HIGH SCHOOL EDITION
Journal of Student Research
Volume 10 Issue 3 (2021)

Heuristic Oncological Prognosis Evaluator (HOPE): Deep-Learning Framework to Detect Multiple Cancers

Anu Iyer[1], Lee Conrad[#], and Fred Prior[#]

[1]Little Rock Central High School, Little Rock, AR, USA
[#]Advisor

ABSTRACT

Cancer is the common name used to categorize a collection of diseases. In the United States, there were an estimated 1.8 million new cancer cases and 600,000 cancer deaths in 2020. Though it has been proven that an early diagnosis can significantly reduce cancer mortality, cancer screening is inaccessible to much of the world's population. Machine learning approaches are increasingly successful in image-based diagnosis, disease prognosis, and risk assessment. A literature search with the Google Scholar and PubMed databases from January 2020 to June 2021 determined that currently, no machine learning model (n=0/417) has an accuracy of 90% or higher in diagnosing multiple cancers. We propose our model HOPE, the Heuristic Oncological Prognosis Evaluator, a transfer learning diagnostic tool for the screening of patients with common cancers. By applying this approach to magnetic resonance (MRI) and digital whole slide pathology images, HOPE 2.0 demonstrates an overall accuracy of 95.52% in classifying brain, breast, colorectal, and lung cancer. HOPE 2.0 is a unique state-of-the-art model, as it possesses the ability to analyze multiple types of image data (radiology and pathology) and has an accuracy higher than existing models. HOPE 2.0 may ultimately aid in accelerating the diagnosis of multiple cancer types, resulting in improved clinical outcomes compared to previous research that focused on singular cancer diagnosis.

Introduction

Artificial Intelligence

In 2007, John McCarthy defined artificial intelligence (AI) as "the science and engineering of making intelligent machines"(McCarthy, 2007). Coined "a simulation of human intelligence", AI aims to mimic human decision-making skills. Looked at another way, can we design a machine that can think? (Turing, 1950). AI is composed of several major subdomains, one of which is machine learning (Figure 1).

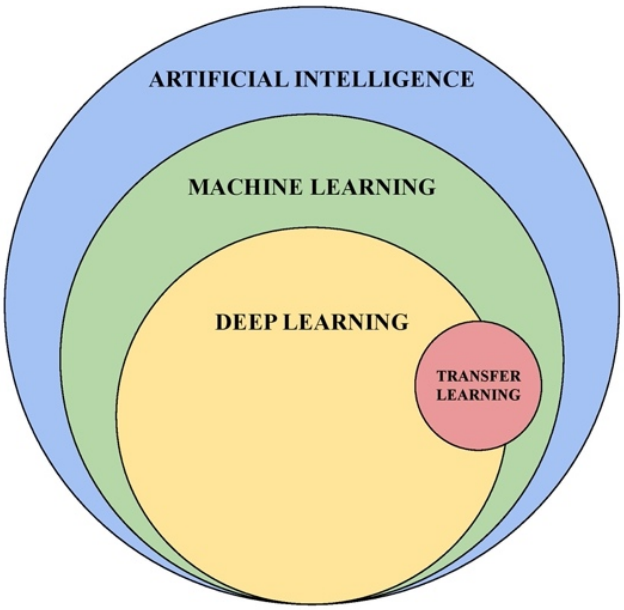

Figure 1. The relationships among AI subfields.

Machine Learning

"Any machine that could think would first have to be able to learn" (Turing, 1950). Machine learning is a very important subdomain of artificial intelligence in which algorithms are designed to learn and adapt without being explicitly programmed. The two main types of machine learning are supervised and unsupervised; a supervised learning algorithm utilizes labeled data whereas an unsupervised algorithm finds patterns in data without labeled examples (Sathya & Abraham, 2013; Shalev-Shwartz & Ben-David, 2014). There are many machine learning approaches in the literature that have been applied to the analysis of medical images, but deep learning algorithms appear to be the most promising.

Deep Learning

Deep learning is a subdomain of machine learning that uses multi-layer neural networks (which model the human brain) to hierarchically extract concepts from data. Deep learning approaches have been used to great success in fields such as radiomics (the extraction of features from a medical image; see Figure 2) and natural language processing (Goodfellow et al., 2016; LeCun et al., 2015).

