DOI: http://doi.org/10.34614/2023IYRC_F61

BREAST HISTOPATHOLOGY IMAGE CLASSIFICATION FOR END-TO-END DIAGNOSIS USING TRANSFORMERS ARCHITECTURE

Mohanad A. Elagan

STEM High School for Boys - 6th of October, Egypt; mohanad.elagan1@gmail.com

ABSTRACT: One of the most significant public health concerns, breast cancer, is regarded as the primary cause of cancer-related deaths in women globally. The clinical benefit of a computer-aided diagnostic system that may reduce uncertainty is significant. Therefore, the method used in this paper is transfer learning for getting good results on downstream tasks, even with little data, to make the diagnostic procedures more successful. The popularity of vision transformers has also expanded the alternatives for picture models that were previously accessible. Using cutting-edge vision transformer models CaiT, ViT, and BeiT, the use of transfer learning in the context of breast histopathology in this research is investigated. The methods for making model behavior and prediction are accessible to humans, so a pathologist may use them to help make a diagnosis. Through the three sections in the results, the experiments show how latent representations and attention mappings may be utilized to analyze model behavior.

Keywords: histopathology, image classification, breast cancer, transfer learning, attention map

1. INTRODUCTION

One of the most critical global public health challenges is cancer. The Global Burden of Disease (GBD) research estimates that there were 9.6 million cancer deaths and 24.5 million new cancer cases globally in 2017 [Fitzmaurice, C., 2019]. According to these figures, the global cancer incidence increased by 33% between 2007 and 2017 [Fitzmaurice, C., 2019]. Therefore, early detection of this condition is essential to stop its development and lower morbidity rates among women.

Traditional pathology diagnoses have great regard among doctors. The pathologist examines tissue slices under a microscope and determines the appropriate cancer diagnosis by examining the slices' tissue composition and cytopathic features. The integrity of the final pathological slice picture may be impacted by the staining density and flatness of the slice, as well as the collecting and storage of pathological slice images. The breast histological pictures' intrinsic complexity and variety make pathologists' diagnostic tasks arduous and time-consuming. [Veta, M., 2014]

These effects have been diminished by the advent of digital pathology, which is beneficial for getting high-resolution photographs [Niazi, M.K.K., 2019]. Digital pathology, as opposed to conventional pathology, employs digital pathology systems to network, process, and digitize pathological materials. Data collecting visualization, long-term archiving, and simultaneous browsing may be employed thanks to the use of big data technologies in the medical industry. Time and place no longer limit how problematic materials are processed after collection. As a result, digital pathology is now extensively employed in pathology-related domains [Yaffe, M.J., 2019].

Various elements, including medical knowledge and tools, influence diagnostic accuracy (IDC). Pathologists concentrate on areas of a sample containing IDC to give a whole mount sample an aggressive grade. Convolutional neural networks (CNNs) have been used in earlier publications [Avishek Choudhury. 2021, Angel Cruz-Roa., 2014, Fei Gao., 2018, Carol E DeSantis., 2017] for data-driven tumorous area detection to speed up this laborious procedure. But more recently, it has been shown that vision transformers perform better on picture categorization tasks than CNNs. Simultaneously, explainability in artificial intelligence is a current study topic. The rise of intricate, large-scale models emphasizes the need for approaches to demystify the mysterious neural network.

Radiology pictures, such as those from mammography, ultrasound imaging, and magnetic resonance imaging (MRI), are first used for clinical screening [Dromain, C., 2013, Wang, L., 2017]. Nevertheless, it's possible that these non-invasive imaging techniques can't accurately identify the malignant regions. In order to do this, a more thorough analysis of the malignancy in breast cancer tissues is often performed using the transfer learning procedure. As part of the transfer learning procedure, tissue samples are collected, mounted on microscopic glass slides, and stained to make the nuclei and cytoplasm more visible [Veta, M., 2014]. The ultimate diagnosis of

breast cancer is subsequently made by pathologists after microscopic examination of these slides [Veta, M., 2014]. Figure 1 shows the complete transfer learning procedure process. The 1,000 categories of natural picture categories in the ImageNet dataset were used to train the classification algorithm CNN-1. CNN-3 is the classification model of tumor molecular subtypes, and CNN-2 is the benign and malignant tumor diagnostic model generated by model transfer and retraining based on the CNN-1 model, respectively. Convolution neural network, or CNN.

In this study, transfer learning is used to apply vision transformers to a dataset of breast histology. The attention maps of the models were examined, which are taken straight from the network's self-attention layers, to understand better how they behave. Vision transformers are demonstrated, which can provide precise predictions, and displaying their attention map and latent representations may provide a further understanding of model behavior.

