define a multi-layer neural network that uses the ReLU activation function

between each layer. At the output layer, we use a sigmoid function to map the inner layer outputs to a specific label, i.e. cancerous/non-cancerous. Training loss was then evaluated using a cross-entropy loss function, and the loss values were backpropagated across the whole network, similar to most supervised learning classifiers. All in all, this provides us with 1351873 parameters to train at various layers of the model.

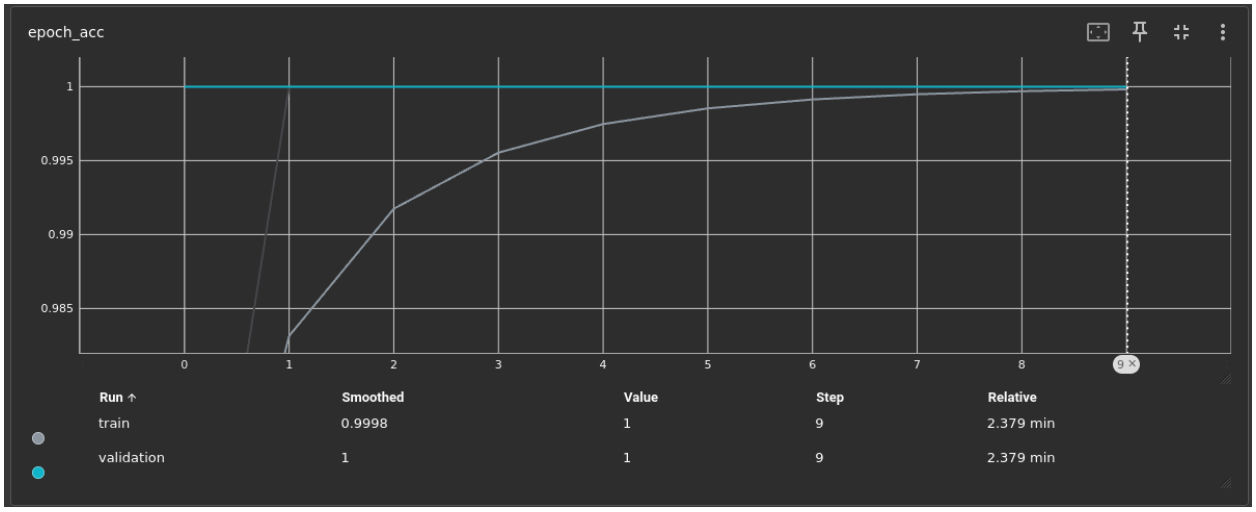

Figure 4: Training and validation accuracy per each epoch of training the 3D convolutional network, for 10 epochs

- **Survey**

In order to address the critical equity impact component of this project, our survey aims to address what contributes to healthcare disparities on such a large scale. The target population of this survey are elderly residents over the age of 60 in Prince George's County, Maryland. Participants are mainly recruited through administrative offices of recreation centers, nursing homes, and clinics, with flyers and tabling at places the elderly frequent (e.g. grocery stores, parks, community centers) to gather more respondents. Disseminating this survey within Prince George's County via Qualtrics captures a wide diversity of potential participants across race, socioeconomic status, and health care experience that allows us to gather an idea of their attitudes and behavior towards seeking out and obtaining healthcare for cancer-related issues. We

focus on questions relating to their experiences with healthcare, cancer detection, and treatments to examine what these populations know about cancer itself. This survey is inspired by the case of Henrietta Lacks, who not only lacked proper resources and care that could have maintained her health before the cancer occurred (treating syphilis) but also did not take steps to be proactive about her health and visit her physicians.

Prior literature has pointed to demographic differences as a catalyst for healthcare disparities. Therefore, we expect race, age, socioeconomic status, and location (urban vs. rural) to have considerable impacts on attitudes and beliefs regarding access to healthcare and cancer detection technologies (Tarver & Haggstrom, 2019). This survey deviates from the traditional surveys by directly asking respondents their opinions about healthcare and thus analyzing their accessibility to healthcare and related technologies against their utilization and understanding of it. We hypothesize that participants that avoid accessing healthcare resources may hold negative beliefs about the healthcare system and related technologies.