Training a deep learning algorithm requires an extensive amount of examples which are frequently difficult or expensive to obtain (Prior et al., 2020). One approach that has been employed to limit the need for domain specific training examples is the use of transfer learning.

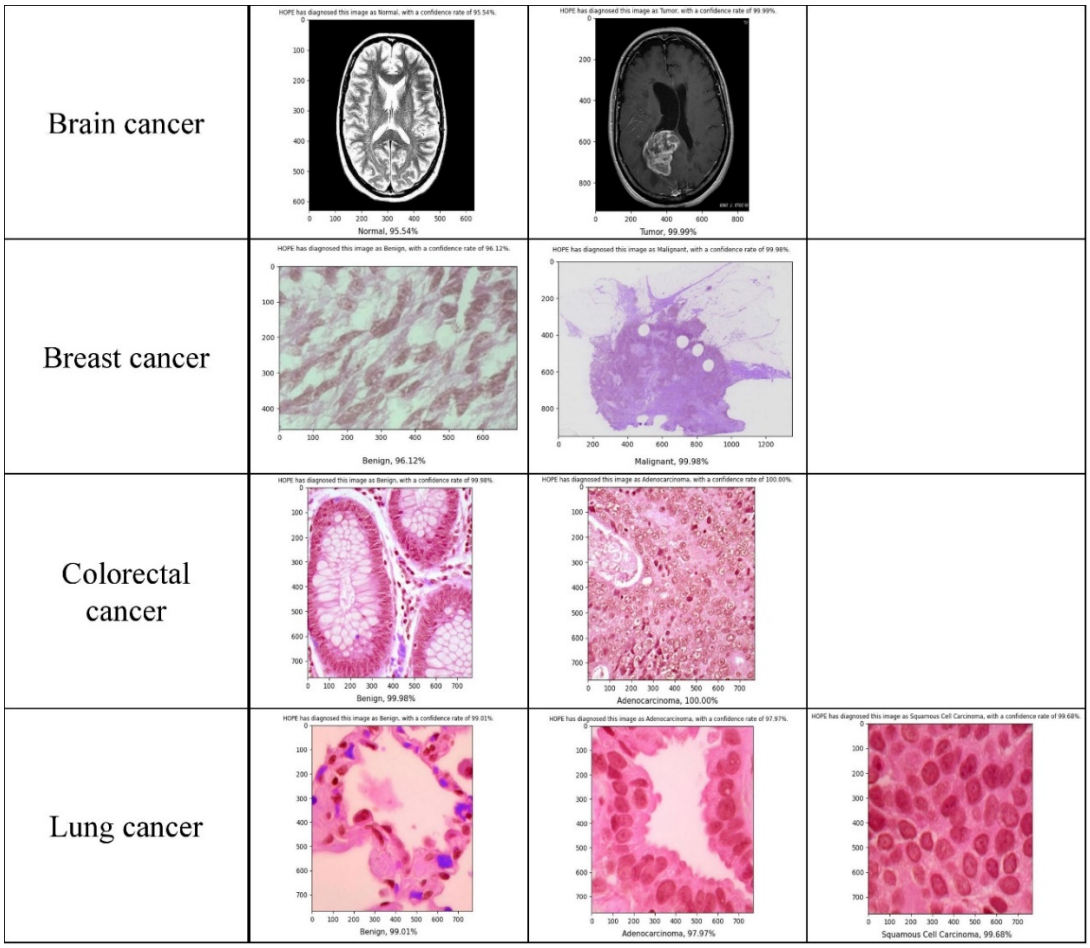

Figure 2. Machine learning enhanced quantitative analysis of radiology and pathology images can assist in disease detection.

Transfer Learning

Transfer learning applies previously learned knowledge to a new task (Raghu et al., 2019; Torrey & Shavlik, 2010). A deep learning algorithm is first trained on a large data set of the same type of information as will be used in the real application, e.g., images drawn from the internet. It is then retargeted to the problem domain, e.g., MRI images of the brain, by additional training using a much smaller labeled data set from the problem domain. Image classification correlates to supervised transfer learning, as it classifies data based on labeled training data. In this examination, a supervised transfer learning algorithm was created to diagnose multiple cancers (Figure 3).