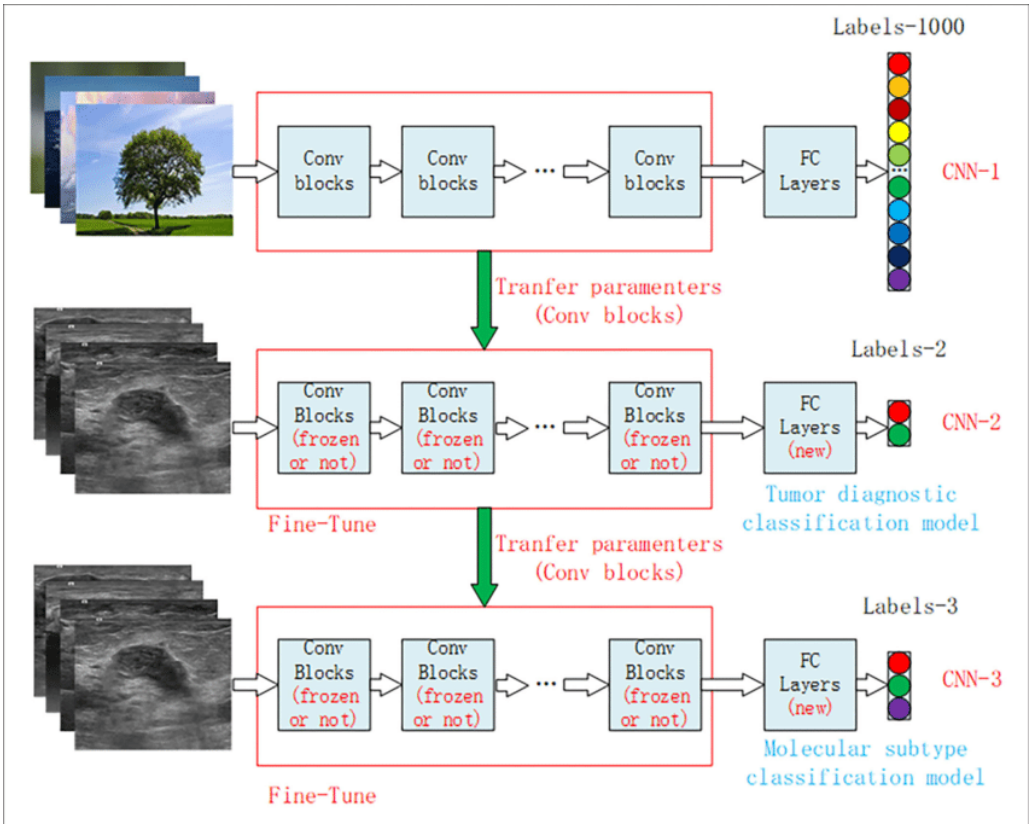


Figure 1. Diagrammatic representation of the transfer learning model for ultrasound pictures of breast cancer.

2. RELATED WORK

Numerous research used handmade features-based algorithms to categorize histopathology pictures linked to breast cancer before the development of machine learning in biomedical engineering.

Additionally, showing interest in the segmentation of nuclei, Filipczuk et al., 2013 retrieved 25 shape- and texture-based characteristics from the segmented nuclei of 737 breast cancer cytology pictures. Four distinct machine learning classifiers, KNN (K-nearest neighbor), NB (Naive Bayes), DT (decision tree), and SVM (support vector machine), were developed to classify these cytological pictures into benign and malignant instances based on these attributes.

Using the BreakHis dataset [Spanhol, F.A., 2021], Bayramoglu et al., 2016 used CNN to categorize the histopathology pictures of breast cancer regardless of their resolution. The scientists specifically suggested single-task and multi-task CNN architectures, the latter of which could predict both the intensity of pictures' magnification

and their malignancy at the same time. This research combined the breast Histopathology Images dataset [Spanhol, F.A., 2021] with other datasets.

2.1 BeiT

BEiT [Hangbo Bao, 2022] was pre-trained in a self-supervised manner with a masked image modeling target, taking inspiration from BERT [Jacob Devlin, 2019]. In a challenge that precisely resembles BERT's masked language modeling job, BeiT was trained to predict the visual tokens from the vector quantization codebook of DALL-E [Aditya Ramesh, 2021] given masked patches. BeiT demonstrated that pre-trained models using self-supervised pretraining techniques can outperform supervised pretraining approaches. To determine if enhanced pretraining methods correspond to better downstream performance, BeiT with ViT are contrasted.