The survey contains 24 items that measure one of the three constructs of the research question; healthcare utilization; health literacy (specifically of cancer detection technologies), and beliefs about the healthcare system. Questions 1 -15 are demographic questions allowing us to summarize our respondent pool. Although a majority of items are ordinal or nominal, they will be re-coded as dummy variables in R once we gather demographic information to analyze how different groups within a demographic respond to modern healthcare. Race, gender, employment status, and health insurance coverage will be the main focus regarding demographics. Questions 16-24 directly measure healthcare utilization, health literacy, and attitudes regarding healthcare. Our main point of interest will be Question 17, an item that is constructed as both a list item and likert scale. Respondents will be asked how frequently they feel supported by providers, and

understand their personal diagnosis and treatments. This item will be re-coded to reflect a scale of 0-4 with “Never” as 0 and “Always” as 4. We will analyze this information through regression analyses to identify what variables are statistically significant in answering the question, “How do the cultural and social backgrounds of PG county residents aged 60+ influence their attitudes and beliefs about cancer technology and its use in cancer treatment?”.

Results

With the use of GradCAM, we aimed to generate visual insights into the significance of the weights associated with the last convolutional layer concerning the input image. Initially, our input array comprised dimensions of 256 x 256 x 64, which was progressively reduced through pooling layers to a final convolutional layer with dimensions of 28 x 28 x 4. Leveraging GradCAM, we computed gradients to produce saliency map outputs. These outputs were then overlaid as heatmaps on 2D slices extracted from the original input, assisting in the identification of regions of interest of the model's decision-making process. Illustrated below is an example extracted from a randomly selected cancerous patient scan. Notably, the visualization reveals heightened gradient weights within the lungs, predominantly concentrated in central areas. This observation underscores the model's focus on specific regions within the lungs, indicative of their significance in the classification process.

Saliency maps like this can be used by clinicians in the field to understand areas of interest within a particular scan. This can help ensure that clinicians are able to understand where exactly in the scan a potential nodule could be, and help ensure that our model does not act like a black box when used in a real-life setting.

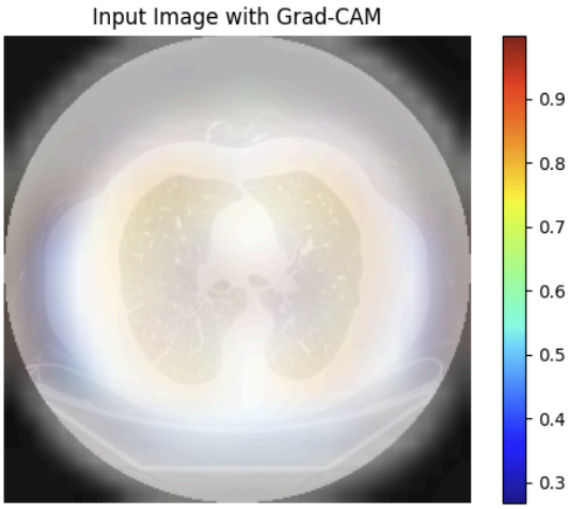

Figure 5: Example of saliency map

- **Model Statistical Evaluation**

In addition to our qualitative assessment using GradCAM, we also performed a manual evaluation of our model using a random sample of test images that were not previously seen by the model during training. This test set consisted of 34 cancerous and 6 normal scans, across a variety of patients. Given the general tendency to make diagnostic tests an overapproximation to err on the side of caution, we felt that it was important to focus on testing abnormal image samples.

Confusion Matrix	Predicted abnormal	Predicted normal	
Actual abnormal	29	5	Sensitivity: 0.853
Actual normal	0	6	Specificity: 1.000
			Accuracy: 0.875

In most cases, our model accurately classified the scan as normal or abnormal (indicating the presence of lung cancer nodules). We are continuing to use saliency techniques, as well as different hyperparameters during the training process to improve the sensitivity and ensure that

the model does not continue to misclassify image samples, particularly in cases where the image shows evidence of cancer. Full results of our manual analysis of the model can be found in the appendix.

NoduleMNIST3D method:	Accuracy:
ResNet-18+2.5D	0.835
ResNet-18 +3D	0.844
ResNet-18 +ACS	0.847
ResNet-50 +2.5D	0.848
ResNet-50 +3D	0.847
ResNet-50 +ACS	0.841
auto-sklearn	0.874
AutoKeras	0.834
Average Accuracy:	**0.846**

The average accuracy of NoduleMNIST3D was 0.846 across the 8 training methods that were used. In comparison, our model had an accuracy of 0.875. However, it is important to note that although the accuracy values seem similar between the two models, NoduleMNIST3D was tested using 310 scans while our model was tested using 40 total scans.