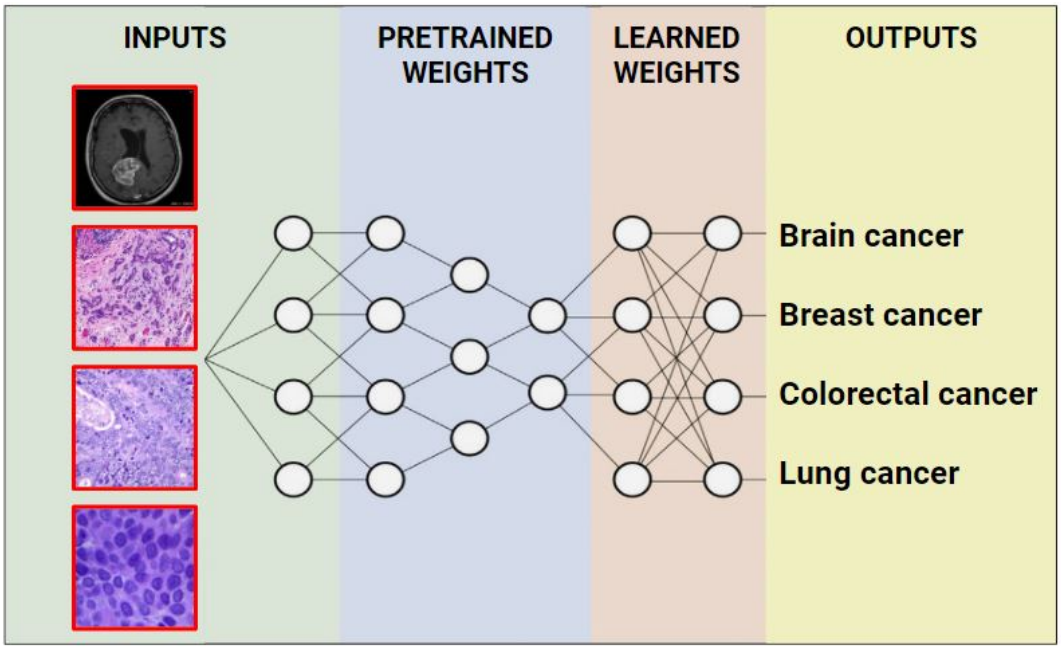

Figure 3. Transfer learning can be utilized to improve cancer detection.

Methods

PRISMA Literature Review

A PRISMA (Moher et al., 2009) literature search was conducted to determine if a machine learning model had an accuracy of 90% or higher in diagnosing multiple cancers. Google Scholar and PubMed with records from January 2020 to June 2021 were analyzed; the keywords used in Google Scholar were "transfer learning cnn brain breast colorectal lung cancer radiology pathology images accuracy", and the keywords used in PubMed were "cancer pathology radiology transfer learning cnn". A total of 459 studies were identified through the database search with Google Scholar (n=444) and PubMed (n=15). After removing duplicates with the Mendeley Reference Manager (n=42), 417 studies remained. These 417 studies were screened based on the title and abstract, and after removal (n=405), 12 studies fit the criteria of the keywords. All 12 studies were either inaccessible (n=2), did not apply transfer learning (n=2), had an accuracy less than or equal to 90% (n=6), or did not describe a model that could detect multiple cancers (n=2). Therefore, the literature search determined that currently, no machine learning model (n=0/417) has an accuracy of 90% or higher in diagnosing multiple cancers (Figure 4).

