Prototypical Network for Lung Cancer Type Detection

Nimesha Hansani Amarasinghe
Department of Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
nimesha.21@cse.mrt.ac.lk

Thanuja D. Ambegoda
Department of Computer Science and Engineering
University of Moratuwa
Moratuwa,Sri Lanka
thanujaa@uom.lk

***Abstract*—Lung cancer is a devastating global health issue that can be fatal if not detected early. To increase the chances of successful treatment and prevent the loss of lives, doctors need to identify the which type lung cancer belongs to. Currently, CT scans are commonly used in medical practice to detect and diagnose lung tumors. However, implementing deep learning models to identify the lung cancer types poses a significant challenge. Because acquiring many medical images for each type of lung cancer can be difficult.The challenge of requiring a large amount of data samples for each category in traditional deep learning models was addressed in this research by implementing a prototypical network, which is a few shot learning technique. This method requires only a few samples per category, and it was used in conjunction with a pre-trained model to extract features from lung CT scans. The accuracy of the model was analyzed based on the number of samples per category. Overall, the results of the study demonstrate that implementing a prototypical network for lung cancer type detection is feasible**

***Index Terms*—Prototypical Network ,Few-shot learning ,VGG16 ,Lung Cancer,CNN**

I. Introduction

Deplorably, lung cancer ranks among the deadliest and most prevalent types of cancer, principally originating within the lung tissues. The American Cancer Society has identified two main classifications of lung cancer: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). Within the domain of NSCLC, there are three significant subgroups: Adenocarcinoma, Squamous cell carcinoma, and enormous cell carcinoma. [1].

On the other hand, SCLC represents a distinct classification. Understanding the differences among these types is urgent for tailoring effective treatment strategies and improving patient endurance rates.

Computer-aided design frameworks are increasingly becoming a preferred tool for healthcare professionals , especially radiologists, in the field of diagnostic radiology and medical imaging. CT imaging stands out as the preferred methodology for lung cancer diagnosis because of its wide availability, cost-effectiveness, and fast image acquisition capacities. It gives exceptionally point by point cross-sectional images of the chest, allowing for the visualization of even little nodules and abnormalities within the lung tissue.

Many of these frameworks are fundamentally trained on datasets that predominantly feature nodal types of lung cancer. While this approach might be effective for identifying and classifying nodules, it doesn't completely align with the clinical and radiological variety of lung cancer types.

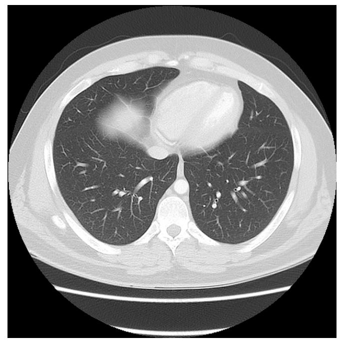


Fig. 1: CT scan of a healthy person

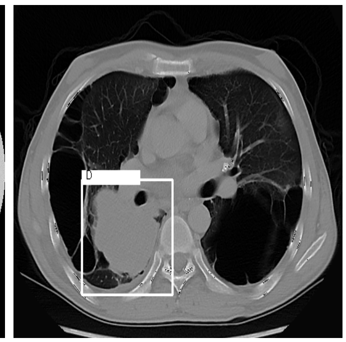


Fig. 2: CT scan of a patient with Large Cell Carcinoma

In the field of image processing and deep learning, there have been many impressive breakthroughs in medical image analysis to diagnose diseases using medical images. Deep learning strategies for lung cancer identification utilizing CT images have been proposed by inspired researchers with a lot of marked data [2], [3], [4], [5], [6].

In any case, detecting the lung cancer type is a truly challenging task, since collecting medical images for each kind is infeasible. In certain situations, researchers will most likely be unable to obtain the necessary extensive data that they expect for their examinations ,because of confidentiality and security concerns. To conquer this challenge, there has been increasing sub-region in machine learning called few-shot learning. The point of this approach is to accomplish great learning results despite having a restricted amount of marked data in the training dataset, which contains instances of inputs matched with their respective results.

