Bridging Domains in Melanoma Diagnostics: Predicting BRAF Mutations and Sentinel Lymph Node Positivity with Attention-Based Models in Histological Images

Carlos Hernández-Pérez, Lauren Jimenez-Martin, Veronica Vilaplana
Image Processing Group - Universitat Politècnica de Catalunya (UPC)
Carrer de Jordi Girona 31, 08034, Barcelona, Spain
{carlos.hernandez.p, lauren.jimenez, veronica.vilaplana}@upc.edu

Abstract

Whole Slide Images (WSIs) have significantly advanced the field of pathology by providing highly detailed views of tissue samples. Integrating Deep Learning (DL) into this area of research, particularly through transformer-based foundational models, has marked a new era in automated image analysis. These foundational models are adept at extracting features from WSIs, an essential step in their analysis process. The subsequent application of weakly supervised learning techniques combines these features to predict critical biomarkers, such as BRAF mutations and sentinel lymph node (SLN) biopsy positivity, which are vital in guiding patient treatment strategies. However, the limited availability of labelled datasets in pathology hinders the usefulness of DL models. Domain adaptation strategies adeptly overcome this hurdle, enabling model knowledge transfer between different tissue types, thus addressing data scarcity. Our study employs a form of domain adaptation by fine-tuning two DINOv2 models, one pre-trained on natural images and the other on WSI of colorectal cancer from the TCGA dataset, adapting them for melanoma analysis. We also incorporate a comparison with features extracted by a third DINOv1 model trained solely on WSIs of breast cancer. With this approach, we find some notable success in detecting BRAF mutations. Nonetheless, predicting SLN positivity presents a more intricate challenge, largely due to the indirect correlation between local histopathological features in WSIs of primary tumours and lymph node metastasis manifestation. This dual-faceted approach not only combats the issue of limited data but also showcases the potential for enhanced accuracy in the field of digital pathology.

1. Introduction

Early detection of biomarkers is a critical component of effective cancer management, significantly impacting both the prognosis and the quality of life of skin cancer patients [12, 27, 34]. Melanoma is the most aggressive form of skin cancer, with a high rate of metastasis and mortality if not detected and treated early. It is the fourth leading cause of cancer-related mortality worldwide [24]. Timely identification of this condition allows for tailored therapeutic approaches, which can considerably enhance the prospects of patient survival and quality of life [13].

Advancements in skin cancer research have significantly benefited from the identification of biomarkers associated with the disease's development. Among them, mutations in the BRAF gene are of particular importance. Involved in cell growth and division, alterations in this gene play a critical role in melanoma progression [6, 8]. The detection of BRAF mutations offers a pathway to more personalized healthcare strategies, enabling clinicians to align surveillance and treatment more closely with individual genetic profiles. This approach opens avenues for integrating targeted therapies that specifically counteract the oncogenic effects of these mutations [6], thereby enhancing the efficacy of skin cancer treatment.

Furthermore, the sentinel lymph node (SLN) biopsy has become a standard procedure in melanoma management [32]. This procedure involves removing a small group of lymph nodes from the armpit or groin to check for cancer cells. The SLN is considered a sentinel node because it is the first node to which cancer cells would likely spread if they were present in the body [3]. A positive SLN biopsy, indicating the presence of cancer cells, guides the treatment plan, including the extent of surgery and the type of radiation therapy to be administered [3, 13].

The gold standard in pathology is the analysis of Whole Slide Images (WSI). These gigapixel images are revolutionizing pathology by providing high-resolution digital envi-

ronments for analysis. Nonetheless, the processing of WSI encounters significant challenges, notably the variability in staining protocols, imaging equipment, and tissue preparation techniques, which can differ widely within and between hospitals and images. This variability, coupled with the vast scale of the images, demands robust computational solutions where the interplay of medical knowledge and technological innovation becomes most evident.