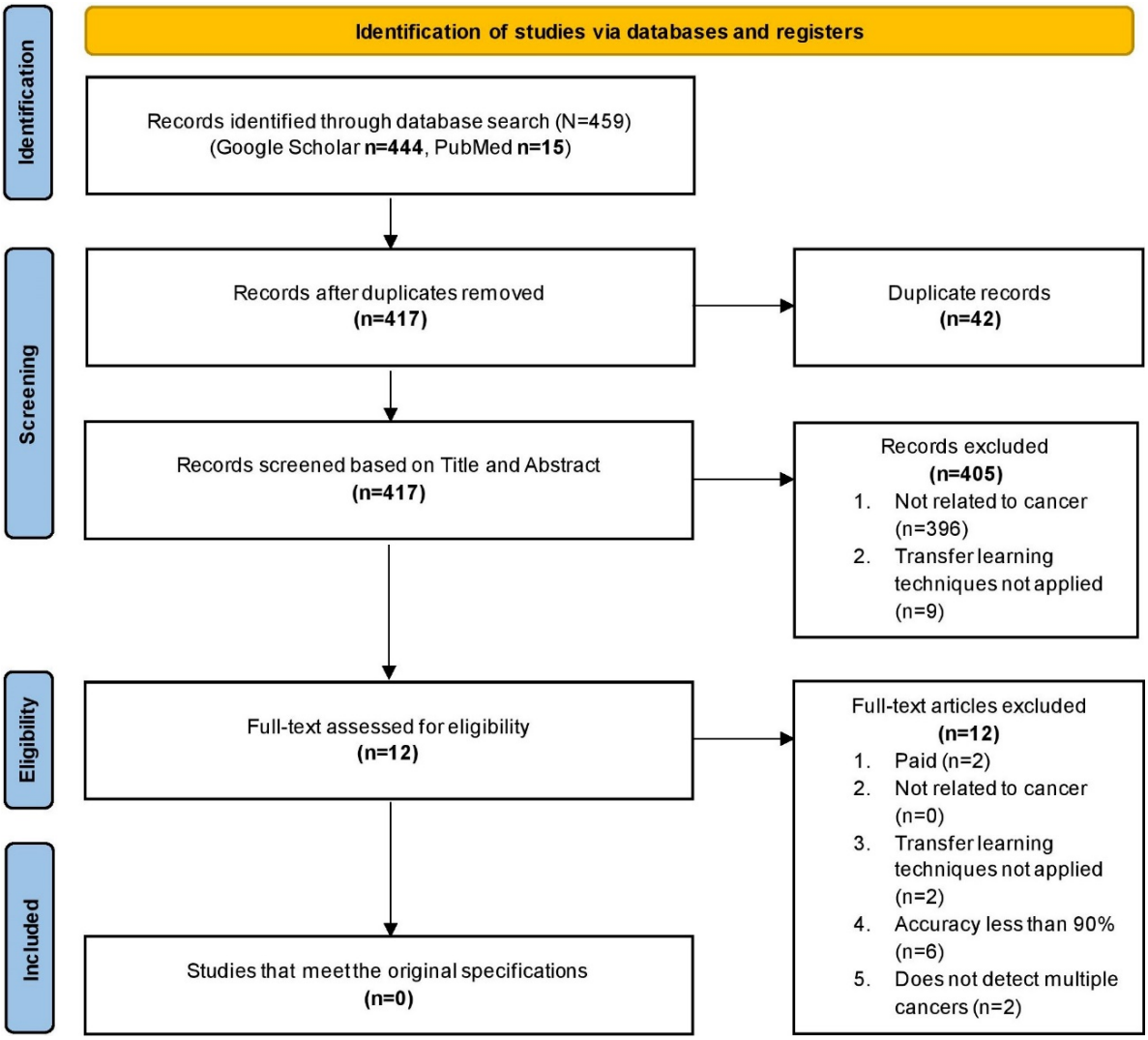

Figure 4. A literature review yielded no current model that can diagnose multiple cancers with a high accuracy.

Dataset Acquisition

After establishing that no machine learning model has met the current specifications, we began to address this issue. Kaggle is an on-line research community that conducts challenge competitions. Their website provides high-quality labeled data from these competitions and was used to locate pathology and radiology images for each of the four cancers. For brain cancer, a total of 11,663 T1 and T2 weighted MRI (radiology) images were found: 6,159 images were pre-labeled as benign, and 5,504 were pre-labeled as malignant(Kaggle, 2020a). For breast cancer, a total of 277,524 pathology images were found: 198,738 images were pre-labeled as benign, and 78,786 were pre-labeled as malignant(Kaggle, 2019). For colorectal cancer, a total of 10,000 pathology images were found: 5,000 images were pre-labeled as benign, and 5,000 were pre-labeled as malignant(Kaggle, 2020b). For lung cancer, a total of 15,000 pathology images were found: 5,000 images were pre-labeled as benign, and 10,000 were pre-labeled as malignant(Kaggle, 2020b). Originally all images were used for the model, which led to overfitting, training on lower-quality images, and an accuracy of 97.56%. To address overfitting, a structural equation modeling (SEM) analysis was

conducted to determine the minimum number of balanced images required to achieve an accuracy of 90% or higher. To address the quality of images, images were examined and chosen based on brightness, contrast, and size.