Prototypical Networks, introduced by Snell et al [7] play had a significant impact in advancing few-shot learning. These networks enable models to learn representations of classes and make predictions considering the similarity of new instances to class prototypes. This approach has found applications in different domains, including image recognition, natural language processing, and medical image analysis [8]

II. Background

Few-shot learning, a subset of machine learning, has arisen as a compelling way to deal with addressing the challenges presented by restricted data availability and improving the generalization capacities of models in different domains. Few-shot learning models can be broadly categorized into two groups known as non-meta learning and meta-learning. The Prototypical Network is a famous meta-learning approach for few-shot learning that utilizes a nearest neighbor strategy, eliminating the need for hyper-parameters in the meta-test stage and resulting in practically insignificant inference time.

Prototypical Networks, introduced by Snell et al [7] play had a significant impact in advancing few-shot learning.These networks enable models to learn representations of classes and make predictions considering the similarity of new instances to class prototypes. This approach has found applications in different domains, including image recognition, natural language processing, and medical image analysis [8], [9], [10], [11], [12].

Few-shot learning has shown guarantee in the field of medical image analysis, as evidenced in the concentrate by Keshani et al. [12].Their work centers around glaucoma diagnosis using a small measured dataset of high-resolution fundus images, emphasizing the potential of few-shot learning techniques in addressing challenges related with restricted medical image data.

Yifan Jian and their research group have introduced an innovative way to deal with Coronavirus CT diagnostics, leveraging regulated domain adaptation techniques [9].This technique offers significant advantages when only a predetermined number of named CT images are accessible, a common challenge in medical imaging applications.

A research team in India [8] proposed an implementation of P-shot n–ways Siamese network, based on deep learning principles combined with prototypical nearest neighbor classifiers. Their approach was to accurately classify COVID-19 infection in lung CT scan slices and the technique investigated that different pre-trained network CNN models affect the performance of multi-class classification of Siamese based networks.

Nicholas and their research group, as documented in their work [13],have introduced an innovative dynamic few-shot learning framework custom fitted for lung cancer lesion segmentation. This framework consistently integrates the force of few-shot learning with the U-Net architecture, presenting a novel way to deal with lung cancer detection and classification. What sets their procedure separated from existing methodologies is the worldwide neighborhood PET/CT methodology fusion, a technique that veers off from conventional strategies, which regularly exploit separate PET and CT features or perform restricted fusion using CNN structures.

A research group in Petersburg, Russia [5],has implemented an aggressive methodology to enhance the differential diagnosis of various lung conditions and create a comprehensive classification system for a broad spectrum of lung cancers.

Domain	Dataset	Performance
Siamese network based model [9]	Covid-19 ct segmentation from url-https://medicalsegmentation.com/covid19	Accuracy 0.8040±0.0356 F1-score 0.7998±0.0384
Prototypical closest neighbors' classifiers combined with a P- shot N-ways Siamese network [8]	Chest CT scans from 1110 patients in medical hospitals from Moscow ,Russia	Accuracy 98.07% F1-Score 95.10%
A model based on a few-shot U-Net architecture [13]	Lung-PET-CT-DX dataset in TCIA database. PET/CT scans from 87 patients	Accuracy 99% Precision 70.62%
Comparison using both Zero-shot learning and Few Shot Learning [14]	LC25000 dataset.25,000 color images in 5 classes	99.87% of accuracy from few-shot setting

TABLE I. Previous Work on Few-Shot Learning in Medical Imaging

Their approach involves the development of a Computer-Aided Diagnosis framework, utilizing a Siamese neural network. This network undergoes training on a specific dataset that includes meticulously segmented and labeled abnormal lung objects, with a focus on tumors. The dataset is carefully categorized into distinct groups such as "typical" peripheral lung cancer (LC), "abnormal" LC, and "not cancer," based on discernible CT image patterns. Notably, the dataset comprises only confirmed tissues validated through precise examinations, ensuring a high level of data integrity.