Team Structure

The subteams our team has split into are AI/Machine learning, Data and Cancer. This was done in an attempt to highlight the strengths and backgrounds of individual members while fostering a learning environment within the team that provides an opportunity for different perspectives.

The AI team was responsible for building and improving our model that our team decided on. This also includes working with the chosen technologies, using the UMIACS cluster, and building our network.

The Data subteam was responsible for data compilation, cleaning, and analysis. This includes finding the best dataset to use for our given model. Also, including the final model results and analyzing the implications of those results.

The Cancer subteam was responsible for cancer specific research and the equity impact of our research as well. The Cancer subteam is integral to providing accurate information about the latest in lung cancer detection methods and contributing both biological and equity based perspectives on the progression of the project. This involves research into lung cancer and the various characteristics that allow us to optimize the machine learning model to target certain biological markers in our data. Furthermore, the Cancer subteam was responsible for headlining the development of our survey. This includes reaching out to audiences impacted by our research to gauge their opinion and views.

Team Data was also involved in much of the research and initial findings, specifically working on the preliminary data. Through the research steps, we narrowed down the datasets that seemed the most relevant and beneficial to our research. After narrowing down to five datasets, we continued to evaluate the pros/cons of each dataset to decide on one specific one to use for our model. This involved meeting with Team Cancer to talk about distinguishing and important factors in these datasets. We used ground-truth metadata from each dataset to help establish a baseline comparison, which can be used in order to determine if our model is more or less accurate than the current standard.

Through our data analysis and training models, we aimed to optimize a machine learning model that can make improved diagnoses and predictions for non-small cell lung cancer (NSCLC). By enhancing the diagnostic for NSCLC, we hope to better the NSCLC survival rates and contribute to a finer quality of life for those with NSCLC. By surveying underserved

populations, we expect to quantify the need for our machine learning model in such communities. The statistics in the literature discussing cancer detection rates as well as the frequency/quality of doctor's visits in these communities demonstrate that there is an apparent need for our research. Our model in tandem with survey data aims to address these concerns, but also highlight why these concerns must be addressed.

For years, there has been a clear distinction between the accessibility and the quality of healthcare in wealthy and represented communities versus communities that are underserved and underrepresented. This is due to a long string of issues and while we would love to address and work on each one, time and budget constraints have encouraged us to focus on a specific subfield: imaging detection. For those in communities where consistent doctor's visits are difficult to maintain and where healthcare institutions themselves lack the resources to provide the best care possible, late cancer detection and diagnosis is often what results in treatable cancer becoming deadly. Having an enhanced model that can efficiently improve detection and diagnosis as well as make informative predictions for physicians and patients of concern would address this issue, serving as a foundation for future cancer detection research. Through our research we aimed and continue to aim to create this model in the most cost-effective way possible to ensure the feasibility of providing underserved communities with such resources; furthermore, it would allow our model to be replicable, opening up the door to further improvement and enhancement.

Equity-Impact Report

After conducting literature review of prior research across NSCLC, available datasets, and current diagnostic methods, many disparities were identified. In response to disparities in healthcare outcomes, Team DOC has undertaken various measures to ensure that our technology is accessible and utilized fairly by all. Firstly, we plan to incorporate saliency methods in our machine learning. These produce visual representations that appear as heat maps of the significant features of medical images that indicate potentially cancerous regions. Not only do saliency maps act as guides for healthcare professionals, but they also increase explainability of the machine learning model so that healthcare professionals can better understand how the model makes its predictions. It allows them to better visualize the model's thought process and understand the complex decision-making process, which furthers collaboration between AI systems and professionals and ultimately improves patient care. This understanding not only increases accuracy of these models but it also decreases bias. Because saliency maps allow healthcare experts to visualize the decision-making process of our machine learning model, those using the model can evaluate whether the model leads to discriminatory outcomes.

Moreover, we have taken steps to mitigate the costs of our technology. This model can be deployed cheaply on current computing infrastructure, thus requiring only an internet connection to get high quality detection results. We also plan to use a CT scanner with internet connection so that our technology can be used for telemedicine practices. This will increase the accessibility of our machine learning model, so that those who are unable to physically be in spaces with our technology can still have access to it.