2.2 ViT

The first effective use of transformers in computer vision was ViT [Alexey Dosovitskiy, 2021]. ViTs suggested using raster-scan ordering to build embeddings utilized as input to the transformer backbone by dividing a picture into patches. ViTs use a classifier token, which is utilized to pool intermediate representations and produce a final prediction, similar to transformers for text classification. ViT is employed as a competitive baseline model in this experiment.

2.3 CaiT

Late categorization token insertion is suggested by CaiT [Hugo Touvron, 2021]. CaiT does self-attention primarily with picture patches for the bulk of its layers and inserts the classification token at the end, in contrast to ViT, which adds the classification token right at the start of the network.

By separating representation learning from the classification problem, these last layers simply calculate attentions between the classification token and the latent picture representations. Using CaiT, it is tested if its architectural upgrades give performance advantages over the traditional ViT architecture.

3. METHODOLOGY

3.1 Data

The data [Paul Mooney, 2017] is presented in patches of 50*50 pixels since breast histopathology pictures are often vast and frequently comprise many sections of cancer cells. Each image has a binary name, coordinate data specifying the position of each pixel, and a specific patient ID. There are 279 patient IDs and 277,524 pictures altogether. The remaining 5% of the data are utilized for validation, leaving 95% for training. The picture is enlarged to 225*225 pixels as part of typical preprocessing, and it is then normalized using the mean and standard deviation data from ImageNet.

3.2 Experimental Setup

This section describes the experimental setup and how the suggested method's evaluation metrics should be interpreted.

3.2.1 Training

It is well known that large transformers may readily overfit to downstream demands. Early halting and data augmentation were used via horizontal and vertical flipping to avoid overfitting. The final adapter classifier and the vision transformer backbone both received distinct learning rates.

The classifier is randomly initialized, while the transformer backbone is started with pre-trained weights, which serve as the basis for this choice. In a similar vein, it experiments with total backbone freezing, where the weight changes are only applied to the adapter layer, and the vision transformer as a whole is completely frozen.

The ratio of negative to positive samples is around three to one. Dataset imbalance is anticipated since IDC cells typically only make up a tiny portion of a whole mount sample. The model would produce false negatives on this dataset if it were naively trained. A weight-adjusted binary cross entropy loss function is employed to reduce this imbalance, which penalizes misclassifying a positive ground truth sample more severely than a negative one according to the imbalance of the dataset.

3.2.2 Implementation

Using Python 3.11.4 (and TensorFlow 2.1.0) installed on a typical PC and with a batch size of 96 for complete fine-tuning and 1024 for backbone-frozen setups, models were trained on an Nvidia GeForce RTX 3070 graphics processing units (GPU) capability. Additionally, this computer contains an overclocked 4.0 GHz Intel Core i7-

11800H CPU with 16 logical threads and 24 MB of cache memory, along with 32.0 GB of RAM. With a learning rate of 1*10^-5 for the final classifier and 1*10^-10 for the backbone in the non-frozen scenario, the AdamW optimizer is employed [Ilya Loshchilov, 2019].

3.2.3 Assessment Metrics

The components of the confusion matrix, also known as the error matrix or contingency table, are crucial to the overall performance of our suggested model. Four terms—True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN)—are included in this assessment matrix. In our example, the TP stands for photos properly identified as carcinoma, whereas the FP stands for images incorrectly identified as non-carcinoma. While the FN stands for carcinoma class pictures that were incorrectly labeled as non-carcinoma, and the TN for properly identified non-carcinoma images. Using the Python sci-kit-learn module, four performance metrics based on the confusion matrix—precision, sensitivity (recall), overall accuracy, and F1-score—were used to assess the classification performance of our proposed model on the testing set. These performance metrics may be computed using the formulas below:

1. Sensitivity, often known as "recall," calculates a model's degree of completeness. It shows the proportion of photos correctly identified as carcinoma compared to all images.
$$Sensitivity = TP/(FN + TP)$$
2. A model is often optimized towards either accuracy or recall using the F1-score, which indicates the harmonic average of precision and recall.
$$F1 - score = (2 * Percision * Recall)/(Percision + Recall)$$
3. Accuracy measures a model's performance and is calculated as the availability of correctly categorized phomages.
$$Accuracy = (TP + FN)/(TN + FN + FP + TP)$$
4. Precision measures a model's accuracy and is expressed as the proportion of cancer photos correctly identified from all anticipated images of the same class.
$$Percision = TP/(TP + FP)$$

4 Results and Tests

Precision (F1 score) and Matthew's Correlation Coefficient (MCC) were the four metrics that are used to evaluate the model's performance. To take advantage of data augmentation during training, test-time augmentation (TTA) is used, a widely utilized approach. Augmentation is used to produce several data points from a single picture rather than feeding the model a single image. In this instance, vertical and horizontal flips are used to produce four pictures per sample. Then, in order to provide a single output, the four projections are averaged. Given post-sigmoid activations between 0 and 1, a cutoff limit of 0.5 is chosen. In Table 1, results are shown.