SEM Analysis

Structural equation modeling (SEM) (Hoyle, 1995) uses different types of models to show relationships between latent and observed variables (cite). The expected effect size was 0.5, the model represents 2 latent variables and 5 observed variables, the p-value was 0.05, and the statistical power was calculated as 0.8. The statistical test indicated that 2,000 images were required for optimal accuracy; so, 2,000 images were evenly distributed (by removal) across all classes of the four cancers (Figure 5).

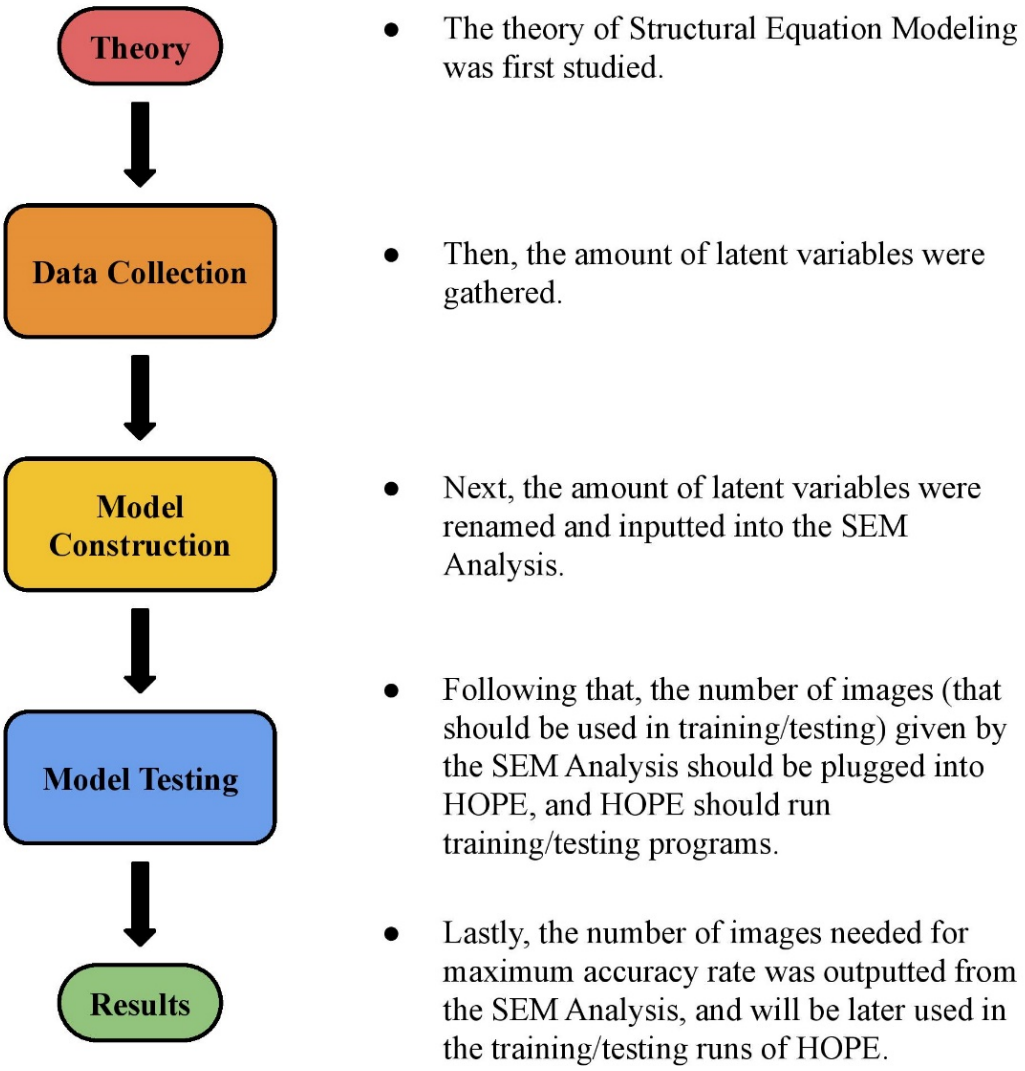

Figure 5. SEM Analysis was used to find the minimum number of images for a high accuracy model.