References

Abedi, I., Vali, M., Bentolhoda Otroshi, Zamanian, M., & Hamidreza Bolhasani. (2024). HRCTCov19-a high-resolution chest CT scan image dataset for COVID-19 diagnosis and differentiation. *BMC Research Notes*, *17*(1). https://doi.org/10.1186/s13104-024-06693-z

Alakwaa, W., Nassef, M., & Badr, A. (2017). Lung cancer detection and classification with 3D convolutional neural network (3D-CNN). *Lung Cancer*, *8*(8), 409.

Alberg, A. J., Brock, M. V., Ford, J. G., Samet, J. M., & Spivack, S. D. (2013). Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest, 143(5 Suppl), e1S–e29S. https://doi.org/10.1378/chest.12-2345

Ardila, D., Kiraly, and A.P., Bharadwaj (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine, 25*, 954-961.

American Cancer Society (2021). Lung Cancer Survival Rates. *American Cancer Society.*

American Cancer Society. (2021, February 2). *Cervical cancer survival rates: Cancer*

5 year survival rates. American Cancer Society. Retrieved December 7, 2021, from https://www.cancer.org/cancer/cervical-cancer/detection-diagnosis-staging/survival.html.

American Cancer Society. (2021, January 12). *Lung cancer statistics: How common is*

lung cancer? American Cancer Society. Retrieved December 7, 2021, from https://www.cancer.org/cancer/lung-cancer/about/key-statistics.html.

Barta JA, Powell CA, Wisnivesky JP. (2019). Global Epidemiology of Lung Cancer. *Ann Glob Health*. ;85(1):8. Published 2019 Jan 22. doi:10.5334/aogh.2419

Best, A. L., Vamos, C., Choi, S. K., Thompson, E. L., Daley, E., & Friedman, D. B. (2017). Increasing Routine Cancer Screening Among Underserved Populations Through Effective Communication Strategies: Application of a Health Literacy Framework. *Journal of Cancer Education : The Official Journal of the American Association for Cancer Education*, *32*(2), 213–217. https://doi-org.proxy-um.researchport.umd.edu/10.1007/s13187-017-1194-7

Commean, P. (2021, January 7). *RIDER lung CT*. Wiki - The Cancer Imaging

Archive
(TCIA) Public Access - Cancer Imaging Archive Wiki. Retrieved October 26, 2021, from https://wiki.cancerimagingarchive.net/display/Public/RIDER+Lung+CT#225127323e88f9778a7d49899473a2a633b68033

Data from NSCLC-Radiomics-Genomics. The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. (n.d.). Retrieved October 25, 2021, from https://wiki.cancerimagingarchive.net/display/Public/NSCLC-Radiomics-Genomics.

DenOtter TD, & Schubert J. (2023). Hounsfield Unit. *StatPearls*. https://www.ncbi.nlm.nih.gov/books/NBK547721/

Detterbeck, F. C., Mazzone, P. J., Naidich, D. P., & Bach, P. B. (2013). Screening for

lung cancer. *Chest*, *143*(5). https://doi.org/10.1378/chest.12-2350

Ellis, Peter M, and Rachel Vandermeer. (2011). Delays in the diagnosis of lung cancer. Journal of thoracic disease vol. 3,3: 183-8. doi:10.3978/j.issn.2072-1439.2011.01.01

Fass L. (2008). Imaging and cancer: a review. Molecular oncology, 2(2), 115–152. https://doi.org/10.1016/j.molonc.2008.04.001

Global Burden of Disease Cancer Collaboration, (2015). The Global Burden of Cancer 2013. JAMA oncology, 1(4), 505–527. https://doi.org/10.1001/jamaoncol.2015.0735

Hamouda, Saeed & Ezz, Reda & Wahed, Mohammed. (2017). Enhancement Accuracy of Breast Tumor Diagnosis in Digital Mammograms. Journal of Biomedical Sciences. 06. 10.4172/2254-609X.100072.

Kirby, J. (2021, February 9). *NSCLC-radiomics*. Wiki - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. Retrieved October 26, 2021, from https://wiki.cancerimagingarchive.net/display/Public/NSCLC-Radiomics

Liao, Zhijun, et al. “Cancer Diagnosis Through IsomiR Expression with Machine Learning Method.” *Current Bioinformatics*, vol. 13, no. 1, 2018, pp. 57–63. *Crossref*, doi:10.2174/1574893611666160609081155.

