Annotated Source List

Sai Siddhish Chandra Sekaran
Due: 20 December 2024

A. R., B., & Kumar R. S., V. (2022). Deep learning-based lung cancer classification of CT images using augmented convolutional neural networks. *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, 21(1). doi.org/10.5565/rev/elcvia.1490

Summary of Article:

This **article** discusses the classification of Lung cancer based on CT images with Deep Learning Convolutional Neural Networks. There are two main forms of lung cancer: Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC). Additionally, there are Adenocarcinoma (ADC) and Squamous Cell Carcinoma (SqCC), which are subtypes of Non-Small Cell Lung Cancer. The image data was from the Lung Imaging Database Consortium-Image Database Resource Initiative. The researchers augmented each image in the data by rescaling, shearing, zooming, and/or a horizontal flip. The model had convolutional and pooling layers. Each node was activated by a non-linear activation function. The model's overall accuracy of 95% is amazing considering the task at hand.

Application to Research:

This article applies to my research as it contains a lot of very useful information. Some of this useful information includes the model used, the layers in the model, the types of lung cancer, the dataset the CT images were pulled from, and the data pre-processing techniques. Based on the authors and references, it is easy to see that this article is a scholarly and credible source. I will request an interview with one or more of the authors.

Ashwin Shanbhag, G., Anurag Prabhu, K., Subba Reddy, N. V., & Ashwath Rao, B. (2022). Prediction of lung cancer using ensemble classifiers. *Journal of Physics: Conference Series*, *2161*(1). doi.org/10.1088/1742-6596/2161/1/012007

Summary of Article:

This **article** talks about detecting Lung Cancer (LC) in its earlier stages with Machine Learning Ensemble techniques and CT scans. Detecting LC is difficult because its symptoms only show up in later stages, but the article uses AI classification to detect LC better. In the article, the final model was an ensemble of a decision tree, multi-layer perceptron (MLP), K-nearest neighbors algorithm (K-NN), logistic regression (LR), and support vector machine (SVM). For pre-processing, OpenCV and Gaussian Blur noise filtration were applied to reduce the noise in each image. Then, the image was divided into segments using the Otsu thresholding technique. Lastly, properties of the labeled regions of the image were extracted (area, perimeter, etc.).The final ensemble classifier had an accuracy of 85%.

Application to Research:

This article was very useful in developing my knowledge of the different AI model statistics (accuracy, precision, recall, and F1-score), which are used widely in AI and machine

learning. Additionally, many aspects of this article may carry into my research as I also plan to detect lung cancer via CT scan images.

Elnakib, A., M. Amer, H., & E.Z. Abou-Chadi, F. (2020). Early Lung Cancer Detection using Deep Learning Optimization. *International Journal of Online and Biomedical Engineering (iJOE)*, 16(06), pp. 82–94. doi.org/10.3991/ijoe.v16i06.13657

Summary of Article:

This **article** presents a new deep learning, computer-aided detection (CAD) system for the earlier detection and diagnosis of lung cancer from low-dose computed tomography (LDCT) images (Elnakib, 2020). The article discusses the cons of CT scans with details about how manual analysis of CT scans is time-consuming and delivers a larger amount of X-ray radiation when compared to LDCT scans that provide higher quality at a lower X-ray radiation exposure. The article details the image pre-processing method, which included improving the contrast of the LDCT scans, extracting deep learning features, optimizing the previously extracted features, and then classifying the screening as normal or cancerous. The article then goes on to describe the models and model architectures used and how they were used. The final architecture settled upon was the VGG19 architecture and an SVM classifier with an accuracy of 96.25%.

Application to Research:

Though this article uses LDCT scans instead of my planned medical imaging (CT scans), the disadvantages discussed in this article are considerably swaying my opinion to switch to researching and using LDCT screening due to its superiority over CT scans. Additionally, the paper covered AI models (such as VGG16) and frameworks I am unfamiliar with and has helped me gain new understanding, insight, and research direction.