Table 1. Results of the **ViT-full** model in comparison with other models.

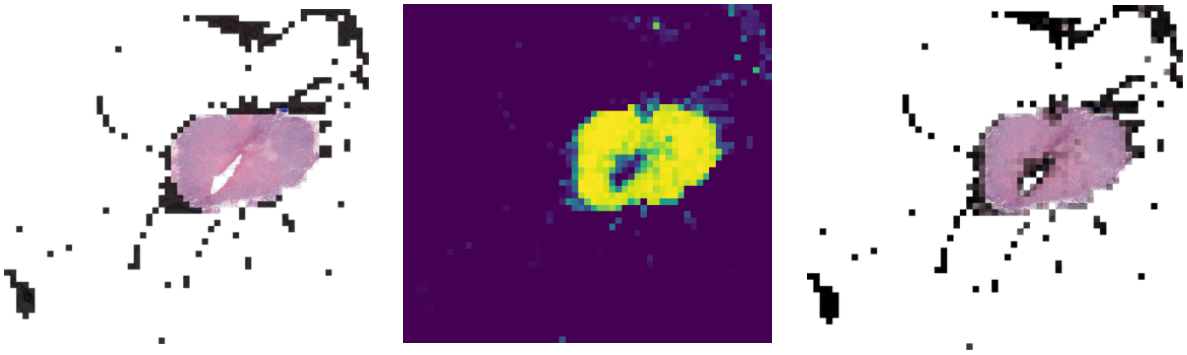
Model	Trainable Params	Recall	F1	Precision	MCC
ViT-full	86M	0.847	**0.789**	**0.736**	**0.722**
ViT-freeze	785	0.843	0.785	0.735	0.717
BeiT-full	86M	0.857	0.785	0.726	0.719
BeiT-freeze	785	0.846	0.783	0.729	0.715
CaiT-full	47M	0.786	0.750	0.717	0.671
CaiT-freeze	387	0.787	0.747	0.711	0.667

ViT, with all of the adjustments made, performed best overall, closely followed by BeiT and CaiT. Thus, the performance of a pre-trained model does not always correspond to the downstream environment. Additionally, CaiT's poor performance is blamed—despite architectural advancements—on its fewer trainable parameters than ViT or BeiT-base. Additionally, It is found that models with frozen and non-frozen backbones perform somewhat differently. Given that backbone freezing significantly restricts model capacity and fine-tuning enables the model to tune its weights to the downstream task with more degrees of freedom, this outcome is predicted.

The model only uses 50*50 photo patches; thus, a full-mount sample can be created by sewing the patches together.

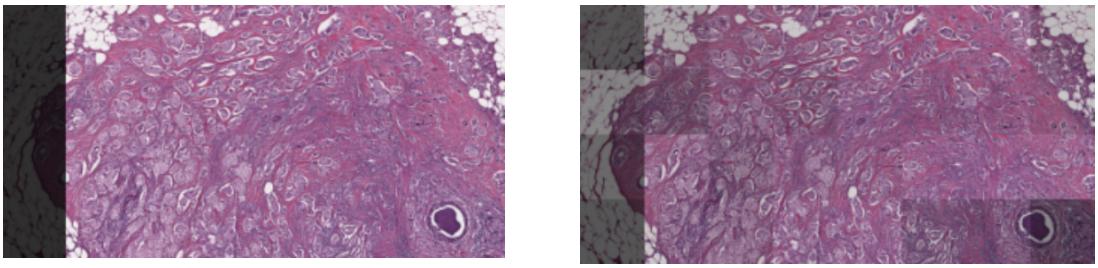
It can be got a heat map of predictions by performing model inference over the full dataset. Figure 2 illustrates how the model properly identifies the malignant area but incorrectly labels the tumorous ring's inside as negative. By looking at these failed situations, One may get more information on misclassification in greater depth. In Figure

3, the model properly detects the sample's far left corner as negative but incorrectly classifies an odd tissue form in the bottom right as negative. These failure nodes may be utilized as signals to direct future data gathering throughout the machine learning pipeline's lifespan.