Experimental Procedure

Google Inception v3 Architecture on Cloud TPU, IDLE (Python 3.7 64-Bit), Python 3.7.3 Shell, Kaggle datasets, and a desktop computer/laptop with 8GB of RAM were used to create a training and testing program in the Python programming language. 200 images of each cancer class were set aside for testing purposes before the training phase began.

Results

Training

During the training phase, epochs (iterations were increased to train the model. After fully training, the detection accuracy was 99.50% for brain and breast cancer, and 100.00% for colorectal and lung cancer: the average training accuracy for the model was 98.54% (Figure 6).

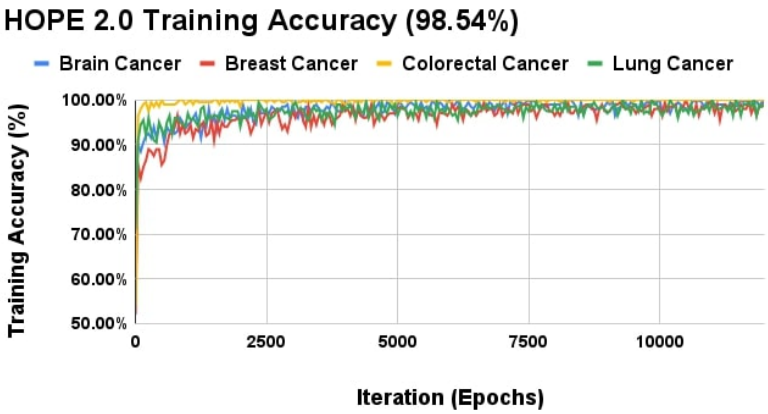

Figure 6. Training accuracy as a function of number of training iterations for each of the four cancers.

Testing

During the testing phase, IDLE was used to conduct test runs with 10 random images of each cancer class (individual tests). After fully testing, the detection accuracy was 100.00% for brain, breast, colorectal, and lung cancer: the average training accuracy for the model was 100.00%. Additionally, independent testing was conducted with images gathered from The Cancer Imaging Archive (TCIA) (Clark et al., 2013). Testing data was drawn from multiple TCIA collections for brain MRI (CPTAC, 2018; Scarpace et al., 2016) and whole slide digital pathology images for breast, lung and colorectal cancer (Saltz et al., 2018). After fully testing, the detection accuracy was 92.00% for brain cancer, 80.00% for breast and colorectal cancer, and 88.00% for lung cancer: the average training accuracy for the model was 92.50% (Figure 7).

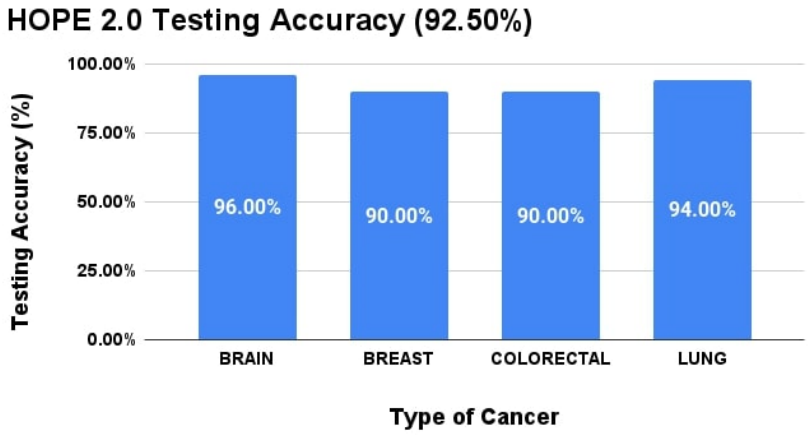

Figure 7. Testing accuracy calculated through independent tests for each of the four cancers.

Discussion

Cancerous tumors are classified as either benign or malignant; benign tumors undergo slow growth and do not spread to the body, whereas malignant tumors undergo rapid growth while spreading to the rest of the body. The four most impactful types of cancer (by prevalence and mortality rate) are brain, breast, colorectal, and lung (Sung et al., 2021).