Flyckt, R. N. H., Sjodsholm, L., Henriksen, M. H. B., Brasen, C. L., Ebrahimi, A., Hilberg, O., Hansen, T. F., Wiil, U. K., Jensen, L. H., & Peimankar, A. (2024, February 14). Pulmonologists-Level lung cancer detection based on standard blood test results and smoking status using an explainable machine learning approach. arXiv. Retrieved January 1, 2025, from https://arxiv.org/pdf/2402.09596

Summary of Article:

This article uses machine learning to detect lung cancer with standard blood tests and smoking status. A dynamic ensemble selection model achieved high sensitivity and specificity, outperforming pulmonologists. Additionally, smoking, age, and blood markers were key predictors of lung cancer. The article states the model built (DES) had a 9% higher accuracy in detecting lung cancer than five pulmonologists. However, it is worth noting the main model (DES) only had a 73% accuracy.

Annotated Source List

Application to Research:

This article introduces non-invasive diagnostic techniques using ML, aligning with my research focus on early lung cancer detection. Its approach offers insights into integrating clinical data with ML models.In addition, it also provides useful pieces of information, such as some of the main factors causing lung cancer.

Kriegsmann, M., Haag, C., Weis, C.-A., Steinbuss, G., Warth, A., Zgorzelski, C., Muley, T., Winter, H., Eichhorn, M. E., Eichhorn, F., Kriegsmann, J., Christopoulos, P., Thomas, M., Witzens-Harig, M., Sinn, P., Stenzinger, A., & Kriegsmann, K. (2020). Deep Learning for the Classification of Small-Cell and Non-Small-Cell Lung Cancer. *Cancers, 12*(6). htttps://doi.org/10.3390/cancers12061604

Summary of Article:

This **article** focuses on the classification and distinction of lung cancer subtypes with the use of Deep Learning Convolutional Neural Networks. The three most common lung cancer subtypes are Small Cell Lung Cancer (SCLC), Adenocarcinoma (ADC), and Squamous Cell Carcinoma (SqCC). To program the AI models, the researchers used the Keras, Tensorflow, and Tidyverse Python packages along with the ImageNet dataset. To optimize these AI models, the RMSProp optimizer was used. The researchers experimented with many different factors, such as different image input sizes for the models, different training batch sizes, and different dropout rates). The research highlights images at 256 x 256 pixels, a batch size of 16, and a dropout rate of 0.5 as being ideal with each having the highest accuracies in their respective part of the experiment.

Application to Research:

This article is relevant and useful to my research as it has a lot of very detailed and precise information I can build on. The mentions of the subtypes of lung cancer, the Python packages used, the exact layers and activation functions used, and the optimizer used are all also good information to have and be able to use.

Li, Y., Wu, X., Yang, P., Jiang, G., & Luo, Y. (2022). Machine learning for lung cancer diagnosis, treatment, and prognosis. *Genomics, Proteomics & Bioinformatics, 20*(5), 850-866. doi.org/10.1016/j.gpb.2022.11.003

Summary of Article:

This article provides an overview of machine learning-based approaches that enhance various aspects of lung cancer diagnosis and therapy, including early detection, auxiliary diagnosis, prognosis prediction, and immunotherapy practice. The study highlights the challenges and opportunities for future applications of machine learning in lung cancer.

Application to Research:

The in-depth review of machine learning applications in lung cancer care offers a broad perspective on current methodologies and future directions, informing the

Annotated Source List

development of my research. It also has several useful references, links/tools, etc. that can aid me in the future.

Makaju, S., Prasad, P., Alsadoon, A., Singh, A.K., & Elchouemi, A. (2018). Lung Cancer Detection using CT Scan Images. *Procedia Computer Science*, *125*. doi.org/10.1016/j.procs.2017.12.016

Summary of Article:

This **article** describes existing lung cancer detection techniques, their strengths, and their weaknesses. The article discusses several stages of their proposed AI model. The researchers grayscaled the images and then used the Median and Gaussian filters to remove noise. Then they dissect and segment the image with watershed segmentation. In the last stage of computer-aided diagnosis, a Support Vector Machine (SVM) model is brought in to predict if the identified cancer nodule is malignant or benign. The final model had an accuracy of 90.1%. One major flaw in the model is the inability to classify the stage of Lung Cancer (e.g. Stage 1, 2, 3, or 4).