Fu-Ming Guo and Ying fang Fan [14],employed a pre-trained Vision Transformer (ViT) model for the classification of lung cancer on histologic slices with multiple labels. They conducted evaluations in both Zero-Shot and Few-Shot settings, comparing the performance of ViT in terms of accuracy, precision, recall, sensitivity, and specificity. Their study revealed that the pre-trained ViT model exhibited excellent performance in the Zero-Shot setting. In the Few-Shot setting with just one epoch, it demonstrated competitive accuracy at 99.87%, and with five epochs, it achieved optimal results with 100.00% accuracy on both the validation and test sets.

III. Methodology

A. Dataset

The implementation of this study used CT scan images obtained from the Lung-PET CT-Dx dataset in The Cancer Imaging File (TCIA), an invaluable asset laid out by the National Cancer Institute [15].These images were complemented by XML Annotation documents that gave vital information regarding the localization of growths through bounding boxes.

The dataset was curated retrospectively, focusing on individuals under suspicion of lung cancer who had undergone both a standard-of-care lung biopsy and PET/CT imaging. It encompasses CT scans of patients diagnosed with different types of lung cancer, including adenocarcinoma, little cell carcinoma, huge cell carcinoma, and squamous cell carcinoma.

B. Image Processing

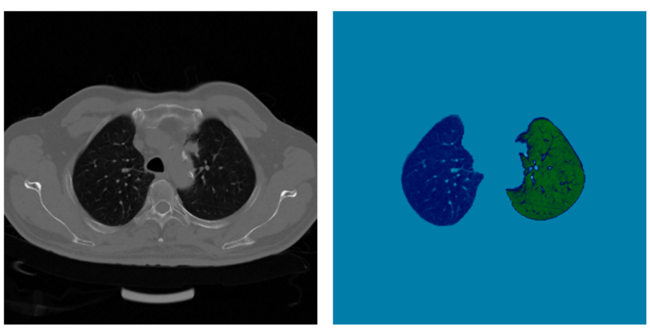
As the first step pixel information from the DICOM images were extracted. This pixel cluster data is then converted into a NumPy exhibit, which works with subsequent preprocessing steps. Conversion of pixel representations into Hounsfield units is the next step. Hounsfield Units represent a standardized scale within the domain of medical imaging, serving as a universal reference for characterizing the radiodensity of explicit tissues or materials. This conversion enhances the likeness and analyzability of images, accordingly, enabling more precise and dependable diagnosis and treatment planning.

In the context of CT scan images, the presence of inherent noise can be especially dangerous, potentially leading to erroneous interpretations of the cancer's stage. Consequently, the evacuation of noise becomes basic to ensure the accuracy and unwavering quality of diagnostic results. In the implementation median filtering was applied to the images to reduce the noise.

C. Image Segmentation

Image segmentation is fundamentally concerned with partitioning an image into discrete regions or classes, with every subset exhibiting internal homogeneity regarding explicit characteristics.

U-net is a modern architecture which has outstanding performance in segmenting returns for money invested [13].Johannes and his group [16] introduced a changed version of the U-net architecture, which they trained using the R-231 Dataset. In their research paper they have featured that at the hour of submission, the U-net(R-231) model attained the second-most elevated score among all participants in the LOLA11 challenge. Furthermore, it covered more cancer volume.



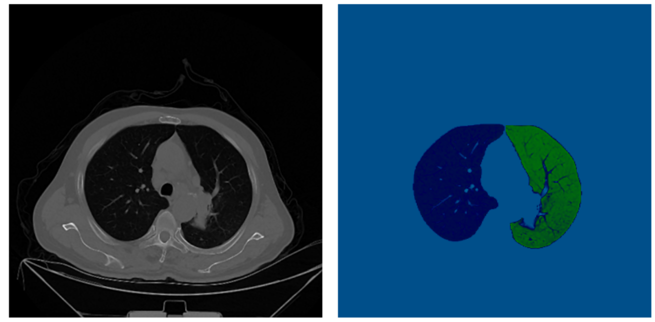
((a)) Before Segmentation ((b)) After Segmentation

Fig. 3: U-Net (R321) applied for CT scan image with Squamous Cell Carcinoma.