Comparison of Models

Computer models have been developed that use image classification for disease detection. Models that target brain cancer (Hollon et al., 2020; Kumar et al., 2020), breast cancer (Sadoughi et al., 2018; Sepandi et al., 2018; Zeng & Zhang, 2020), colorectal cancer (Y. Li et al., 2019; Nartowt et al., 2019), or lung cancer (X. Li et al., 2019; Zhao et al., 2020) achieve accuracies ranging between 81.44% and 100%, but they only focus on one specific type of image data. HOPE 2.0 not only achieved high accuracy (95.53%) but also detected multiple types of cancer through multiple image types (radiology and pathology).

Limitations

This study has potential limitations. Selection bias was presented, as the researcher decided which images were fit to be used for training and testing. The datasets acquired might have mislabeled data, which impacts the accuracy. Applicability is also posed as a limitation; currently, the model is saved directly on a computer and is not publicly accessible.

Conclusion

After importing approximately 18,000 images into the model, the Inception v3 HOPE 2.0 classified brain, breast, colorectal, and lung cancer with an overall accuracy of 95.52% $\left(\frac{98.54+[\frac{100+85}{2}]}{2}\right)$. This is a unique model, as it possesses the ability to analyze multiple types of image data: radiology and pathology images. As most algorithms possess an accuracy in the 80-90% range and focus on one specific type of image data, HOPE 2.0 is a state-of-the-art model. HOPE 2.0 may ultimately aid in accelerating the diagnosis of these cancers, resulting in improved clinical outcomes, when compared to previous research which focused on singular cancer diagnosis.

Acknowledgments

I would like to take this opportunity to thank Dr. Fred Prior and Ms. Lee Conrad. Dr. Prior (Professor and Chair, Department of Biomedical Informatics, Professor of Radiology, The University of Arkansas for Medical Sciences (UAMS)) is an expert in the Radiology field. Dr. Prior mentored me towards helping to validate the problem statement, the diagnosis classification, and the efficacy of the solution. Dr. Prior also provided images for independent testing of my model. My chemistry teacher Ms. Conrad guided me towards creating a scholarly presentation, validating assumptions, and mapping the appropriate methodology for the quantitative evaluation of my model.

This material is based upon work supported by the National Science Foundation under Award No. OIA-1946391. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., & Pringle, M. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. Journal of digital imaging, 26(6), 1045-1057. https://doi.org/10.1007/s10278-013-9622-7

CPTAC. (2018). Radiology Data From the Clinical Proteomic Tumor Analysis Consortium Lung Squamous Cell Carcinoma [Cptac-Lscc] Collection [Data Set]. https://doi.org/https://doi.org/10.7937/k9/tcia.2018.3rje41q1

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

Hollon, T. C., Pandian, B., Adapa, A. R., Urias, E., Save, A. V., Khalsa, S. S. S., Eichberg, D. G., D'Amico, R. S., Farooq, Z. U., Lewis, S., Petridis, P. D., Marie, T., Shah, A. H., Garton, H. J. L., Maher, C. O., Heth, J. A., McKean, E. L., Sullivan, S. E., Hervey-Jumper, S. L., Patil, P. G., Thompson, B. G., Sagher, O., McKhann, G. M., Komotar, R. J., Ivan, M. E., Snuderl, M., Otten, M. L., Johnson, T. D., Sisti, M. B., Bruce, J. N., Muraszko, K. M., Trautman, J., Freudiger, C. W., Canoll, P., Lee, H., Camelo-Piragua, S., & Orringer, D. A. (2020). Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. Nature Medicine, 26(1), 52-58. https://doi.org/10.1038/s41591-019-0715-9

Hoyle, R. H. (1995). The structural equation modeling approach: Basic concepts and fundamental issues. Sage.