Lung cancer statistics: How common is lung cancer? American Cancer Society. (n.d.). Retrieved December 5, 2021, from https://www.cancer.org/cancer/lung-cancer/about/key-statistics.html.

Mayo Foundation for Medical Education and Research. (2021, April 27). *Cancer.* Mayo Clinic. Retrieved December 8, 2021, from https://www.mayoclinic.org/diseases-conditions/cancer/diagnosis-treatment/drc-20370594.

Midthun D. E. (2016). Early detection of lung cancer. F1000Research, 5, F1000 Faculty Rev-739. https://doi.org/10.12688/f1000research.7313.1

Mazières, J., Pujol, J. L., Kalampalikis, N., Bouvry, D., Quoix, E., Filleron, T., Targowla, N., Jodelet, D., Milia, J., & Milleron, B. (2015). Perception of lung cancer among the general population and comparison with other cancers. Journal of thoracic oncology : official publication of the International Association for the

Study of Lung Cancer, 10(3), 420–425. https://doi.org/10.1097/JTO.0000000000000433

Minna JD, Roth JA, Gazdar AF. (2002) Focus on lung cancer. Cancer Cell: Volume 1. Cell Press, http://wxjs.chinayyhg.com/upload/Files/Cancer-Cell/2002-(Volume-1)/Issue-1-(1-108)/49-52.pdf

Nasim, Md Abdullah & Shah, Faisal & Hossain, Tonmoy & Ashraf, Mohsena & Shishir, Fairuz. (2019). Brain Tumor Detection using Convolutional Neural Network. 10.13140/RG.2.2.15562.52163.

National Cancer Institute. (2021, May 5). *What is cancer?* National Cancer Institute. Retrieved December 7, 2021, from https://www.cancer.gov/about-cancer/understanding/what-is-cancer.

O'Shea, K., & Nash, R. (2015). *An Introduction to Convolutional Neural Networks* (arXiv:1511.08458). arXiv. http://arxiv.org/abs/1511.08458

Patel A. Benign vs Malignant Tumors. *JAMA Oncol.* 2020;6(9):1488. doi:10.1001/jamaoncol.2020.2592

PDQ® Adult Treatment Editorial Board. PDQ Non-Small Cell Lung Cancer Treatment. Bethesda, MD: National Cancer Institute. Updated 2021. Available at: https://www.cancer.gov/types/lung/hp/non-small-cell-lung-treatment-pdq. . [PMID: 26389304]

Purandare, N. C., & Rangarajan, V. (2015). Imaging of lung cancer: Implications on staging and management. The Indian journal of radiology & imaging, 25(2), 109–120. https://doi.org/10.4103/0971-3026.155831

Selvaraju, R. R.; Cogswell, M.; et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.

TensorFlow Developers. (2024). *TensorFlow* (v2.15.1) [Computer software]. [object Object]. https://doi.org/10.5281/ZENODO.4724125

U.S. National Institute Of Health, National Cancer Institute. SEER Cancer Statistics Review, 1975–2015.

Vendt, B. (2021, May 24). *Lidc-idri*. Wiki - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. Retrieved October 26, 2021, from https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI resolved

Tarver, W. L., & Haggstrom, D. A. (2019). The Use of Cancer-Specific Patient-Centered Technologies Among Underserved Populations in the United States: Systematic Review. *Journal of Medical Internet Research*, *21*(4), e10256. https://doi-org.proxy-um.researchport.umd.edu/10.2196/10256

Wang, X., Chen, H., Gan, C., Lin, H., Dou, Q., Tsougenis, E., Huang, Q., Cai, M., & Heng, P. (2020). Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis. *IEEE Transactions on Cybernetics, 50*, 3950-3962.

W. De Wever, J. Coolen, J.A. Verschakelen (Jun 2011). Imaging techniques in lung cancer. *Breathe.*7 (4) 338-346; DOI: 10.1183/20734735.022110

Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., & Ni, B. (2023). MedMNIST v2—A large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Scientific Data*, *10*(1), 41. https://doi.org/10.1038/s41597-022-01721-8

Zhu, W., Liu, C., Fan, W., & Xie, X. (2018). DeepLung: Deep 3D Dual Path Nets for Automated Pulmonary Nodule Detection and Classification. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 673–681. https://doi.org/10.1109/WACV.2018.00079

Zhou, B.; Khosla, A.; et al. Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2921–2929.

