Detection of the PIK3CA Mutation in Breast Cancer

Ulysse Trin
EPITECH
ulysse.trin@hotmail.fr

Matteo Sammut
Paris-Dauphine
matteo.sammut@dauphine.eu

Abstract

Breast cancer is a complex disease that poses significant challenges for diagnosis and treatment. The **PIK3CA** gene plays a crucial role in tumor development, and its mutation is a key biomarker for personalized therapies. However, traditional mutation detection methods rely on manual histopathological review and DNA sequencing, which are both time-consuming and resource-intensive. In this work, we propose a novel weakly supervised learning approach using multiple instance learning (MIL) to predict the presence of the PIK3CA mutation directly from histopathological slides. By extracting tiles from whole-slide images and employing an adaptive attention-based MIL pooling mechanism, our approach aggregates local tile features into a global bag representation. Furthermore, we incorporate domain adaptation techniques to address the heterogeneity across data originating from different medical centers. Ensemble learning and focal loss are also used to further boost performance. Our method has shown promising results, with an AUC of 0.73 on the public test set, and we discuss further strategies to push performance towards an AUC of 0.80.

1 Introduction

Histopathology is a cornerstone in the diagnosis and treatment planning of cancers, including breast cancer. The analysis of tissue slides allows pathologists to identify morphological features that correlate with genetic mutations. Among these, mutations in the **PIK3CA** gene are of particular interest due to their implications for targeted therapies. However, conventional approaches for mutation detection require extensive manual annotation and specialized sequencing techniques, which can limit their widespread adoption.

In response to these challenges, automated methods based on machine learning have been proposed. In particular, the Owkin Data Challenge formulated the problem as a weakly supervised binary classification task, where each whole-slide image (WSI) is associated with a single label (mutated or wild-type) while being composed of thousands of tile images. This setup naturally leads to the use of multiple instance learning (MIL), where a bag of instances (tiles) is assigned one label.

2 Methodology

2.1 Multiple Instance Learning (MIL)

In the standard supervised learning framework, each instance $x \in \mathbb{R}^D$ is associated with a label $y \in \{0, 1\}$. In the MIL setting, however, we are given a bag of instances

$$X = \{x_1, x_2, \ldots, x_K\},$$

with a single bag-level label $Y \in \{0, 1\}$. The MIL assumption is that

$$Y = \begin{cases} 0, & \text{if } \sum_{k=1}^{K} y_k = 0, \\ 1, & \text{otherwise,} \end{cases}$$

or equivalently,

$$Y = \max_{k \in \{1, \dots, K\}} y_k.$$

Since the instance-level labels y_k are not available during training, the model must learn to aggregate information from all instances in a permutation-invariant manner.

2.1.1 MIL Pooling and Permutation Invariance

The bag probability $\theta(X)$ must be invariant to the order of instances. By the Fundamental Theorem of Symmetric Functions, any permutation-invariant function can be decomposed as

$$S(X) = g\left(\sum_{x \in X} f(x)\right),$$

where f transforms each instance and g aggregates these transformed features into a global score. A common choice is the max-operator; however, this is often problematic for gradient-based optimization. Instead, we propose an attention-based pooling mechanism.

2.2 Attention-Based MIL Pooling

Our approach uses a weighted average of the instance embeddings:

$$z = \sum_{k=1}^{K} a_k h_k,$$

where h_k are the feature embeddings of tile k and the weights a_k are computed via an attention mechanism:

$$a_k = \frac{\exp\left\{w^\top \tanh\left(V h_k\right)\right\}}{\sum_{j=1}^{K} \exp\left\{w^\top \tanh\left(V h_j\right)\right\}}.$$

To enhance expressiveness, we also integrate a gating mechanism:

$$a_k = \frac{\exp\left\{w^\top \left[\tanh(V h_k) \odot \sigma(U h_k)\right]\right\}}{\sum_{j=1}^{K} \exp\left\{w^\top \left[\tanh(V h_j) \odot \sigma(U h_j)\right]\right\}},$$

where U is an additional learnable parameter, $\sigma(\cdot)$ is the sigmoid function, and $\odot$ denotes elementwise multiplication.

2.3 Domain Adaptation

A critical challenge in our task is the heterogeneity among data from different centers. To address this, we integrate domain adaptation via a Domain Adversarial Neural Network (DANN). A gradient reversal layer (GRL) is applied to the bag-level features before passing them to a domain classifier. This adversarial training forces the learned features to be invariant to the center (domain) information, thereby mitigating overfitting due to data disparity.

2.4 Ensemble Learning and Focal Loss

To further boost performance, we employ an ensemble of models trained with different random seeds. At inference, predictions from each model are averaged to yield the final output. Moreover, to focus on hard examples and address class imbalance, we use Focal Loss for the main classification task:

$$\mathrm{FL}(p_t) = -\alpha(1 - p_t)^\gamma \log(p_t),$$

where p_t is the predicted probability corresponding to the true label, γ is the focusing parameter, and α is the balancing weight.

3 Experiments and Results

Our method was evaluated on the Owkin Data Challenge dataset. Each whole-slide image is decomposed into 1000 tiles, with features extracted using a pre-trained Wide ResNet-50-2 (MoCo v2 features). Our training strategy includes:

- Using AdamW optimizer with a low learning rate (5×10^{-5}) and weight decay (1×10^{-4}).
- Early stopping with a patience of 5 epochs based on validation AUC.
- A ReduceLROnPlateau scheduler that decreases the learning rate on plateauing AUC.
- An ensemble of three models with different random seeds.

Our baseline logistic regression approach achieved an AUC of 0.60, while our proposed method reached an AUC of 0.73 on the public test set. We further discuss strategies (such as increasing attention dimensions, additional domain adaptation levels, and hyperparameter tuning) to boost the AUC towards 0.80.

Figure **??** shows the evolution of AUC during the training process.

4 Conclusion

We have presented a weakly supervised MIL approach for detecting the **PIK3CA** mutation directly from histopathological slides. By integrating an attention-based MIL pooling mechanism with domain adaptation techniques, and leveraging ensemble learning along with focal loss, our method significantly improves upon baseline performance. Our results indicate that further improvements—such as multi-level domain adaptation and advanced regularization strategies—could potentially boost the AUC to 0.80 or higher, thereby offering a more reliable tool for clinical decision-making.

References

[1] Dietterich, T. G., Lathrop, R. H., & Lozano-Pérez, T. (1997). *Solving the multiple instance problem with axis-parallel rectangles.* Artificial Intelligence, 89(1-2), 31–71.

[2] Maron, O., & Lozano-Pérez, T. (1998). *A framework for multiple-instance learning.* Advances in Neural Information Processing Systems, 570–576.

[3] Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). *Learning and transferring mid-level image representations using convolutional neural networks.* In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1717–1724.

[4] Lin, T.-Y., et al. (2017). *Focal Loss for Dense Object Detection.* Proceedings of the IEEE International Conference on Computer Vision, 2980–2988.

[5] Dauphin, Y. N., et al. (2016). *Language Modeling with Gated Convolutional Networks.* In International Conference on Machine Learning, 933–941.

[6] Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). *PointNet: Deep learning on point sets for 3D classification and segmentation.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 652–660.