Detection and Measurement of Lymph Nodes Using Artificial Intelligence

Authors: Nathan Patton, Kaya Oguz, Daniel Song

Word count: 2692

Figure(s): 7

Table(s): 0

Equations(s): 0

Supplements(s): 2

References(s): 13

Advisor Approval:_____Date___05/05/2024_____

Xue Feng, Department of Biomedical Engineering

Detection and Measurement of Lymph Nodes Using Artificial Intelligence

Nathan Patton[a,1], Kaya Oguz[a,2], Daniel Song[a,3], Xue Feng[b]

[a] UVA Department of Biomedical Engineering
[b] Carina Medical
[1] Correspondence: njp2cyf@virginia.edu, (703) 479-6913
[2] Correspondence: klo9xq@virginia.edu, (703) 994-3338
[3] Correspondence: djs2jr@virginia.edu, (703) 615-5642

Abstract

Lymphoma is a type of blood cancer, affecting roughly 80,000 individuals annually in the United States. However, the current method of lymph node examination for lymphoma diagnosis is manual and relies on the radiologist's discretion. To this end, an introduction of a more automated method of scan analysis would allow the radiologists more time to review the scan as a whole, reducing fatigue and error rates. This work first aimed to develop an algorithm that will quickly process the input scan and output the results of the analysis, showing the numerical change in tumor burden as well as what was identified as tumors. The finalized interface will require minimal steps for implementation, allowing a seamless integration into the radiologists existing workflow. Additionally, the primary objective of any medical tool or technology is to ensure patient safety. Thus, assessing the safety of this tool, while minimizing potential risks, such as misdiagnosis, and enhancing the overall well-being of patients was another aim. To begin, methods for malignancy classification were identified, and machine learning capabilities and radiomics were explored. Various EfficientNet models, a Keras image classification model, were trained, utilizing lymph node data, while also identifying potential features for malignancy classification via radiomics, an image feature extraction method using data characterization algorithms. Radiomics analysis was performed on features and integrated into scikit-learn machine learning models to be trained and tested. 2D EfficientNet models hold great potential in breast cancer classification, but improvements need to be made before considering it reliable. For lung nodule testing, our 2.5D model achieved 77% validation accuracy. Our model using radiomics analysis achieved 79% testing accuracy, identifying substantial features for lung nodule classification. These results demonstrate the potential of artificial intelligence in malignancy classification, and lays the framework for its application in clinical practice.

Keywords: Radiology, Artificial Intelligence, Lymph Nodes, Malignancy Classification

Introduction

Significance

Cancer is a group of diseases caused by the uncontrolled growth and spread of irregular cells that can ultimately lead to death if not treated properly. In the United States, cancer is the second highest cause of death with heart disease being the leading cause. The American Cancer Society has estimated that in the US alone, approximately 2 million cancer cases are expected to be diagnosed in 2024 while over 610,000 cancer-related deaths are expected[1].

Lymphoma is a type of blood cancer that starts in the white blood cells. It is distinguished as a heterogeneous group of malignant tumors and is characterized by the abnormal proliferation of mature lymphoid cells[2]. The abnormal proliferation leads to complications within the lymphatic system, compromising the defense against harmful substances and leading to serious illness. The most important contributor to the system's immune defense against these substances are lymph nodes (Figure 1). Lymph nodes play a vital role within the body's immune system, filtering substances which travel through the lymphatic fluid. Figure 1 depicts the lymph node structure, where lymphatic fluid enters the node through the afferent

lymphatic vessels and exits through the efferent lymphatic vessel. There are hundreds of lymph nodes throughout the human body which are connected to one another, with each node containing lymphocytes that constantly help fight infection and various diseases by destroying harmful substances[3]. Lymph nodes can become enlarged due to several reasons, including infection or malignancy. Recognizing processes in which lymph nodes increase in size is crucial for the understanding of lymphoma pathology and diagnosis. Enlarged lymph nodes can also display unique characteristics in lymphoma since they can be rubbery, painless, and may feel fixed in their position. The size and location of swollen nodes can vary based on the type and stage of lymphoma. So, depending on the stage of cancer, time frame, and dimensions of the node, it can have a profound effect on the process of diagnosis and results.