(a) Ground truth (b) Prediction heat map (c) Overlaid

Figure 2. Prediction using the BeiT-freeze whole mount sample. The model recognizes the center of the cancerous area with accuracy, however around its inner ring there are some false negatives.



(a) Ground truth (b) Prediction model

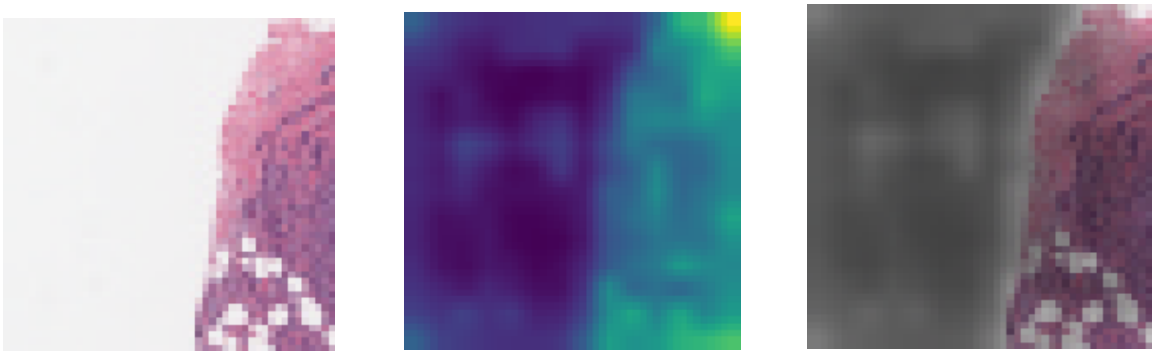
Figure 3. ViT-freeze's per-patch anticipate on a mixed area. The model properly labels the bottom right as also being negative but only with a limited degree of confidence.

4.1 Aggregate Attention Map

Utilizing vision transformers has the advantage of offering a direct method of studying model behavior via self-attention matrices. The attention score mix of the i-th layer is written as $CapCapA_i \in R^{L*L}$ given a sequence length of L. Matrix multiplication is cumulatively applied on the attention score matrices to acquire the overall model attention across all levels. Let $\hat{A}$ e whole attention matrix. Then, this matrix may be represented geometrically as $\hat{A} = \Pi_{i=1}^{N} no(\hat{A} + \Pi_L)$, where II stands for an identity matrix and N for the quantity of transformer encoder layers. The aggregate impact of any attention score matrix may be intuitively comprehended by looking at its product if it reflects a change. The outcome is normalized to retain the attention matrix condition that each row adds to 1 in order to account for any remaining connections between the layers.

The top row of the aggregate attention matrix is indexed, which comprises attention scores between the classification token and every other picture patch, as this study aims to better understand how the model generates a prediction. The row is scaled and restructured to match the original image's proportions to create an attention map.

The attention map in Figure 4 demonstrates which area of the picture the model focused on in relation to the categorization token. The area of the picture with the cell tissues receives the majority of the model's attention, as predicted, whereas vacant space is avoided. Nonetheless, practically, the attention map often lacks the depth and granularity needed to provide a better understanding of which texture or patterns were primarily responsible for the output. Aggregate attention maps, however, may act as a fast and easy sanity check for ViT models.



(a) Ground truth (b) Aggregate attention (c) Overlaid

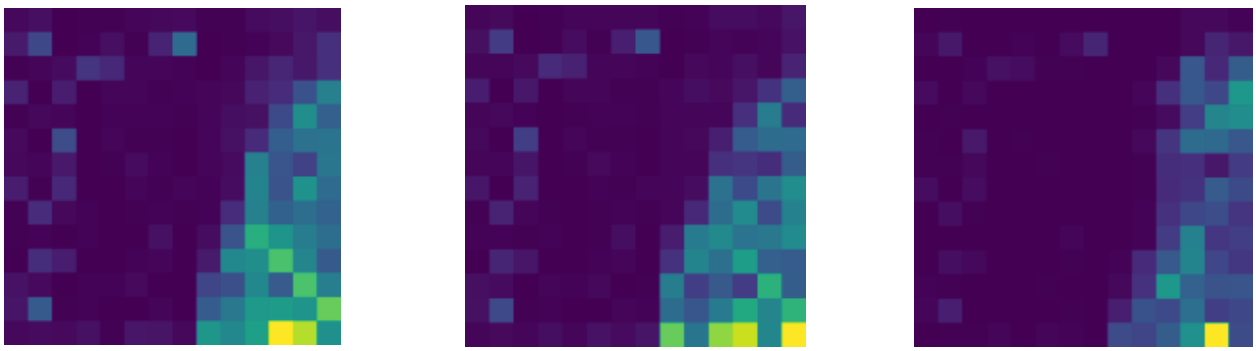
Figure 4. Map of collective attention taken from ViT-freeze. While appropriately focusing on the important area of the picture, the model overlooks the majority of the vacant space.