Because of the remarkable performance of the U-net(R-231) model ,it was applied as the segmentation method for preprocessed CT images.

D. Feature Extraction

Pre-trained CNNs in the feature extraction period of medical image analysis stands as a significant advancement, enabling the programmed acquisition of intricate image features and leveraging the progressive representations learned during training.



((a)) Before Segmentation ((b)) After Segmentation

Fig. 4: U-Net (R321) applied for CT scan image with Adenocarcinoma.

Utilization of pre-trained CNNs conveys the benefit of these models having been trained on different sets of images, promoting generalization, and mitigating the risks of overfitting, particularly when confronted with restricted medical data.

Four distinct pre-trained CNNs, namely VGG16 [17], ResNet50 [18], DenseNet [19] and CNN, were harnessed for the feature extraction process in the analysis of lung CT images in this methodology.

E. Prototypical Network

The D-dimensional feature vector of each example will be output by a pre-trained feature extraction model. To create a prototype for each lung cancer type ,the average of all feature vectors in a class label will be calculated. This prototype serves as a representative for the label and is used to compare with the features of new examples during inference.

Next key component in prototypical network is calculation of distance matrix. The distance matrix is used to measure the similarity between the support set and the query set. It is shown that Euclidean distance performs better than cosine similarity when calculating distance [7]

F. Episodic Learning

Episodic learning has been applied to improve the performance of the model. In this approach, a few shot classification tasks are directly tuned for the model. One Iterative cycle ,here considered as "Epoch".

During each epoch, the whole training set is feeding forward and backward and it will be partitioned into several parts which are called "Batch". Consequently, the number of iterations per epoch are all batches that feed into the model. The support set is in the form of N-way k-shot random samples, and the query set consists of q random samples for each of the N support set classes and these two data sets would be created during each episode. The prediction error over episodes is used to update the meta-learner. The meta-learner learns to learn from the limited dataset throughout a series of episodes. This stage is known as meta-learning.

For this development, the CrossEntropyLoss() function is used to define the loss criterion for the classification task. Adam() function is used to define the optimizer for updating the model's parameters during training. The learning rate was set to 0.001,which determines the step size taken by the optimizer in the parameter space.

IV. RESULTS & DISCUSSION

The experiment involved the utilization of four pre-trained CNN models as the feature extractors. The performance evaluation of each pre-trained model is conducted by assessing metrics such as Precision, Recall, F-Score, and Accuracy score. This evaluation is conducted under varying conditions where the number of shots per class in the support set is altered. The purpose is to gauge how well the models perform when faced with different scenarios of data availability for each class. Table II,III,IV and V present the Precision ,Recall and F1-Score for four different pre-trained model ,when applied to the prototypical network.

Shots per class	Precision	Recall	F1-Score
1-Shot	0.74	0.68	0.68
2-Shot	0.82	0.81	0.81
3-Shot	0.86	0.81	0.82
4-Shot	0.77	0.75	0.74
5-Shot	0.69	0.66	0.65

TABLE II. Evaluation metrics of VGG16 pre-trained model

Shots per class	Precision	Recall	F1-Score
1-Shot	0.6	0.56	0.54
2-Shot	0.88	0.87	0.87
3-Shot	0.68	0.68	0.68
4-Shot	0.95	0.93	0.93
5-Shot	0.87	0.83	0.82

TABLE III. Evaluation metrics of CNN model

Shots per class	Precision	Recall	F1-Score
1-Shot	0.79	0.68	0.66
2-Shot	0.95	0.93	0.93
3-Shot	0.88	0.87	0.87
4-Shot	0.87	0.83	0.82
5-Shot	0.9	0.9	0.9