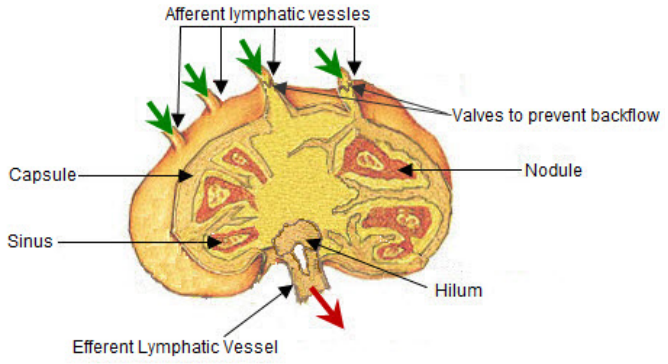

Figure 1. The figure above is a diagram of a lymph node. Green arrows indicate flow of lymphatic fluid entering the lymph node, with valves to keep fluid direction constant. The red arrow shows flow of lymphatic fluid exiting the lymph node[4].

Current Detection and Measurement Methods

A variety of indicators point towards the development of lymphoma, a common one being enlargement. Identification and assessment of the change in tumor burden (size) is considered an important feature in the clinical evaluation of cancer therapeutics, however, there is still room for improvement. Imaging methods for finding these indicators include utilizing X-rays, computed tomography, magnetic resonance imaging, ultrasound and positron emission tomography. Radiologists and certified clinicians use these tools to recognize different cell types by looking at the shape, size, orientation, and arrangement of cells within lymphoid tissue. Furthermore, tissue samples from biopsies assist in cancer diagnosis and lymphoma subtype determination. More specifically, radiologists measure lymph nodes using the length, volume, width, and thickness as an indicator for diagnosis in combination with other indicators. Response Evaluation Criteria in Solid Tumors (RECIST) is the current methodology to evaluate and assess changes in tumor burden. It is a guideline that provides a standard approach for solid tumor measurement and provides definitions for the assessment of tumor size changes[5]. To calculate the change in tumor burden between images from distinctive time points, the dimension of the select lesions, referred to as target lesions, are used. This calculation is then categorized as complete response (CR), partial response (PR), stable disease (SD), or progressive disease (PD). CR is complete disappearance of tumor (-100%), PR is a change between -100% and -30%, SD is a change between -30% and +20%, and PD is an increase of 20% or greater[6]. Currently, radiologists must manually measure and analyze these changes in order to determine the tumor burden.

Issues with Current Methods

Tumor burden identification is entirely dependent on the radiologists experience and ability, leaving space for human error as shown in a study which cites a day-to-day average error rate of 3-5% as well as a retrospective error rate of 30% from radiological studies[7]. In the United States alone, diagnostic errors are responsible for roughly 40,000 to 80,000 deaths annually[8]. In addition, close to 75% of medical malpractice claims against radiologists are associated with diagnostic errors[9]. Even with these errors, The average cost for lymphoma treatment of these complications can reach $200,000 for patients without health insurance[10]. This current process is also heavily time consuming. A study by Alexander et. al. states that an average MRI scan generates over 1000 images and within a normal work day, leaving radiologists with less than 1 second to read each image[11].

Solution

Early detection of lymphoma can significantly increase the chances of successful treatment and improve longitudinal quality of life. Our solution utilized an AI-based tool to automate the image analysis process and provide for the radiologist to view. This would enhance the current existing workflow by decreasing the amount of manual analysis required and increasing the amount of time radiologists can analyze each set of scans, resulting in a more thorough analysis. This work has the potential to be integrated into any radiologist clinical workflow. In addition, this work would advance the field of radiology by: 1) Improving the efficiency of lymph node detection while matching or improving the results of current

methodologies, 2) Improving the quality of life for radiologists, reducing mental fatigue and exhaustion from manual image analysis, and 3) Enhancing the existing clinical workflow and setting a new standard for medical image analysis. Specifically, there were two major aims: 1) Develop an AI-based lymph node detection measurement tool and 2) Evaluate the safety and effectiveness of the tool in a radiological workflow.

Results

2D Network Structure for Breast Cancer Application

The results of the EfficientNetB0 model are poor, yet promising. As seen in the confusion matrix shown in Figure 2A, the true negatives are high, which is good. The false positives are also relatively high, which is not great, but most importantly, the number of false negatives is too large (29) and true negatives is low (3). The model failed to predict 29 out of the 32 total cancer patients, which is not acceptable. Additionally, as seen in Figure 2B, the accuracies are high, but since the data is quite imbalanced, metrics such as F1 score are more important. The high F1 scores for the normal class and low F1 scores for the cancer class show that the model is much better at predicting the normal class, which further supports the findings from the confusion matrix.