4.2 Last Attention Map

A straight application of aggregate attention maps to CaiT is not possible due to late classifier insertion. Therefore, alternate applications for CaiT's attention output was investigated. First, the output of the final layer from CaiT is extracted, drawing inspiration from DINO [Mathilde Caron, 2021], which creates attention map visualizations using the output of the last layer's output on attention. In the unrolled series of picture patches, this attention output—which comprises attention scores between the classification token and each other token—is a vector, not a matrix. The attention vector is resized and interpolated in the manner of DINO; then it is overlaid on top of the input picture. A threshold value of = 0.6 is defined to clip very low attention scores.

Applying DINO's attention visualization technique to CaiT results in interpretable attention maps, as shown in Figure 5. The bottom right corner of the picture, which has erratic patterns indicative of IDC, seems to get the most out of the model's focus. Compared to the aggregate attention approach, which created essentially homogenous maps with little variation over the attended area, the outcome is different. It is proposed that late token categorization for disentangled representation learning facilitates the retrieval of attention mappings from the network's final layer.

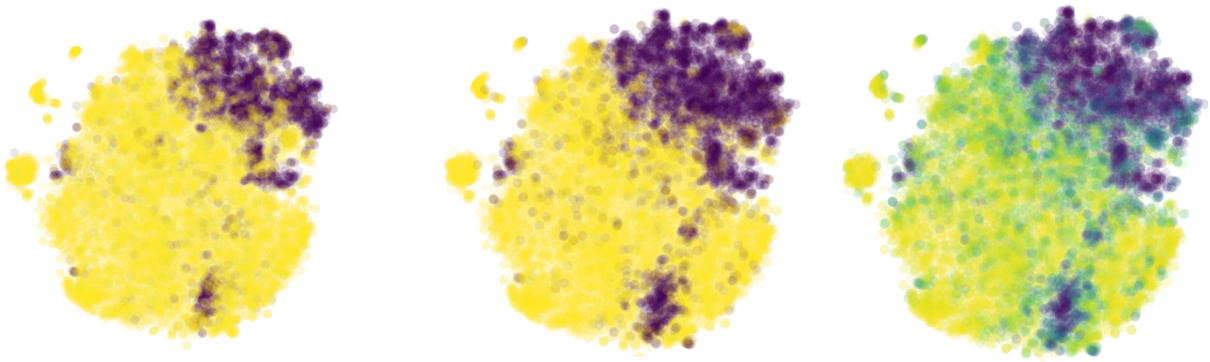
4.3 Latent Representations



(a) Head mean (b) Head 1 (c) Head 2

Figure 5. Using the ground truth picture from Figure 4a, the last attention map was recovered from CaiT-freeze. It can be seen 2 of the 6 self-attention heads.

The pooled latent representations of the model serves as the final classifier layer's inputs, to investigate various theories for explaining model behavior. The latent may be expressed specifically as the $H \in R^{d_h}$, where d_h stands for the model's hidden size. t-SNE is used [Laurens van der Maaten, 2008] to decrease d_h's dimensionality to R^2 since it often has a big value, such as 768. Figure 6 displays the outcomes.



(a) Ground truth (b) Predictions at 0.5 cutoff (c) Post-sigmoid activations

Figure 6. Ground truth, predictions, and logits from Beit-Freeze are used to create t-SNE visualizations. Near cluster borders, predictions become inaccurate.

Positive data points are largely grouped on the upper right, with a smaller cluster on the bottom, as seen by the colors assigned to each point in two-dimensional space using ground truth labels. Additionally, there are some beneficial aspects that do not fit into any one group. A similar image is produced when the model's anticipated labels are used. Nevertheless, it can be seen that the model makes unreliable predictions close to cluster borders. A rise in false positives that are not part of any cluster also seems to be occurring. It can be more easily observed that the model is less sure at cluster borders and more definite about negative predictions on the bottom left with the brightest yellow hue when coloring the points in accordance with the output of the final sigmoid activation function. In addition to offering greater insight into model behavior than scalar measurements, these visualizations may be utilized to choose a decision threshold that is more appropriate for the choice between positive and negative samples. [Lindroos, J., 1970]

5. CONCLUSION

This study uses transfer learning to vision transformers to identify breast cancer. By using model attention outputs and latent representations, it is concentrated on the interpretability of the model. The aggregate attention map approach, which may be used with typical ViT variations, aims to explain the model's overall attention throughout the network's layers. The last attention map visualization is used, inspired by DINO for CaiT, which suggests late classification token insertion to decouple representation learning from the classification challenge. When used in conjunction with a dimensionality reduction method, latent representations may be a valuable tool for examining model behavior. These findings point to the potential interpretable use of vision transformer models to support IDC diagnosis.