TABLE IV. Evaluation metrics of pre-trained DenseNet model

Shots per class	Precision	Recall	F1-Score
1-Shot	0.33	0.43	0.33
2-Shot	0.22	0.31	0.25
3-Shot	0.76	0.68	0.69
4-Shot	0.71	0.62	0.6
5-Shot	0.98	0.98	0.98

TABLE V. Evaluation metrics of pre-trained ResNet50 model

The results demonstrate how the model's performance changes with the number of labeled example per class. Figure 4 displays the accuracy for each pre-trained model with the number of lung CT images per class. The highest accuracy scores are attained for all models with more than 3-shot settings.

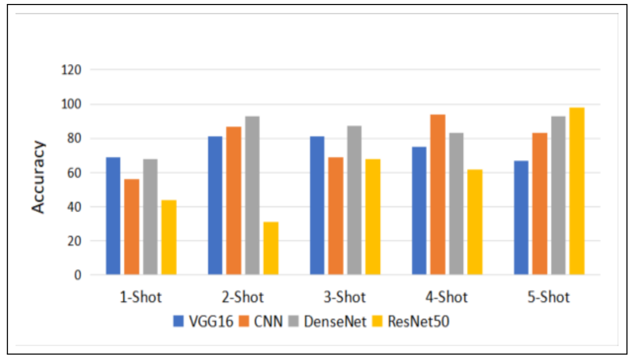


Fig. 5: Accuracy comparison for each pre-trained model.

Notably, the highest accuracy observed as 97% for ResNet50 model with 5-Shot setting for each class. Overall, the models achieve their best performance when there are more shots per class, especially with 4 or 5-Shot classes, indicating the importance of having sufficient training examples to achieve higher accuracy and F1-Scores. The findings underline the importance of considering both the decision of model and the availability of marked data when aiming for ideal few-shot learning results.Below graphs show the training accuracy and training loss were tracked over multiple episodes for each model. Model was set to learn for 15 episodes.

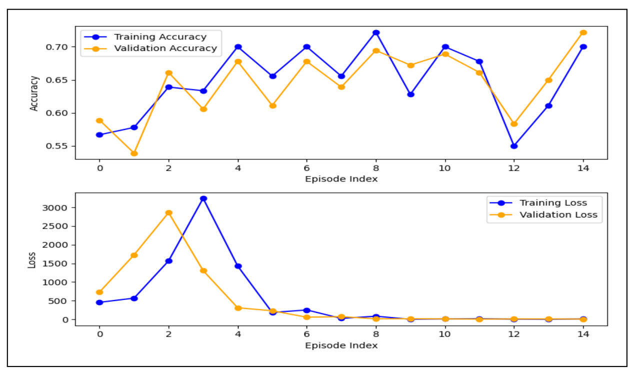


Fig. 6: Episodic Learning with Pre-trained VGG16

When examining the training progress for each pre-trained model from Figure 6 to 10, which was completed 15 episodes with 4 shots for every class, interesting patterns arise. Looking at the VGG16 model, the training accuracy fluctuates all through the episodes. This indicates that the model's learning cycle isn't totally steady, because of variations in the accessible data or the intricacy of the task.

For the ResNet50 model, the training accuracy initially wavers around 0.6 prior to showing improvement in episodes two and three. In any case, the accuracy is inconsistent across episodes, and a declining trend is seen towards the end of the training. Conversely, the training loss begins higher however diminishes with slight fluctuations. The CNN model,