Zunair, H., Rahman, A., Mohammed, N., & Cohen, J. P. (2020). *Uniformizing Techniques to Process CT scans with 3D CNNs for Tuberculosis Prediction* (arXiv:2007.13224). arXiv. http://arxiv.org/abs/2007.13224

Appendix

I: Test Results

File Name (cancerous datasets):	Normal (percent confidence)	Abnormal (percent confidence)
100.000000-NA-47868-patient-0215.npy	15.67	84.33
1613.000000-NA-96652-patient-0201.npy	8.26	91.74
2.000000-NA-34236-patient-0291.npy	56.87	43.13
3.000000-NA-92139-patient-0204.npy	37.98	62.02
3000514.000000-NA-56951-patient-0057.npy	29.85	70.15
3000525.000000-NA-47372-patient-0020.npy	17.95	82.05
3000568.000000-NA-87015-patient-0041.npy	24.06	75.94
3000629.000000-NA-31276-patient-0043.npy	9.79	90.21
3000641.000000-NA-56733-patient-0112.npy	18.86	81.14
3000670.000000-NA-70157-patient-0102.npy	4.14	95.86
3000705.000000-NA-92959-patient-0111.npy	39.85	60.15
3000709.000000-NA-50417-patient-0010.npy	7.97	92.03
3000712.000000-NA-13937-patient-0056.npy	8.14	91.86
3000720.000000-NA-12443-patient-0030.npy	2.21	97.79
3000774.000000-NA-03667-patient-0092.npy	30.42	69.58
3000776.000000-NA-23312-patient-0121.npy	28.56	71.44
3000908.000000-NA-34320-patient-0055.npy	1.63	98.37
3000938.000000-NA-36565-patient-0031.npy	1.61	98.39
3000927.000000-NA-19268-patient-0015.npy	7.98	92.02
3001578.000000-NA-15293-patient-0082.npy	2.61	97.39
3001597.000000-NA-31510-patient-0058.npy	2.23	97.77
3001608.000000-NA-25501-patient-0092.npy	7.99	92.01
30083.000000-NA-95222-patient-0075.npy	35.78	64.22
3185.000000-NA-66614-patient-0150.npy	58.89	41.11
4.000000-NA-15280-patient-0241.npy	29.99	70.01
4.000000-NA-58748-patient-0240.npy	33.52	66.48
4.000000-Recon 3-02075-patient-0088.npy	8.87	91.13
5408.000000-NA-86657-patient-0151.npy	53.66	46.34
5591.000000-NA-44791-patient-0107.npy	62.46	37.54
5597.000000-NA-08531-patient-0104.npy	58.68	41.32
3210.000000-NA-19980-patient-0084.npy	11.67	88.33
4.000000-NA-00450-patient-0196.npy	61.62	38.38
4.000000-Recon 3-88650-patient-0072.npy	7.34	92.66
5.000000-NA-47975-patient-0216.npy	30.93	69.07
Averages:	24.06	75.94

File Name (normal datasets):	Normal (percent confidence)	Abnormal (percent confidence)
020-patient-ge497g00.npy	78.25	21.75
021-patient-ge497g00.npy	82.71	17.29
022-patient-ge497g00.npy	80.17	19.83
023-patient-ge497g00.npy	93.49	6.51
(203) 0.000000-CT141836RC	61.87	38.13
(204) 0.000000-CT135949Re	71.30	28.70
Average:	77.97	22.04

Confusion Matrix:	Predicted Abnormal	Predicted Normal		
Actual Abnormal	29	5	**Sensitivity:**	0.853
Actual Normal	0	6	**Specificity:**	1.000
			Accuracy:	0.875

Final Model Tests

II: Survey- Full Text of the Questions and Responses

1. Consent to Participate
   a. Yes
      i. If you have selected “Yes”, please click next to continue.
      ii. Here is a downloadable version of the consent form
   b. No
      i. You may stop the survey here.