6. FUTURE WORK

In digital pathology, WSI is often used to forecast gene mutations, molecular subtypes, and clinical consequences. As a result, they are often separated into patches for training neural networks and prediction models. However, it was not possible to categorize each patch since patch-level tags are often absent properly. Numerous studies have used gene expression patterns to comprehend the molecular properties of breast cancer throughout the last several decades, thanks to the quick development of high-throughput technology for microarrays and gene expression analysis. A preliminary investigation was carried out by Van de Vijver, 2009, to successfully estimate the prognosis of breast cancer using the gene expression profile. Data from gene expression profiles were grouped and associated with prognostic values. The accuracy of prognostic and diagnostic prediction models may be increased by the combination of gene expression profile data with clinical data [Khademi, M., 2015]. Each patient's microarray data comprises roughly 25,000 genes and is high-dimensional. The accuracy of prognosis prediction for breast cancer may be improved by possible correlations between several genes [Lee, E.S., 2008]. There are several breast cancer-related genes known. Oncogene and tumor suppressor gene mutation and aberrant amplification are major contributors to the formation and growth of malignancies. For instance, the human epidermal growth factor receptor 2, commonly known as c-erbB-2, and two well-known breast cancer risk genes, BRCA1 and BRCA2, are major breast cancer carcinogens.

The probability graph model (PGM) [Khademi, M., 2015], developed by Khademi et al., integrates two independent microarray models with clinical data to predict and diagnose breast cancer. Prior to building a depth

confidence network to extract the feature representation of the data, they used principal component analysis (PCA) to minimize the dimensionality of the microarray data. They also used structural learning algorithms on clinical data at the same time. However, there is currently a lot of work to directly provide slide-level prediction through deep learning, and digital whole image (WSI) may provide a computationally effective and efficient method to quantitatively characterize the heterogeneity of cancer specimen cell level, motivated by the successful application of deep learning methods in the cv field and the enormous contribution of multidimensional data to cancer prognosis prediction. WSIs are often used by pathologists to determine nuclear characteristics, determine the stage of cancer, and assess the histopathological grading of cancer tissues. According to preliminary research, the use of deep learning techniques can automatically identify different cancer subtypes [Stone, P.C., 2007], predict lung and liver cancer mutations [Martin, L.R., 2005], classify mesotheliomas [Khademi, M., 2015], find DNA methylation patterns [Sun, Y., 2007], determine the human epidermal growth factor receptor status in breast cancer [Gevaert, O., 2006], and forecast the overall prognosis of patients for cancer. However, pan-cancer research cannot fully describe breast cancer histology, mutation, and pathway activity level [Nguyen, C., 2013]. By employing biomarkers that are now invisible to doctors, DL based on CNN may currently predict the gene mutation status in H&E-stained WSIs and has the potential to enhance cancer prediction and therapy. Gene mutation prediction may be used as a prescreening to increase the cost-effectiveness before next-generation sequencing, enhancing precision medical treatment's performance, even if artificial intelligence cannot totally replace humans in practice.

REFERENCES

Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In International Conference on Learning Representations, 2021.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In International Conference on Learning Representations, 2021.

Angel Cruz-Roa, Ajay Basavanhally, Fabio González, Hannah Gilmore, Michael Feldman, Shridar Ganesan, Natalie Shih, John Tomaszewski, and Anant Madabhushi. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In Medical Imaging 2014: Digital Pathology, volume 9041, page 904103. SPIE, 2014.

Avishek Choudhury and Sunanda Perumalla. Detecting breast cancer using artificial intelligence: Convolutional neural network. Technology and Health Care, 29(1):33–43, 2021.

Bayramoglu, N.; Kannala, J.; Heikkila, J. Deep learning for magnification independent breast cancer histopathology image classification. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancún, Mexico, 4–8 December 2016.

Carol E DeSantis, Jiemin Ma, Ann Goding Sauer, Lisa A Newman, and Ahmedin Jemal. Breast cancer statistics, 2017, racial disparity in mortality by state. CA: a cancer journal for clinicians, 67(6):439–448, 2017.