Similarly, when looking at the results from the EfficientNetB5 model, notice how it is predicting all cases to be normal, as seen in the confusion matrix shown in Figure 2C. Therefore, it failed to predict all cases of cancer, which is debatably worse than the first model with several false positives. Figure 2D might make it seem like the model performs well in classifying this data due to the high accuracy. However, the additional metrics say otherwise. The validation and test loss are both high, 0.631 and 0.632 respectively. Breaking it down further, the model performs very well for the normal class. The precision is approximately 97.70%, recall is 100%, and F1 score is 98.84%. For the cancer class, the precision, recall, and F1 score are all 0, suggesting that the model failed to correctly identify any instances. Thus, the overall F1 score is 0, indicating that the model is poor at identifying the minority class (cancer).

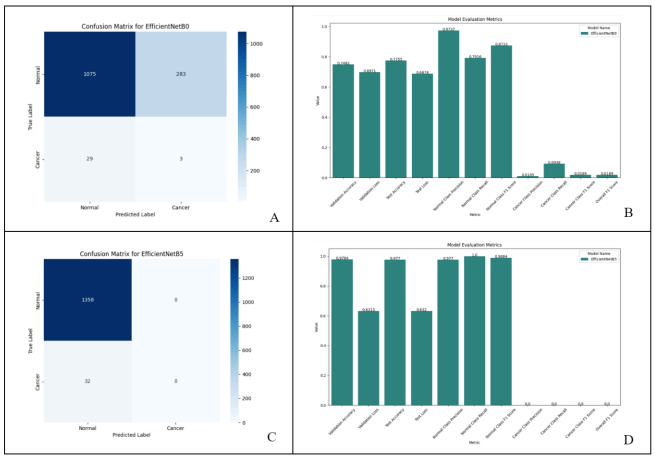

Figure 2. A) The confusion matrices for EfficientNetB0 B) Evaluation metrics for EfficientNetB0 C) The confusion matrices for EfficientNetB5 D) Evaluation metrics for EfficientNetB5

2.5D Network for Lung Nodule Classification

This task aimed to develop a 2.5D model for malignancy classification. After preprocessing steps were taken for lung nodule data to create the training data, the EfficientNet model was trained and performance was monitored. K-folds cross validation was utilized when evaluating model performance due to the minimal number of samples available for model training. K-folds cross validation was performed for five folds, with the best model performance for each fold being shown in Figure 3. Validation accuracy is the metric which displays the performance of the model when tested on unseen data. Five folds of K-folds cross validation displayed an average validation accuracy score of 77%. The highest validation accuracy of 87% was displayed in fold one 52% while the lowest was observed in fold four.

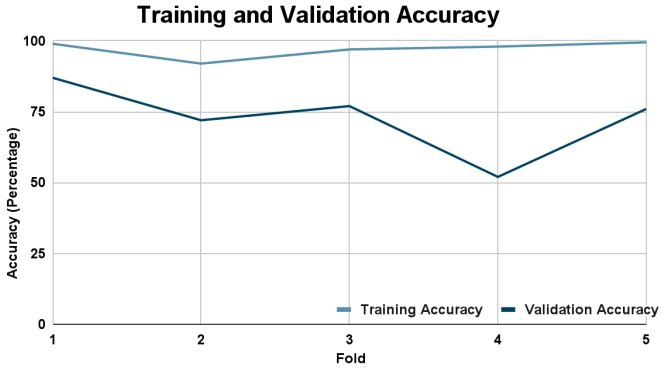

Figure 3. Training and Validation accuracy results after each fold. Average training percentage was 97% while average validation percentage was 77%.