Self-supervised learning (SSL) has risen as a tool to obtain robust feature extractors by leveraging the intrinsic patterns and structures present in unlabeled data to learn meaningful representations [7, 17] in the presence of heterogeneous data. Researchers [10, 21] have utilized self-supervised contrastive learning [7] to improve the extracted representations but the field is moving towards transformer-based architectures [22, 30, 31]. In this context, Vision Transformers (ViT) [11] have been increasingly adopted, with the DINO framework [5, 20] representing the forefront of this transition to attention-based systems. These ViT systems produce better results and provide an additional layer of explainability.

A way to profit from data from different tissue types is the use of domain adaptation. This technique is a particular and popular type of transfer learning that facilitates the transfer of knowledge from a source domain, normally abundant in labelled data, to a target domain where labels are sparse or partially available [15, 29]. It plays a crucial role in the field of histopathology due to the aforementioned variability and scarcity of labelled data found in WSIs. Adapting models from one tissue type to another can help in capturing subtle, domain-specific features that might be critical for accurate disease diagnosis and prognosis. By using domain adaptation, the same model can be used across multiple datasets, reducing the need for large, annotated datasets specific to each new domain.

Given that WSIs exceed the processing capacities of most systems when taken as a whole, they are typically divided into numerous smaller segments or patches. A significant challenge arises when these patches, which may number in the hundreds or thousands per WSI, must be analyzed under a single slide-level label. Multiple Instance Learning (MIL), a variant of weakly supervised learning, has been increasingly adopted to navigate this issue in WSI analysis [18, 23, 33, 35]. Within this framework, each WSI is conceptualized as a 'bag', with its constituent patches viewed as 'instances.' This paradigm allows for slide-wide classification based on the collective feature representation of patches, obviating the need for individual patch labels. This approach enables the model to identify distinctive patterns within the aggregated patch representations. Recent years have seen a surge in the application of attention mechanisms within MIL [16] to enrich feature aggregation from WSIs, concurrently offering interpretability by spotlighting the contributory weight of each patch [9, 19, 25]. While efforts like those in [28] harness advanced attention techniques to heighten model discrimination, methodologies such as ACMIL [35] explore regularization strategies aimed at preventing overfitting, particularly in scenarios of limited training data.

In this paper, we propose a framework that combines SSL, MIL, and transformer-based models to predict BRAF and SLN positivity. We study the influence of different domain-trained feature extractors as well as an array of MIL methodologies, aiming to effectively utilize the limited labelled melanoma WSI data available for public use.

2. Materials and Methods

In this study, we present an integrated computational framework designed to detect BRAF mutations and predict SLN positivity. Our methodology leverages finetuned ViTs for feature extraction and various weakly-supervised classification strategies.

This section describes the data used and the steps involved in image preprocessing and feature extraction. With regard to the classification strategies, we discuss different methodologies such as specific bag aggregators, ACMIL, as well as our new proposal ACTrans.

2.1. Dataset

The TCGA Skin Cutaneous Melanoma (TCGA-SKCM) dataset is part of the PanCancer Atlas initiative, which was designed to address broad, overarching questions about cancer. This dataset includes 475 slide-labeled WSI for samples of skin cutaneous melanomas from primary tumours. The distribution of positive and negative labels for the BRAF and SLN positivity can be seen in Table 1. All the data used in this study is publicly accessible through the TCGA data portal[1].

Marker	Positive Cases	Negative Cases
BRAF Positivity	250	225
SLN Positivity	223	252

Table 1. Distribution of WSI in the dataset according to BRAF and SLN positivity.

In this work, the labels for BRAF and SLN positivity are defined based on established clinical and pathological criteria. A positive BRAF case is identified through the presence of V600 mutations in the BRAF gene [6], which are indicative of alterations that may influence cell growth, and signaling pathways commonly associated with melanoma. Conversely, a negative BRAF case lacks these mutations. SLN positivity is determined by the presence of melanoma