Application to Research:

This article was very useful to my research as it closely aligns with what I plan to do. I could build upon the identified strengths and weaknesses of the ML model built and analyzed here. Additionally, the article introduced me to several new techniques, such as Median filtering and Watershed Segmentation. Along with that, I learned about different features of lung cancer nodules, such as eccentricity.

Mamun, M., Farjana, A., Al Mamun, M., & Ahammed, M. S. (2022). Lung cancer prediction model using ensemble learning techniques and a systematic review analysis. *IEEE Xplore*, *World AI IoT Congress (AIIoT)*. doi.org/10.1109/AIIoT 54504.2022.9817326

Summary of Article:

This **article** talks about detecting Lung Cancer with Machine Learning Ensemble techniques using tabular data. It states that lung cancer is dangerous cells that spread and grow fast and out of control. The researchers used Machine Learning Ensemble techniques to improve the accuracy of lung cancer classifiers. Additionally, their data included many parameters such as “age, smoking, yellow fingers, anxiety,” and more taken from Kaggle (Mamun et al., 2022, p. 1). For data preprocessing, they cleaned up the data, handled missing values, and balanced the dataset, among other preprocessing methods. Their results showed that XGBoost outperformed the other techniques with an accuracy of 94.42%.

Application to Research:

This article was descriptive and had a lot of useful information. The ensemble architectures of XGBoost, LightBGM, Bagging, and AdaBoost are things I will definitely be looking into for my research. Additionally, their descriptions of lung cancer and ensemble modeling also helped to improve my understanding and comprehension in this area. The one notable downside, however, is that their models used tabular data; I plan to use CT scans.

Mamun, M., Mahmud, M. I., Meherin, M., & Abdelgawad, A. (2023, April 10).

Annotated Source List

LCDctCNN: *Lung cancer diagnosis of CT scan images using CNN based model*. arXiv. Retrieved January 1, 2025, from https://arxiv.org/pdf/2304.04814

Summary of Article:

This article proposes a deep learning model-based Convolutional Neural Network (CNN) framework for the early detection of lung cancer using CT scan images. The model outperformed other architectures, achieving an accuracy of 92%, indicating its potential for clinical application.

Application to Research:

The proposed CNN framework provides a basis for developing deep learning models in lung cancer detection, an architecture and design relevant to my research. This also strengthens my plan to use CNNs in experiments in the future and allows me to cite this source as a strong recommendation.

Maurya, S. P., Sisodia, P. S., Mishra, R., & Singh, D. P. (2024). Performance of machine learning algorithms for lung cancer prediction: A comparative approach. *Scientific Reports, 14*(1). doi.org/10.1038/s41598-024-58345-8

Summary of Article:

This article compares twelve machine learning algorithms applied to clinical data containing eleven symptoms of lung cancer, along with two major habits. The goal was to evaluate the accuracy and ability of these algorithms in predicting lung cancer. The study found that certain algorithms (namely the K-Nearest Neighbor model and the Bernoulli Naive Bayes model) outperformed others in terms of accuracy and reliability, highlighting the potential of machine learning in early lung cancer detection.

Application to Research:

This article is applicable to my research as it provides a comparative analysis of various machine learning models, including non-deep learning approaches, for lung cancer detection. The insights into algorithm performance can help me with the selection of appropriate models for my end product.

Nasser, I. M., & Abu-Naser, S. S. (2019). Lung Cancer Detection Using Artificial Neural Network. *International Journal of Engineering and Information Systems (IJEAIS)*, *3*(3). papers.ssrn.com/sol3/papers.cfm?abstract_id=3369062

Summary of Article:

The **article** starts by discussing cells and cancer as a whole. Cancerous cells are defined as rapidly growing and splitting cells that don't die; they form tumors and are usually dangerous. The article also discusses the train-test split where the AI model developed was trained on 80% of the available data and tested on the other 20%. The dataset they used comes from the data world website. The dataset had information on gender, age, smoker status, and much more. Additionally, they did data preprocessing where they looked at the table and transformed the

values. In the end, the final model had an acceptable accuracy of 96.67% and showed that the age attribute had the most effect on lung cancer presence.