Radiomics Analysis for Lung Nodule Classification

In this task, radiomic image feature testing of lung nodule datasets was explored. Preprocessed Lung Nodule Data from 2.5D Network Model approach were evaluated using radiomics to calculate various image features such as voxel volume, least axis length, gray level emphasis, and surface area. Threshold testing was performed for each image feature to manually classify lung nodules. Plotting of Receiver Operating Characteristic (ROC) curves determined the best image features for classification, as shown in Figure 4. Run entropy ended up being the most accurate predictor of lung nodule type with an area under the curve of approximately 0.690. Large Dependence Emphasis and Low Gray Level and High Gray Level features were ranked lowest in accuracy of detection with an area of about 0.660. It's also worth noting the minimal difference of 0.03 between areas for the most and least accurate feature, indicating that altering extracted features is insensitive in predictive performance. These radiomics image features were extracted and integrated into logistic regression and random forest models to be trained and tested. Unfortunately, the random forest model was only able to reach an accuracy of roughly 65%. On the other hand, the logistic regression model showed greater promise, yielding an accuracy of approximately 79%.

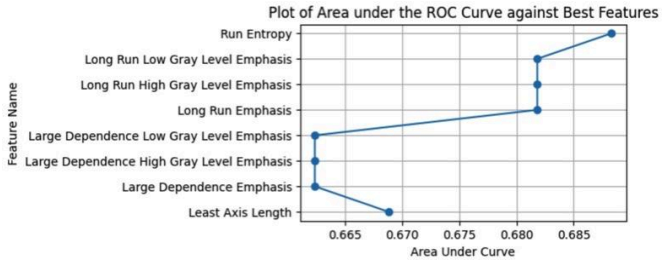

Figure 4. Area under Curve for Image Features used in Logistic Regression Model indicating detection accuracy. The y-axis indicates the feature name and x-axis indicates the area under the curve extracted from ROC curves.

Safety and Effectiveness in Radiological Workflow

This project aimed to monitor and evaluate the safety and effectiveness of a tool developed in Aim 1. Due to the sequential nature of Aims 1 and 2, implementation of a tool within a radiological workflow was not conducted because the tool proposed in Aim 1 was not entirely completed. Thus, Aim 2 is a piece of the future work that needs to be done.

Discussion

The 2D network breast cancer application indicated incomplete results. The EfficientNetB0 model suggested that it has significant potential to be utilized in classification, but there are still several opportunities for improvement. Additional data, as well as refining the model architecture, hyperparameter tuning, and optimizing the training strategies, are important things to consider when working with this model in the future. The EfficientNetB5 model's inability to identify any instances of the cancer class is a significant issue. Similar to the other model, there are improvements that can be made. Due to the increased complexity of this model, the training time (number of epochs) might need to be increased. Allowing the model to train longer could potentially increase the learning of the model, preventing it from predicting just the normal class and achieving more reliable evaluation metrics.

The 2.5D model task of the project showed an average validation accuracy of 77% which was achieved by K-folds cross validation. This result is likely due to the small dataset available for model training as research states that model performance is directly related to training set size[8]. The limitation to the dataset was due to the lack of time for the data to be formatted and shared with the team. Future studies should explore methods that allow for high model performance with smaller datasets. Additional steps in data augmentation can increase the quantity of data available for model training. In addition, it is encouraged to continue using K-folds cross validation or a similar method as the procedure is commonly utilized in analysis involving small sample sizes.

The task using a radiomics approach showed a testing and validation accuracy of 79%. Since this task utilized the same data set for training and testing as the 2.5D network model task, the model accuracy can be attributed for the same reasons such as a limited data set and the time constraint. In the future, more research and testing should be conducted on various feature selection and modeling using different methods for the logistic regression model. For example, procedures like recursive feature elimination and deep learning-based feature extraction can alleviate issues with low training accuracy[12]. Furthermore, integrating more layers for model design and implementing more image features can give the model further predictive parameters and improve the classification performance of the model.

Next Steps

To reiterate the recommended steps mentioned for each task previously, a larger quantity of data should be collected and utilized to train each task in order to improve model performance. Afterwards, it is recommended that all three tasks are incorporated into the development of a model for lymph node detection and measurement. Incorporating a balanced model, identifying ideal features for malignancy detection, and implementing an effective method for training and model development will result in a ready-model to be implemented within a clinical workflow.

To evaluate the safety and effectiveness of the AI tool in radiology, a retrospective study should be conducted. The hybrid design of a retrospective approach involves selecting a group of patients who have already been exposed to a radiological tool (MRI and/or CT), and whose lymph nodes have been manually detected and measured using the preexisting method (control). The results of the existing process should then be compared to those using the experimental AI tool. The outcomes and safety measures for both the control and experimental tools should be compared, such as diagnostic accuracy, patient outcomes, or safety profiles. The null hypothesis used for this experiment would be that there is no difference between the two diagnostic techniques. Statistical analysis should be performed to determine if there is a significant difference in analysis between the two techniques. Based on the statistical results, a cost-effectiveness analysis can be performed to evaluate the financial implication of using the experimental tool.