Dromain, C.; Boyer, B.; Ferré, R.; Canale, S.; Delaloge, S.; Balleyguier, C. Computed-aided diagnosis (CAD) in the detection of breast cancer. Eur. J. Radiol. 2013, 82, 417–423.

Evaluating the Accuracy of Breast Cancer and Molecular Subtype Diagnosis by Ultrasound Image Deep Learning Model - Scientific Figure on ResearchGate.

Fei Gao, Teresa Wu, Jing Li, Bin Zheng, Lingxiang Ruan, Desheng Shang, and Bhavika Patel. Sd-cnn: A shallow-deep cnn for improved breast cancer diagnosis. Computerized Medical Imaging and Graphics, 70:53–62, 2018.

Filipczuk, P.; Fevens, T.; Krzyzak, A.; Monczak, R. Computer-Aided Breast Cancer Diagnosis Based on the Analysis of Cytological Images of Fine Needle Biopsies. IEEE Trans. Med. Imaging 2013, 32, 2169–2178.

Fitzmaurice, C. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: A systematic analysis for the global burden of disease study. *JAMA Oncol.* **2019**, *5*, 1749–1768.

Gevaert, O.; Smet, F.D.; Timmerman, D.; Moreau, Y.; Moor, B.D. Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. Bioinformatics 2006, 22, e184–e190.

Hangbo Bao, Li Dong, and Furu Wei. Beit: BERT pre-training of image transformers. In International Conference on Learning Representations, 2022.

Hugo Touvron, Matthieu Cord, Alexandre Sablayrolles, Gabriel Synnaeve, and Hervé Jégou. Going deeper with image transformers, 2021.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In International Conference on Learning Representations, 2019.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2019.

Khademi, M.; Nedialkov, N.S. Probabilistic graphical models and deep belief networks for prognosis of breast cancer. In Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 9–11 December 2015; pp. 727–732.

Laurens van der Maaten and Geoffrey E. Hinton. Visualizing data using t-SNE. Journal of Machine Learning Research, 9:2579–2605, 2008.

Lee, E.S.; Son, D.S.; Kim, S.H.; Lee, J.; Jo, J.; Han, J.; Kim, H.; Lee, H.J.; Choi, H.Y.; Jung, Y.; et al. Prediction of recurrence-free survival in postoperative non–small cell lung cancer patients by using an integrated model of clinical information and gene expression. Clin. Cancer Res. 2008, 14, 7397–7404.

Lindroos, J. (1970, January 1). Transformers for breast cancer classification. JYX. https://jyx.jyu.fi/handle/123456789/81505

Martin, L.R.; Williams, S.L.; Haskard, K.B.; DiMatteo, M.R. The challenge of patient adherence. Ther. Clin. Risk Manag. 2005, 1, 189.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In International Conference on Computer Vision, 2021.

Nguyen, C.; Wang, Y.; Nguyen, H.N. Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. J. Biomed. Sci. Eng. 2013, 6, 31887.

Niazi, M.K.K.; Parwani, A.V.; Gurcan, M.N. Digital pathology and artificial intelligence. Lancet Oncol. 2019, 20, e253–e261.

Paul Mooney. Breast histopathology images. https://www.kaggle.com/paultimothymooney/breast-histopathology-images.

Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. A Dataset for Breast Cancer Histopathological Image Classification. IEEE Trans. Biomed. Eng. 2016, 63, 1455–1462.

Stone, P.C.; Lund, S. Predicting prognosis in patients with advanced cancer. Ann. Oncol. 2007, 18, 971–976.

Sun, Y.; Goodison, S.; Li, J.; Liu, L.; Farmerie, W. Improved breast cancer prognosis through the combination of clinical and genetic markers. Bioinformatics 2007, 23, 30–37.

Van De Vijver, M.J.; He, Y.D.; Van't Veer, L.J.; Dai, H.; Hart, A.A.; Voskuil, D.W.; Schreiber, G.J.; Peterse, J.L.; Roberts, C.; Marton, M.J.; et al. A gene-expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. 2002, 347, 1999–2009.

Veta, M.; Pluim, J.P.; Van Diest, P.J.; Viergever, M.A. Breast cancer histopathology image analysis: A review. IEEE Trans. Biomed. Eng. 2014, 61, 1400–1411.

Wang, L. Early Diagnosis of Breast Cancer. Sensors 2017, 17, 1572.

Yaffe, M.J. Emergence of "Big Data" and Its Potential and Current Limitations in Medical Imaging. Semin. Nucl. Med. 2019, 49, 94–104.