2. Biological Sex
   a. Male
   b. Female
   c. Prefer not to answer

3. Gender Identity
   a. Man

b. Woman
c. Genderqueer/Non-binary
d. _____ (fill in the blank)
e. Prefer not to answer

4. Are you of Hispanic, Latino, or Spanish origin?
a. No, not of Hispanic, Latino, or Spanish origin
b. Yes, Mexican, Mexican Am., Chicano
c. Yes, Puerto Rican
d. Yes, Cuban
e. Yes, another Hispanic, Latino, or Spanish origin
f. Prefer not to answer

5. What race/ethnicity do you identify as?
a. White
b. Black or African American
c. Asian
d. Hispanic or Latino/Latina
e. Native American or Alaska Native
f. Native Hawaiian or Pacific Islander
g. Middle Eastern or North African
h. Multiracial or Biracial
i. Other
j. Prefer not to answer

6. Highest level of education
a. Some high school or less
b. High school graduate or equivalent
c. Other post high school vocational training
d. Completed some college, but no degree
e. Associate's degree
f. Bachelor's degree
g. Master's or professional degree
h. Doctorate degree
i. None of the above

7. What is your current annual household income?
a. Less than $30,000
b. Between $30,000 and $60,000
c. Between $60,000 and $100,000

d. Over $100,000
e. Prefer not to answer

8. Employment status
a. Employed full-time
b. Employed part-time
c. Unemployed and looking for work
d. Unemployed and not looking for work
e. Retired
f. Prefer not to answer

9. If you are retired, how many years has it been since you stopped working full-time or part-time for pay?
a. 1 to 5 years
b. 5 to 10 years
c. 10 to 15 years
d. 15+ years
e. Prefer not to answer
f. Not applicable

10. How long have you been in the workforce?
a. 0-20 years
b. 20-40 years
c. 40+
d. Prefer not to answer
e. Not applicable

11. Do you currently have health insurance?
a. Yes, I do
b. No, I do not

12. Were you without health insurance for any amount of time in the past 12 months?
a. Yes, I was
b. No, I was not

13. Who pays for your health insurance?
a. Local government
b. Current employer
c. National government
d. Former employer

e. State government
f. Self funded
g. Prefer not to answer
h. Not applicable

14. In which state do you permanently reside?
a. Drop down box

b. If Maryland, in which county do you reside?
i. Drop down box

15. Please confirm that you are over the age of 60.
a. Yes, I am over the age of 60

---

16. How long has it been since you last saw or talked to a doctor/healthcare provider about your own health?
a. A few days
b. A few weeks
c. A few months
d. A few years
e. More than 5 years

17. Please indicate the frequency with which you have experienced any of the following events. (Never, Rarely, Occasionally, Sometimes, Often).(Trust, Understanding Results, Cost, Transparency)

	Never	Rarely	Occasionally	Sometimes	Often
1. At the end of a doctor visit, I understand what the doctor has discussed with me.					
2. I consider/research healthcare prices prior to using a healthcare service or provider.					

3. My concerns are acknowledged throughout a doctor visit.					
4. My needs are put first by the physician.					
5. The doctor is interested in making me feel comfortable as a patient.					
6. It is difficult to make time to travel to a specialty healthcare service or provider					
7. Visiting the doctor makes me feel nervous.					
8. The physician provides clear and thorough explanations throughout my visit.					
9. The doctor provides visual aids that are helpful to my understanding.					
10. The physician uses simple language to break down medical terminology.					
11. I feel I am given enough information about medical interventions (such as procedures, medications, referrals, etc..) to provide informed consent.					

17. Have you heard about cancer technology?
1. Yes
2. No
3. Not Sure

18. Have you ever used cancer technology before?
1. Yes
2. No
3. Not Sure

19. If you have used cancer technology before, did you find it helpful in your treatment?
1. Yes, definitely
2. Yes, somewhat
3. Not sure
4. No, somewhat
5. No, definitely not

20. To what extent do you believe cancer technology can be helpful in the detection of cancer?
1. Not at all
2. To a slight extent
3. Moderately
4. To a strong extent
5. Very strong extent

21. How important is it for you to have access to the latest cancer technology?
1. Very important
2. Somewhat important
3. Neutral
4. Not very important
5. Not at all important

22. Do you think cancer technology is too expensive?
a. Yes, definitely
b. Yes, somewhat
c. Not sure
d. No, somewhat
e. No, definitely not

23. Do you feel that cancer technology is accessible to everyone who needs it?

1. Yes, definitely
2. Yes, somewhat
3. Not sure
4. No, somewhat
5. No, definitely not

24. Do you think cancer technology can help improve the quality of life for cancer patients?

1. Yes, definitely
2. Yes, somewhat
3. Not sure
4. No, somewhat
5. No, definitely not