Materials and Methods

Data Description

Breast and lung lymph nodule scans were provided by Xue Feng and Carina Medical. The breast cancer nodule data consisted of over 54,706 images where each nodule was isolated as multiple 2D dcm images. See Figure 5 to the right for training images for both a normal patient and a cancer patient. Additional data, including patient id, views, lateralities, age, and class labels (benign or malignant) were provided in a csv (Figure S2), which was later mapped to the images.

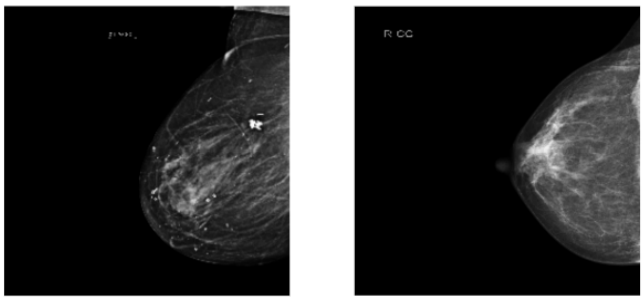

Figure 5. Shown here are two dcm breast images. The cancerous nodule is on the left and normal on the right. Notice the bright white spec in the cancerous image.

For the Lung Nodule data, each nodule was isolated as a 64x64 pixel 3D image consisting of 48 slices. Training data set consisted of 37 nodule samples and the testing set consisted of 18 lung nodule samples. Nodule Mask data was provided within a numpy array, consisting of nodule area per slice as well as a numpy array of malignancy indicators. Indicators were labeled as “0” or “1”, “0” indicating a benign lung nodule and “1” indicating a malignant nodule.

2D Network Structure for Breast Cancer Application

Data Exploration and Processing

Digital Imaging and Communications in Medicine (DICOM) images were loaded in by batches from a specified images folder and underwent several transformations, including normalization of pixel values from 0 to 1, inverting monochrome pixel values, resizing (256 x 256 pixels), and converting to PNG format. A subset of the original data, 2000 patients, was selected randomly by patient ID and split into training, validation, and testing sets using a 70-15-15 split respectively. Splitting by patients rather than images was crucial for maintaining the independence between the sets. For each of these sets, they were further split by their class labels (i.e., “normal” and “cancer”) which facilitated separate analysis during model training and evaluation. Lastly, data to image mapping ensured that each row in the csv (Figure S2) corresponds to the correct PNG file in the class folders.

Model Training and Testing

Images were resized for each model based on the target resolutions shown below in Figure 6. For each EfficientNet model, the script creates a CSV file for storing the model metrics, loads the split data with the specific input shape, and computes class weights for the

training data for handling class imbalance. The EfficientNet model is then initialized with the ImageNet weights, layers are added, and compiled. Data augmentation techniques and callbacks (Early Stopping and Reduce Learning Rate om Plateau) were utilized, and the model was trained using the 'model.fit' function in Jupyter Notebook.

Base model	resolution
EfficientNetB0	224
EfficientNetB1	240
EfficientNetB2	260
EfficientNetB3	300
EfficientNetB4	380
EfficientNetB5	456
EfficientNetB6	528
EfficientNetB7	600

Figure 6. The figure here shows the desired resolutions for each model complexity. These are not required, but are recommended for optimal model performance[13].

Model Evaluation

The model was evaluated on validation and test sets, which consisted of accuracy and loss metrics. Precision, recall, and F1 score of the test set were included for each class, "normal" vs "cancer", and for the overall model performance. In addition to these, confusion matrices were generated to provide a visual of the models performance in terms of true positives, false positives, true negatives, and false negatives.

2.5D Network for Lung Nodule Classification

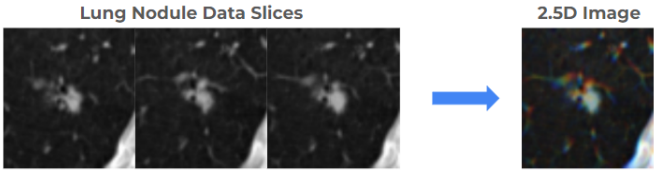

Figure 7. Shown here is the conversion of lung nodule slices to a 2.5D image.