Kaggle. (2019). BreakHis-v1. https://www.kaggle.com/rangan2510/breakhisv1

Kaggle. (2020a). Brain Tumor Segmentation. https://www.kaggle.com/leaderandpiller/brain-tumor-segmentation

Kaggle. (2020b). Lung and Colon Cancer Histopathological Images. https://www.kaggle.com/andrewmvd/lung-and-colon-cancer-histopathological-images

Kumar, R., Gupta, A., Arora, H. S., Pandian, G. N., & Raman, B. (2020). CGHF: A Computational Decision Support System for Glioma Classification Using Hybrid Radiomics- and Stationary Wavelet-Based Features. IEEE Access, 8, 79440-79458. https://doi.org/10.1109/ACCESS.2020.2989193

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539

Li, X., Hu, B., Li, H., & You, B. (2019). Application of artificial intelligence in the diagnosis of multiple primary lung cancer. Thoracic cancer, 10(11), 2168-2174. https://doi.org/ https://doi.org/10.1111/1759-7714.13185

Li, Y., Eresen, A., Shangguan, J., Yang, J., Lu, Y., Chen, D., Wang, J., Velichko, Y., Yaghmai, V., & Zhang, Z. (2019). Establishment of a new non-invasive imaging prediction model for liver metastasis in colon cancer. Am J Cancer Res, 9(11), 2482-2492.

McCarthy, J. (2007). What is artificial intelligence? Retrieved 9/11/21 from http://35.238.111.86:8080/jspui/bitstream/123456789/274/1/McCarthy_John_What%20is%20artificial%20intelligence.pdf

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS medicine, 6(7), e1000097. https://doi.org/

https://doi.org/10.1371/journal.pmed.1000097
Nartowt, B. J., Hart, G. R., Roffman, D. A., Llor, X., Ali, I., Muhammad, W., Liang, Y., & Deng, J. (2019). Scoring colorectal cancer risk with an artificial neural network based on self-reportable personal health data. PloS one, 14(8), e0221421. https://doi.org/10.1371/journal.pone.0221421
Prior, F., Almeida, J., Kathiravelu, P., Kurc, T., Smith, K., Fitzgerald, T. J., & Saltz, J. (2020). Open access image repositories: high-quality data to enable machine learning research. Clinical Radiology, 75(1), 7-12. https://doi.org/https://doi.org/10.1016/j.crad.2019.04.002
Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. arXiv preprint arXiv:1902.07208.
Sadoughi, F., Kazemy, Z., Hamedan, F., Owji, L., Rahmanikatigari, M., & Azadboni, T. T. (2018). Artificial intelligence methods for the diagnosis of breast cancer by image processing: a review. Breast Cancer (Dove Med Press), 10, 219-230. https://doi.org/10.2147/bctt.S175311
Saltz, J., Gupta, R., & Hou, L. (2018). Tumor-infiltrating lymphocytes maps from tcga h&e whole slide pathology images. https://doi.org/https://doi.org/10.7937/K9/TCIA.2018.Y75F9W1
Sathya, R., & Abraham, A. (2013). Comparison of supervised and unsupervised learning algorithms for pattern classification. International Journal of Advanced Research in Artificial Intelligence, 2(2), 34-38.
Scarpace, L., Mikkelsen, L., Cha, T., Rao, S., Tekchandani, S., Gutman, S., & Pierce, D. (2016). Radiology data from the cancer genome atlas glioblastoma multiforme [TCGA-GBM] collection. https://doi.org/https://doi.org/10.7937/K9/TCIA.2016.RNYFUYE9
Sepandi, M., Taghdir, M., Rezaianzadeh, A., & Rahimikazerooni, S. (2018). Assessing Breast Cancer Risk with an Artificial Neural Network. Asian Pac J Cancer Prev, 19(4), 1017-1019. https://doi.org/10.22034/apjcp.2018.19.4.1017
Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge university press.
Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 71(3), 209-249. https://doi.org/ https://doi.org/10.3322/caac.21660
Torrey, L., & Shavlik, J. (2010). Transfer learning. In Handbook of research on machine learning applications and trends: algorithms, methods, and techniques (pp. 242-264). IGI global.
Turing, A. M. (1950). Computing Machinery and Intelligence. Mind, 49, 433-460. https://www.csee.umbc.edu/courses/471/papers/turing.pdf
Zeng, Y., & Zhang, J. (2020). A machine learning model for detecting invasive ductal carcinoma with Google Cloud AutoML Vision. Computers in Biology and Medicine, 122, 103861. https://doi.org/https://doi.org/10.1016/j.compbiomed.2020.103861
Zhao, L., Bai, C. X., & Zhu, Y. (2020). Diagnostic value of artificial intelligence in early-stage lung cancer. Chin Med J (Engl), 133(4), 503-504. https://doi.org/10.1097/cm9.0000000000000634