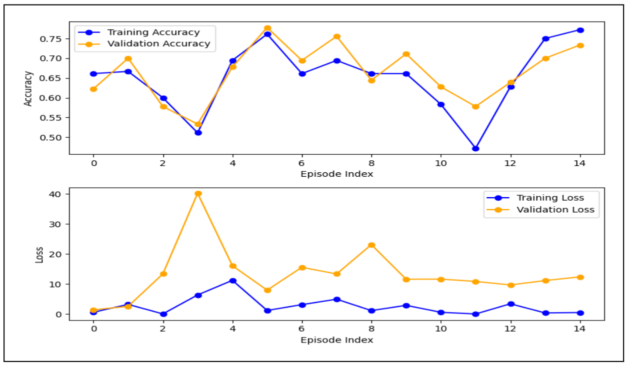


Fig. 7: Episodic Learning with Pre-trained ResNet50

in contrast, displays consistent advancement regarding training accuracy, with consistent improvement throughout the span of the episodes.

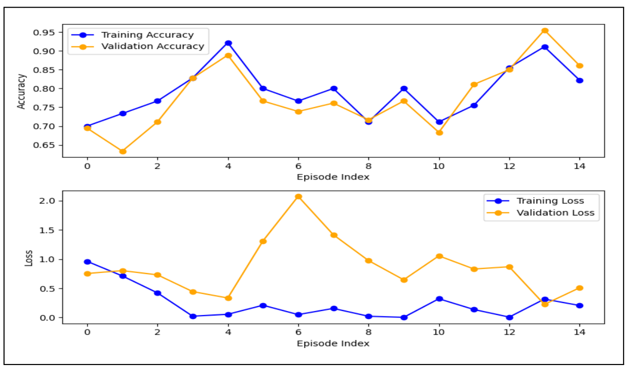


Fig. 8: Episodic Learning with DenseNet

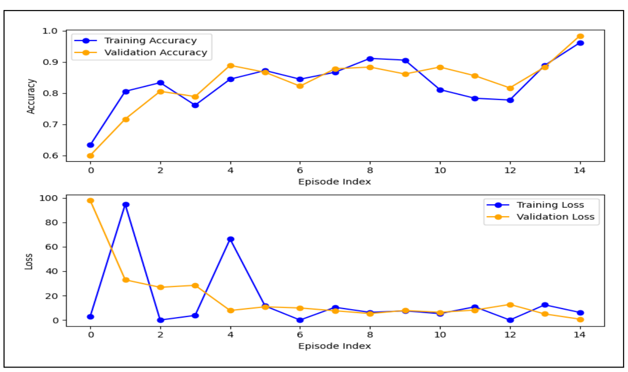


Fig. 9: Episodic Learning with CNN

This indicates a steady learning process, potentially owing to the model's architecture and data distribution. Likewise, the training loss shows a consistent decline, indicative of effective learning.Finally, the DenseNet model shows an initially fluctuating training accuracy, however it balances out as training advances. The training loss diminishes, with minor fluctuations, suggesting that the model is adapting to the data effectively. These findings recommend that different pre-trained models display assorted training ways of behaving. While certain models demonstrate sporadic patterns, others show more steady and consistent learning trends.

These observations feature the mind-boggling nature of training neural networks, influenced by factors like model architecture, data distribution, and convergence dynamics. In the conducted study, the most noteworthy accuracy was accomplished by a model employing Convolutional Neural Networks (CNNs) as the feature extractor. The experiment encompassed 15 training episodes, where every episode consisted of 4 support images from each class and 3 query images from each class. The model attained an impressive accuracy of 98

V. CONCLUSION & FUTURE WORK

The entire postulation project spun around identifying four main types of lung cancer using CT scan images, with an essential spotlight on determining the most fitting image processing philosophy for medical images when it is restricted to train data. The project basically used the prototypical network, a notable few-shot learning model, to accomplish this objective. Different pre-trained models and a CNN model were utilized to extract a superior feature space from lung CT images.