Data Exploration and Preprocessing

HDF5 files containing the 3D images were loaded in by batches into 48 image slices per sample of 64x64 pixel size from a specified images folder and were normalized to an intensity range of 0 to 255 pixels. Images were then resized to 300x300 pixels and converted to PNG format.

2.5D Image Stacking

For the slices of each nodule data, any slice that did not contain a part of the nodule was removed. Each remaining slice was processed into grayscale format, in which there is only one channel of image information. RGB images contain three channels, therefore three consecutive image slices were placed into one of three channels within an RGB image, allowing for an image that contains more information compared to a standard grayscale image slice. This process continued until all image slices of the available grayscale nodule data were converted into the RGB images (Figure 7).

Model Training and K-fold Cross Validation

Model training generally followed the outline shown in Figure S1, however there were slight deviations. Transfer learning was utilized to create the base model to be trained. The model used for transfer learning was EfficientNetV2 B3. Due to the small quantity of data available for model training, K-fold cross validation was implemented. K-fold cross validation allows the data to be divided into smaller sets, with one of the divided sets being isolated from the others. The divided set is used as a validation set, being utilized after one fold of training to evaluate performance of the single training split. For the next split, a different isolated set is selected and the process continues as indicated. For this research, the training was conducted for five folds, or splits, with each training split operating for 10 epochs each. The model was trained upon the RGB images created in the previous step. The best validation accuracy of each split was recorded and then once the K-fold cross validation was completed, the average of the validation accuracies was calculated to display model performance.

Radiomics Analysis for Lung Nodule Classification

Data Exploration and Preprocessing

Similar preprocessing steps for the 2.5D Network approach were taken. HDF5 files were loaded in by batches and normalized to a similar intensity range of 0 to 255 pixels. Contrary to the 2.5D Network task, images were not resized or converted to PNG format. Maintaining HDF5 file format allowed for greater resolution and image quality for radiomics analysis.

Threshold Testing

Using PyRadiomics library in Jupyter notebook, image features were extracted from lung nodule scans indicating features like voxel volume, surface area of nodule, least axis length, and HU intensity. Quantitative analysis was performed to evaluate the average value of each feature for both malignant images and for benign images. Next, a threshold range was determined for statistically significant differences in averages for benign and malignant data sets. Thresholds were tested for success rate under each threshold in this range to create the ROC

curve for each feature. Areas under these curves were calculated and plotted against respective image features to determine which features were most accurate in distinction, with larger areas indicating better performance and classification.

Model Training and Improvement

Image features from threshold testing were implemented into random forest and logistic regression models based on features that had higher areas under the ROC curve. The structure of the logistic regression model involved weight estimation using coefficients for each image feature and using an intercept as a bias term, creating a linear estimation. A sigmoid function is used for prediction metrics between 0 and 1 for this model. Trial and error with various feature combinations were used to examine any improvements with accuracy for the model. Plots of benign and malignant threshold classification for each feature on testing data sets were created. This indicated any clear difference in features for malignant and benign nodules that could be potentially utilized in the model. These features were tested and implemented into the model, consequently increasing its performance.

Model Evaluation

The random forest and logistic regression models were tested using the lung nodule testing data set consisting of 18 lung nodule samples, each either benign or malignant. Success rates were determined by calculating the ratio of total correctly diagnosed lung nodules to the total number of testing image files.

Detection and Measurement Tool Performance Testing

After the development of a lymph node detection and measurement tool, the tool was to be integrated into a radiological workflow to observe performance within a workflow and to gauge user satisfaction. Accuracy and measurement values were to be compared to the radiologists assessment and a survey was to be conducted afterwards. However, this step was not implemented due to the incomplete development of the detection and measurement tool. The recommended approach can be seen in the Next Steps section of the Discussion.

End Matter

Author Contributions and Notes

The authors declare no conflict of interest.

Acknowledgments

As a team, we would like to thank our advisor Xue Feng and the team at Carina Medical for their support and guidance through the course of this project.