[1] https://portal.gdc.cancer.gov

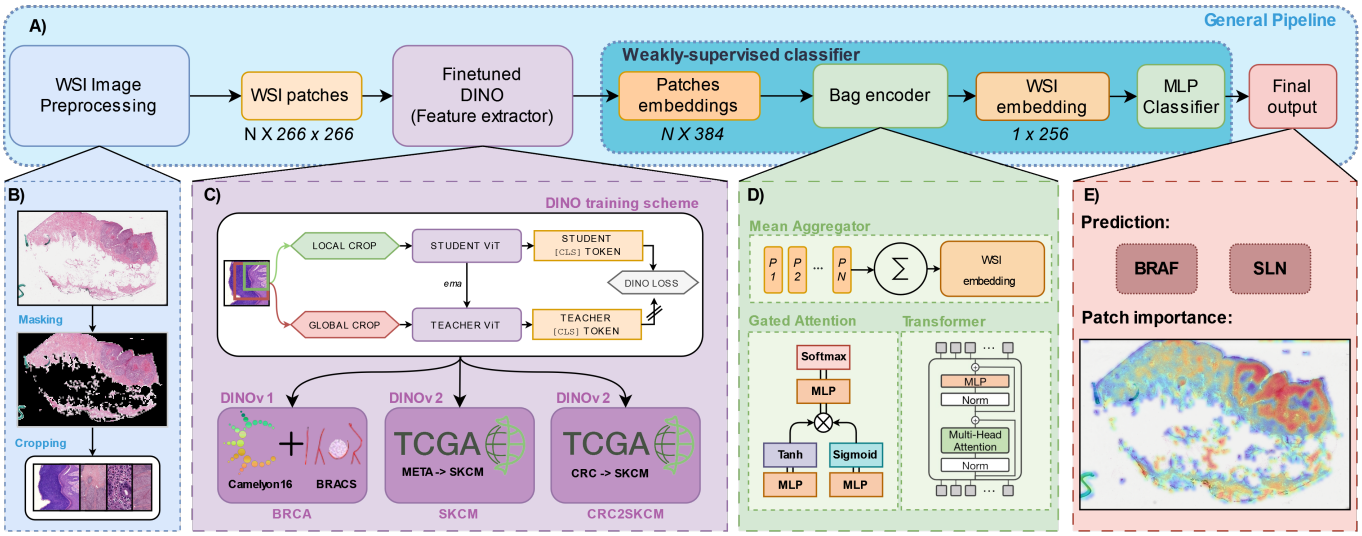

Figure 1. Overview of the computational pipeline for Whole Slide Image analysis. Panel A) outlines the whole process from WSI preprocessing to model output. Panel B) exemplifies the preprocessing of histopathological images. Panel C) details the knowledge distillation approach between a student and teacher Vision Transformer (ViT) on datasets from TCGA Melanoma, TCGA Colorectal, and the combination of CAMELYON16, and BRACS. Panel D) depicts the architecture of the MIL bag encoder that integrates patch embeddings into a global WSI embedding. Panel E) shows the prediction task for BRAF mutation and sentinel lymph node (SLN) status along with a heatmap of patch importance for interoperability for the BRAF positivity.

cells in the sentinel lymph node biopsy [32], which is a prognostic indicator of the likelihood of lymphatic spread. A negative SLN case shows no evidence of melanoma cells in the sentinel lymph nodes.

2.2. WSI Preprocessing

Our data pre-processing follows the method detailed in [19], involving a series of steps to separate tissue from background in WSIs. Initially, we downsample the images for ease of processing and convert them from RGB to HSV colour space. To create smoother edges, median blurring is applied, followed by binarization using a threshold on the saturation channel. We use morphological closing to fill in small spaces eliminating holes, and then filter the contours of foreground objects based on their area. Finally, patches are extracted from the identified tissue areas.

2.3. Feature Extractors

As SSL has been proven to be state-of-the-art for extracting robust features within the field of histopathological image analysis [21, 22, 28, 31], our study employs the DINO [5] and DINOv2 [20] architectures as primary feature extractors. In the DINO model, there is a 'student' ViT trained on smaller, localized image crops, and a 'teacher' network that utilizes bigger image crops. The student model's objective is to mimic the teacher's output, with a focus on matching the `[CLS]` tokens, which encapsulate global image information. Contrary to most distillation methodologies [14], in DINO, the teacher's parameters are also updated through an exponential moving average of the student's parameters.