The insights generated by this research give a foundation to additional exploration and refinement in the field of lung cancer identification using CT scans and few-shot learning techniques. An intriguing avenue for future review is the amalgamation of Siamese Neural Networks and Prototypical Networks. This fusion can possibly yield an additional strong and precise model for classifying types of lung cancer. Siamese Networks [20] succeed in discerning differences, while Prototypical Networks are proficient at categorization using models. Combining their strengths could address challenges that each network faces in isolation. Mastering the specialty of consistently integrating these networks and addressing any challenges that might arise represents a captivating issue to be settled.

REFERENCES

[1] "American cancer society," https://www.cancer.org/cancer/lung-cancer.

[2] P. Chaturvedi, A. Jhamb, M. Vanani, and V. Nemade, "Prediction and classification of lung cancer using machine learning techniques," *IOP Conference Series: Materials Science and Engineering*, vol. 1099, p. 012059, 03 2021.

[3] S. Makaju, P. P.W.C, A. Alsadoon, A. Singh, and A. Elchouemi, "Lung cancer detection using ct scan images," *Procedia Computer Science*, vol. 125, pp. 107–114, 01 2018.

[4] L. Wang, "Deep learning techniques to diagnose lung cancer," *Cancers*, vol. 14, no. 22, 2022. [Online]. Available: https://www.mdpi.com/2072-6694/14/22/5569

[5] A. A. Meldo and L. V. Utkin, "A new approach to differential lung diagnosis with ct scans based on the siamese neural network," *Journal of Physics: Conference Series*, vol. 1236, no. 1, p. 012058, jun 2019. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/1236/1/012058

[6] G. Xiuhua, S. Tao, W. huan, and L. Zhigang, "Prediction models for malignant pulmonary nodules based-on texture features of ct image," in *Theory and Applications of CT Imaging and Analysis*, N. Homma, Ed. Rijeka: IntechOpen, 2011, ch. 4. [Online]. Available: https://doi.org/10.5772/14766

[7] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," 03 2017.

[8] S. Ahuja, B. K. Panigrahi, N. Dey, A. Taneja, and T. K. Gandhi, "Mcs-net: Multi-class siamese network for severity of covid-19 infection classification from lung ct scan slices," *Applied Soft Computing*, vol. 131, p. 109683, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494622007323

[9] Y. Jiang, H. Chen, H. Ko, and D. Han, "Few-shot learning for ct scan based covid-19 diagnosis," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2021-June, pp. 1045–1049, 2021.

[10] R. Agier, S. Valette, R. Kéchichian, L. Fanton, and R. Prost, "Hubless keypoint-based 3d deformable groupwise registration," 09 2018.

[11] D. Liu, X. Gao, and Q. Shen, "Prototypical network for radar image recognition with few samples," *Journal of Physics: Conference Series*, vol. 1634, p. 012116, 09 2020.

[12] J. Kotia, A. Kotwal, R. Bharti, and R. Mangrulkar, *Few Shot Learning for Medical Imaging*, 01 2021, pp. 107–132.

[13] N. Protonotarios, I. Katsamenis, S. Sykiotis, N. Dikaios, G. Kastis, S. Chatziioannou, M. Metaxas, N. Doulamis, and A. Doulamis, "A few-shot u-net deep learning model for lung cancer lesion segmentation via pet/ct imaging," *Biomedical Physics Engineering Express*, vol. 8, 02 2022.

[14] F. Guo and Y. Fan, "Zero-shot and few-shot learning for lung cancer multi-label classification using vision transformer," *ArXiv*, vol. abs/2205.15290, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:249192263

[15] "The cancer imaging archive," Year. [Online]. Available: https://www.cancerimagingarchive.net

[16] J. Hofmanninger, F. Prayer, J. Pan, S. Röhrich, H. Prosch, and G. Langs, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *European Radiology Experimental*, vol. 4, p. 50, 08 2020.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition." Computational and Biological Learning Society, 2015, pp. 1–14.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[20] Y. Li, C. Chen, and T. Zhang, "A survey on siamese network: Methodologies, applications and opportunities," *IEEE Transactions on Artificial Intelligence*, vol. PP, pp. 1–21, 12 2022.