References

1. Common Cancer Sites - Cancer Stat Facts. SEER. Accessed May 6, 2024. https://seer.cancer.gov/statfacts/html/common.html

2. Definition of lymph node - NCI Dictionary of Cancer Terms - NCI. Published February 2, 2011. Accessed May 6, 2024. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/lymph-node

3. Storck K, Brandstetter M, Keller U, Knopf A. Clinical presentation and characteristics of lymphoma in the head and neck region. *Head Face Med*. 2019;15(1):1. doi:10.1186/s13005-018-0186-0

4. Lymph Nodes | SEER Training. Accessed May 6, 2024. https://www.training.seer.cancer.gov/anatomy/lymphatic/components/nodes.html

5. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer Oxf Engl 1990*. 2009;45(2):228-247. doi:10.1016/j.ejca.2008.10.026

6. Change in Tumor Size by RECIST Correlates Linearly With Overall Survival in Phase I Oncology Studies | Journal of Clinical Oncology. Accessed May 6, 2024. https://ascopubs.org/doi/10.1200/JCO.2011.36.4752

7. Brady AP. Error and discrepancy in radiology: inevitable or avoidable? *Insights Imaging*. 2017;8(1):171-182. doi:10.1007/s13244-016-0534-1

8. Newman-Toker DE, Pronovost PJ. Diagnostic errors--the next frontier for patient safety. *JAMA*. 2009;301(10):1060-1062. doi:10.1001/jama.2009.249

9. Berlin L, Berlin JW. Malpractice and radiologists in Cook County, IL: trends in 20 years of litigation. *AJR Am J Roentgenol*. 1995;165(4):781-788. doi:10.2214/ajr.165.4.7676967

10. Kutikova L, Bowman L, Chang S, Long SR, Arning M, Crown WH. Medical costs associated with non-Hodgkin's lymphoma in the United States during the first two years of treatment. *Leuk Lymphoma*. 2006;47(8):1535-1544. doi:10.1080/10428190600573325

11. Alexander R, Waite S, Bruno MA, et al. Mandating Limits on Workload, Duty, and Speed in Radiology. *Radiology*. 2022;304(2):274-282. doi:10.1148/radiol.212631

12. Munasinghe K, Karunanayake P. Recursive Feature Elimination for Machine Learning-based Landslide Prediction Models. In: *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. ; 2021:126-129. doi:10.1109/ICAIIC51459.2021.9415232

13. Team K. Keras documentation: EfficientNet B0 to B7. Accessed May 6, 2024. https://keras.io/api/applications/efficientnet/

Supplemental Information

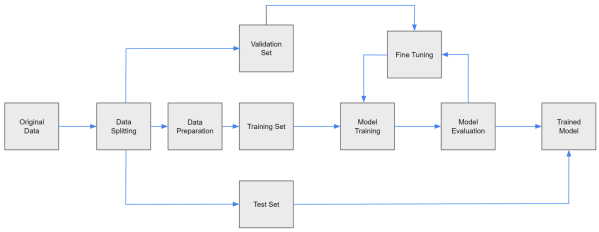

Figure S1. The above supplementary figure outlines the general approach carried out to train a deep learning model.

	site_id	patient_id	image_id	laterality	view	age	cancer	biopsy	invasive	BIRADS	implant	density	machine_id	difficult_negative_case
0	2	10006	462822612	L	CC	61.0	0	0	0	NaN	0	NaN	29	False
1	2	10006	1459541791	L	MLO	61.0	0	0	0	NaN	0	NaN	29	False
2	2	10006	1864590858	R	MLO	61.0	0	0	0	NaN	0	NaN	29	False
3	2	10006	1874946579	R	CC	61.0	0	0	0	NaN	0	NaN	29	False
4	2	10011	220375232	L	CC	55.0	0	0	0	0.0	0	NaN	21	True
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
54701	1	9973	1729524723	R	MLO	43.0	0	0	0	1.0	0	C	49	False
54702	1	9989	63473691	L	MLO	60.0	0	0	0	NaN	0	C	216	False
54703	1	9989	1078943060	L	CC	60.0	0	0	0	NaN	0	C	216	False
54704	1	9989	398038886	R	MLO	60.0	0	0	0	0.0	0	C	216	True
54705	1	9989	439796429	R	CC	60.0	0	0	0	0.0	0	C	216	True

54706 rows × 14 columns

Figure S2. Shown here is the csv file containing the breast nodule training data that was mapped to the images.