We use three distinct feature extractors named BRCA, SKCM, and CRC2SKCM, each optimized through dataset-specific finetuning of the DINO ViT foundational model. Initially, we employed the Breast Cancer (BRCA) feature extractor proposed in [35], a DINOv1 model adapted to the combined CAMELYON16 [1] and BRACS [2] datasets. The second feature extractor was obtained by finetuning the base DINOv2 model [20] with the TCGA-SKCM dataset to derive the SKCM feature extractor. Finally, we also finetuned the model proposed by [22], another DINOv2 model which was pre-trained on TCGA's Colorectal Cancer (CRC) cohort. The resulting feature extractor was called CRC2SKCM. These procedures are illustrated at the bottom of panel C) in Figure 1.

2.4. Weakly-Supervised Classifier

To be able to process the WSI, the patches need to be encoded to reduce their dimensionality. Using a feature extractor, each WSI's patch is independently processed to acquire its respective embedding. After obtaining the embeddings, a bag encoder is applied followed by a classifier.

Bag Encoder. Within the MIL framework, the embeddings from each patch are treated as instances of the bags of a WSI. By adopting this strategy, the bag encoder applies an aggregation technique to integrate these discrete patch embeddings into a WSI embedding. The objective is to encapsulate patch-specific information into a single vector. Let N

represent the number of tokens per bag, $\mathbf{h}_i$ the embedding for the i^{th} patch computed by the patch encoder, and $\mathbf{z}$ the WSI embedding. The following aggregation schemes are considered:

- **Mean Aggregation:** The simplest approach, where the final representation of the WSI is computed as the arithmetic mean of the computed representations across all patches: $\mathbf{z}_{\text{Mean}} = \frac{1}{N}\sum_i^N \mathbf{h}_i$. This strategy assumes equal contribution from each patch towards the final prediction.
- **Gated Attention Mechanism:** this method employs a Gated Attention mechanism based on [16], to dynamically weigh the importance of each patch embedding based on its relevance to the outcome.

$$\mathbf{z}_{\text{GA}} = \sum_{i=1}^{N} a_i \mathbf{h_i}; \tag{1}$$
$$a_i = \text{Softmax}\left\{\mathbf{w}^\top \left(\tanh(\mathbf{V}\mathbf{h_i}^\top) \circ \text{sigm}(\mathbf{U}\mathbf{h_i}^\top)\right)\right\}$$

Where a_i is the attention weight for the *i-th* patch embedding, and $\mathbf{w} \in \mathbb{R}^{1\times L}$ and $\mathbf{V}$, $\mathbf{U} \in \mathbb{R}^{L\times M}$ are learnable parameters.

- **Transformer-based Encoding:** Based on [28], this approach leverages an additional Transformer encoder to process the collection of patches embeddings. Like the initial patch encoder phase, this encoder generates a new `[CLS]` token that encapsulates the collective information of the entire bag of patch embeddings. By doing so, it enables a deeper contextual analysis of the patch-level features in relation to each other.

$$\begin{aligned}
\mathbf{z}_{\text{T}} &= \mathbf{x}^{(L)}_{\texttt{[CLS]}};\\
\mathbf{x}^{(l)} &= \text{LN}(\mathbf{y}^{(l)}) + \mathbf{x}^{(l-1)};\\
\mathbf{y}^{(l)} &= \text{FF}(\text{LN}(\mathbf{m}^{(l)})) + \mathbf{m}^{(l)};\\
\mathbf{m}^{(l)} &= \text{MHSA}(\mathbf{x}^{(l-1)}) + \mathbf{x}^{(l-1)};
\end{aligned} \tag{2}$$

Where $l \in \{1, \ldots, L\}$ and $\mathbf{x}^{(0)} = \{\mathbf{h}_{\texttt{[CLS]}}, \mathbf{h}_1, \ldots, \mathbf{h}_N\}$, L is the total number of Transformer layers, LN represents Layer Normalization, FF represents the application of a Feed-Forward network and MHSA represents the application of the Multi-Head Self-Attention mechanism [26].