Application to Research:

This article was very useful to me as it told me all about Artificial Neural Networks, a model I had never even considered implementing for Lung Cancer detection. The article also used tabular data and had a good accuracy from it, which has made me consider using tabular data as well. I also got a better understanding of the cell cycle and cancerous cells from this article, which is very important to my research on lung cancer detection.

Nazir, I., Haq, I. U., AlQahtani, S. A., Jadoon, M. M., & Dahshan, M. (2023). Machine learning‐based lung cancer detection using multiview image registration and fusion. *Journal of Sensors,* (1). doi.org/10.1155/2023/6683438

Summary of Article:

This article introduces a methodology for designing an effective machine learning classification model to identify and classify lung cancer efficiently. The approach involves multiview information fusion and radiomics, integrating various data sources to enhance detection accuracy. The proposed model demonstrated improved performance in distinguishing between benign and malignant lung cancer nodules.

Application to Research:

The article's focus on multiview information fusion and radiomics offers useful insights into combining different data types for improved lung cancer detection. This approach aligns with my research objective of improving early detection accuracy using diverse data sources.

Pathan, R. K., Shorna, I. J., Hossain, M. S., Khandaker, M. U., Almohammed, H. I., & Hamd, Z. Y. (2024). The efficacy of machine learning models in lung cancer risk prediction with explainability. *PLOS ONE, 19*(6). doi.org/10.1371/journal.pone.0305035

Summary of Article:

The article addresses the relationship between lung cancer factors and early symptoms. Four machine learning models are tuned to detect low, medium, or high lung cancer risk levels. The research emphasizes the importance of early detection and the potential of machine learning models in stratifying lung cancer risk. They place importance on unveiling what's behind the typical black-box of AI. After hyperparameter tuning, all four models had nearly 100% accuracy.

Application to Research:

The article's exploration of machine learning models for risk stratification provides a foundation for developing predictive tools in lung cancer detection, which is directly applicable to my research focus.

Pour, E. S., & Esmaeili, M. (2023, December 5). *Lung cancer detection from CT*

Annotated Source List

scan images based on genetic-independent recurrent deep learning. arXiv. Retrieved January 1, 2025, from https://arxiv.org/pdf/2312.03185

Summary of Article:

This article introduces a new lung cancer detection model in CT images using machine learning methods. The model comprises three steps: noise reduction, segmentation using Independent Recurrent Neural Networks (IndRNN), and optimization with a Genetic Algorithm. The results indicate that the proposed method can accurately detect the exact area of nodules in CT images. The IndRNN architecture used here achieved a 98.97% accuracy.

Application to Research:

This new and novel approach is something for me to consider as another model to build and experiment with in the future. In addition, the technologies used here (such as segmentation) are very important to any AI-based experiment and will be very useful later on.

Saha, A., Ganie, S. M., Pramanik, P. K. D., Yadav, R. K., Mallik, S., & Zhao, Z. (2024). VER-Net: A hybrid transfer learning model for lung cancer detection using CT scan images. *BMC Medical Imaging, 24*(1). doi.org/10.1186/s12880-024-01238-z

Summary of Article:

This article discusses a new transfer learning model (VER-Net) by stacking three different transfer learning models to detect lung cancer using lung CT scan images. The hybrid model achieved improved accuracy compared to individual models, showing the effectiveness of transfer learning in medical image analysis

Application to Research:

The hybrid transfer learning approach detailed in this article offers a promising direction for enhancing lung cancer detection models. Building this type of model and running experiments on it against individual and ensemble models could provide useful data.

