Deep Learning for Breast Cancer Classification Using Convolutional Neural Networks

Sérgio Barreto
Centro de Informática (UFPE)
Recife, Brasil
slbp@cin.ufpe.br

Yves Emmanuel
Centro de Informática (UFPE)
Recife, Brasil
yefo@cin.ufpe.br

I. Goals

Developing a neural network model that classifies tumor as benign or malignant can have significant benefits for the medical community. One of the main advantages of this technology is that it can provide a more accurate and reliable diagnosis of breast cancer. With the help of a well-trained machine learning model, medical professionals can more accurately identify cancerous regions in tissue samples, resulting in more effective treatment options for patients. Moreover, an automated classification system can reduce the possibility of human error and misdiagnosis, leading to faster and more accurate diagnosis.

Furthermore, developing a computer-aided diagnosis system can help address the issue of limited access to medical professionals in rural or remote areas. With the help of this technology, healthcare providers in such regions can gain access to a reliable and efficient diagnosis system that can assist them in identifying cancerous tissues. This can result in improved patient outcomes and better overall healthcare for communities that have historically been underserved.

In summary, this model development can have a wide range of benefits for the medical community. It can improve the accuracy and efficiency of breast cancer diagnosis and reduce healthcare costs. As such, this project has the potential to make a significant impact on the lives of millions of people around the world.

Justification

Breast cancer is one of the most common types of cancer in women worldwide, with an estimated 2.3 million new cases diagnosed in 2020 alone (3). Early detection of breast cancer is crucial for successful treatment and improved patient outcomes. Histopathological analysis of breast tissue samples is one of the most widely used methods for breast cancer diagnosis, but it requires skilled pathologists and is time-consuming.

Deep learning models, specifically CNNs, have demonstrated high accuracy in identifying and categorizing breast cancer from histopathological images (9). By creating a precise deep learning model that distinguishes between benign and malignant breast tumor images, this project could help clinicians in detecting and diagnosing cancer early, resulting in better patient results.

The project can also contribute to the development of more efficient and automated methods for breast cancer diagnosis, reducing the dependence on human expertise and increasing the speed and accuracy of diagnosis. Furthermore, the project can pave the way for the development of computer-aided diagnosis systems that can assist pathologists in identifying cancerous regions in the tissue samples, leading to more accurate and efficient diagnosis.

II. Methodology

The methodology for this project involves a series of steps to develop a machine learning model that accurately classifies breast tumors from histopathological images. The following section will discuss each step in detail.

1) Dataset Collection: collect histopathological images of breast tumors from publicly available datasets such as BreaKHis (6), CAMELYON16 (2), and TUPAC16 (7). These datasets are well-established in the medical imaging community and provide a large number of high-quality histopathological images of both benign and malignant breast tumors.
2) Data preprocessing: those techniques have been demonstrated to improve the accuracy of image classification tasks. Pal and Sudeep (2016) (4), for instance, showed that preprocessing techniques can enhance the performance of a deep learning model for image classification. Their study specifically demonstrated that image enhancement techniques such as color correction, normalization and gray scaling can significantly improve the classification accuracy of a convolutional neural network.
3) Model development: this is an essential step in the deep learning process for breast cancer detection, as stated by Wang et al. (2016) (8). To achieve this, we plan to explore different CNN architectures such as VGG, ResNet, Inception, LeNet, AlexNet, and DenseNet and test them on the same datase, so this will help us determine the most appropriate model architecture for the given dataset. Alghamdi et al. (2019) (1) and Hameed et al. (2020) (10) have demonstrated the importance of testing multiple CNN architectures to find the best one for a given dataset.

4) Comparison: the proposed deep learning approach for breast cancer detection using histopathological images can be evaluated by comparing its performance with other state-of-the-art models. This would help identify areas for improvement, strengths, and weaknesses of the proposed model. Such a comparison would provide insights into the model's potential to improve diagnostic accuracy and its applicability in real-world scenarios.

Schedule of Activities

- **Week 1:**
  - Choose and evaluate the best dataset of histopathological images from the medical center
  - Collect and preprocess histopathological images from the medical center dataset
  - Research existing deep learning models and techniques for breast cancer classification
  - Design and implement a baseline CNN model for breast cancer classification
- **Week 2:**
  - Train and validate the baseline CNN model using a portion of the dataset
  - Explore and experiment with different CNN architectures and hyperparameters
  - Evaluate the baseline model and compare its performance with other models
- **Week 3:**
  - Fine-tune the best-performing model on the full dataset
  - Assess the model's accuracy and generalization ability through cross-validation
  - Conduct statistical analysis of the results
- **Week 4:**
  - Prepare a report summarizing the project's findings and results
  - Create a presentation to share the project's outcome
  - Evaluate the limitations and potential future directions of the project
  - Finalize and submit the project report

References

[1] Alghamdi, W., Hussain, M., Hussain, A. (2020). *Analysis of Convolutional Neural Network Architectures for Chest Radiography Classification.* Scientific Reports, 10(1), 1-13. https://doi.org/10.1038/s41598-020-70479-z

[2] Bejnordi, B. E., Veta, M., van Diest, P. J., van Ginneken, B., Karssemeijer, N., Litjens, G., ... Geessink, O. G. (2017). **Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer**. JAMA, 318(22), 2199-2210. doi: 10.1001/jama.2017.14585. Available at: https://jamanetwork.com/journals/jama/fullarticle/2665774

[3] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., Jemal, A. (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. doi: 10.3322/caac.21660.

[4] Pal, K. K. and Sudeep, K. S. (2016). Preprocessing for image classification by convolutional neural networks. In *2016 IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, pages 1778-1781. doi: 10.1109/RTEICT.2016.7808140. Available at: https://ieeexplore.ieee.org/document/7808140.

[5] S. Vani Kumari and K. Usha Rani, "Analysis on Various Feature Extraction Methods for Medical Image Classification," *Advances in Computational and Bio-Engineering*, pp. 19-31, 2020, Springer International Publishing. doi: 10.1007/978-3-030-46943-6_3.

[6] Spanhol, F. A., Oliveira, L. S., Petitjean, C., Heutte, L. (2016). **Breast cancer histopathological image classification using convolutional neural networks**. In *International Joint Conference on Neural Networks (IJCNN)* (pp. 2560-2567). IEEE. doi: 10.1109/IJCNN.2016.7727519. Available: https://ieeexplore.ieee.org/document/7727519

[7] Veta M, Heng YJ, Stathonikos N, Bejnordi BE, Beca F, Wollmann T, Rohr K, Shah MA, Wang D, Rousson M, Hedlund M, Tellez D, Ciompi F, Zerhouni E, Lanyi D, Viana M, Kovalev V, Liauchuk V, Phoulady HA, Qaiser T, Graham S, Rajpoot N, Sjöblom E, Molin J, Paeng K, Hwang S, Park S, Jia Z, Chang EI, Xu Y, Beck AH, van Diest PJ, Pluim JPW. **Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge**. Med Image Anal. 2019 May;54:111-121. doi: 10.1016/j.media.2019.02.012. Epub 2019 Feb 27. Erratum in: Med Image Anal. 2019 Aug;56:43. PMID: 30861443. Available at: https://pubmed.ncbi.nlm.nih.gov/30861443/

[8] Wang, D., Khosla, A., Gargeya, R., Irshad, H., Beck, A. H. (2016). Deep learning for identifying metastatic breast cancer. arXiv preprint arXiv:1606.05718. Available at: https://arxiv.org/abs/1606.05718.

[9] Xie, J., Liu, R., Luttrell, J., Zhang, C. (2019). **Deep Learning Based Analysis of Histopathological Images of Breast Cancer**. Frontiers in Genetics, 10, 80. doi: 10.3389/fgene.2019.00080. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00080

[10] Z. Hameed, S. Zahia, B. Garcia-Zapirain, J. Javier Aguirre, and A. María Vanegas, *Breast Cancer Histopathology Image Classification Using an Ensemble of Deep Learning Models,* Sensors, vol. 20, no. 16, p. 4373, Aug. 2020, doi: 10.3390/s20164373.