Shah, A. A., Malik, H. A. M., Muhammad, A., Alourani, A., & Butt, Z. A. (2023). Deep learning ensemble 2D CNN approach towards the detection of lung cancer. *Scientific Reports, 13*(1). doi.org/10.1038/s41598-023-29656-z

Summary of Article:

The article uses a Convolutional Neural Network (CNN) to detect lung nodules (which have the potential to be cancerous) from different CT scan images. The study demonstrates that the ensemble approach improves detection accuracy, suggesting its potential in clinical applications.

Annotated Source List

Application to Research:

The article's focus on deep learning ensemble methods provides insights into advanced techniques for lung cancer detection, which is the core of my research and is in my research question itself. This article is a strong source as to why to use ensemble modeling.

Shimazaki, A., Ueda, D., Choppin, A., Yamamoto, A., Honjo, T., Shimahara, Y., & Miki, Y. (2022). Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method. *Scientific Reports, 12*(1). doi.org/10.1038/s41598-021-04667-w

Summary of Article:

This article developed and validated a deep learning-based model using the segmentation method to detect lung cancer on chest radiographs. The model demonstrated significant accuracy in identifying cancerous lesions, indicating its potential utility in clinical settings.

Application to Research:

The development of a segmentation-based deep learning model for lung cancer detection aligns with my research idea, giving me a potential methodology to explore further.

Technology, I. (2021, October 6). *What are Convolutional Neural Networks (CNNs)?* [Video]. YouTube. www.youtube.com/watch?v=QzY57FaENXg

Summary of Video:

The video states that Convolutional Neural Networks (CNNs) are a type of deep learning specializing in pattern recognition. A CNN is a part of an Artificial Neural Network (ANN). The convolutional layers of the ANN are made up of filters that perform pattern recognition. A filter is applied and compared to every possible 3x3 block on an image and an output array of similarity scores is produced. One can combine the outputted arrays from multiple filters through a process called pooling. Progressing further through the layers of a Convolutional Neural Network will reveal more abstract and powerful filters.

Application to Research:

This video is useful to my research as it gives a high-level overview of CNNs in simpler terms, making it easier to understand. Additionally, the video is by IBM Technology, a reputable and credible company in Computer Science. Maybe I could contact someone there for an interview?

Tensorflow. (2019, September 11). *Introducing convolutional neural networks (ML Zero to Hero - Part 3)* [Video]. YouTube. www.youtube.com/watch?v=x_VrgWTKkiM

Summary of Video:

Annotated Source List

According to the video, the primary principle of a Convolutional Neural Network is applying filters to an image and extracting features before training a Deep Neural Network and then reconsidering the extracted features. Outputs from multiple filters can then be combined and simplified with pooling. Filters are learned by a Convolutional through a process called Feature Extraction where the model starts with randomly initialized filters and the filters that give the best outputs will be kept and learned.

Application to Research:

This video was very useful to my research as it went more in-depth than the other bolded source; however, it stayed relatively simple and easy to understand. The video was filled with example images and filters with their multipliers. The most useful part of this video was near the end where actual CNN code was shown on screen and discussed part by part.

Wang, L. (2022). Deep learning techniques to diagnose lung cancer. *Cancers*, *14*(22). doi.org/10.3390/cancers14225569

Summary of Article:

This **article** looks at applying various deep-learning techniques to medical imaging for lung cancer detection. The article then discussed medical imaging technologies such as "X-ray, positron emission tomography (PET), [and] magnetic resonance imaging (MRI)," among others (Wang, 2022). Different AI models, such as convolutional neural networks, and types are mentioned. The article then talks about pre-processing images by denoising them with the Wiener filter, median filter, and non-local means filter (Wang, 2022). The second topic is normalization with methods such as min-max normalization. Performance-gauging metrics such as accuracy, sensitivity, specificity, precision, F1 score, and Error, among many others, are also discussed. Lastly, it provides a table of potential datasets that could be used for creating AI models for earlier lung cancer detection.

Application to Research:

This article will be very useful to my research as it contains pages of in-depth AI and lung cancer information that will help with my ventures. Things like medical imaging techniques, specific details about those techniques, different types of AI models, ways to pre-process data, and ways to evaluate models, among other things will become useful later on.