MLP Classifier. After applying an aggregation strategy, a Multi-Layer Perceptron (MLP) takes the WSI embedding and outputs the predicted diagnostic label.

2.4.1 ACMIL

The Attention Challenging MIL (ACMIL) methodology was introduced in [35] as a proposal to address overfitting in scenarios with limited data availability such as for WSI analysis. ACMIL's weakly-supervised classifier integrates two techniques: Multiple Branch Attention (MBA) to capture more discriminative instances, and Stochastic Top-K Instance Masking (STKIM) to consider more instances beyond those with the top-K saliency.

- **MBA:** The MBA initially identifies M patterns and subsequently aggregates their embeddings to formulate predictions. Each pattern is identified through a gated attention branch using Equation 1. To preserve the discriminative quality of the patterns while ensuring semantic diversity among them, two regularization techniques were introduced: semantic regularization and diversity regularization. The semantic regularization, aimed at capturing distinctive patterns, is implemented by adding an MLP layer behind each pattern embedding, with a cross-entropy loss function:

$$L_{\text{p}} = -\frac{1}{M}\sum_{i=1}^{M} \mathbf{Y}\log(\hat{\mathbf{Y}}_i) + (1-\mathbf{Y})\log(1-\hat{\mathbf{Y}}_i) \tag{3}$$

where $\hat{\mathbf{Y}}_i = g_i(z_i)$ is the prediction based on i-th pattern embedding z_i and $\mathbf{Y}$ is the bag label.
To avoid learning similar patterns and get more discriminative information, a diversity loss is introduced as follows:

$$L_d = \frac{2}{M(M-1)\sum_{i=1}^{M}\sum_{j=i+1}^{M}} cos(\mathbf{a}_i, \mathbf{a}_j) \tag{4}$$

where $\mathbf{a}_i$ compounds all the attention values of the i-th pattern, $\mathbf{a}_i = \{a_{i1}, ...a_{iN}\}$. The $cos(.)$ function measures the similarity in attention across different branches.
Once the patterns are obtained, they are aggregated as $\mathbf{a} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{a}_i$ where $\mathbf{a}$ is the aggregated attention of the whole bag, with dimension N. The bag embedding is obtained by aggregating the instance features using averaged attention $\mathbf{a}$.

- **STKIM.** This method incorporates a masking operation into the attention mechanism, positioned before feature aggregation and following attention value generation. It stochastically masks out a portion of instances with the highest attention values (top-K) with a probability p and redistributes their attention values among the remaining instances.

The ACMIL model is trained by minimizing a combined loss function, which includes the sums of L_d, L_a, and a cross-entropy loss calculated between the model's slide-level predictions and the corresponding labels

2.4.2 ACTrans

We made a variation to the ACMIL method explained in Section 2.4.1 where the features obtained from MBA are aggregated using the transformer aggregator described in Section 2.4. We call this new approach Attention Challenging Transformer (ACTrans).

2.5. Patch Importance Visualization

The attention-based aggregators identify which patches of the WSI were more relevant for the models' prediction. The relative importance of the extracted patches is quantified by converting the attention scores associated with the model's predicted class into percentiles. These scores are derived from the attention and transformer bag aggregators (Section D of Figure 1). The quantified scores are then mapped to their respective spatial locations within the WSI. This process yields a heatmap, which is superimposed on the WSI to illustrate the distribution of patch importance across the image. This method is not available for the mean bag aggregator.

3. Experiments and Results

Several experiments were performed to compare the influence of the different feature extractors and aggregators for BRAF and SLN classification tasks. The implementation details and experiment design used are the same for each task.

Every feature extractor (Section 2.3) is combined with each presented weakly-supervised classifier following these steps: 1) the WSIs are processed and tessellated into patches; 2) patch features are extracted using every feature extractor; 3) these extracted features serve as the input for each weakly-supervised classifier.

We combine every feature extractor (Section 2.3) with each presented weakly-supervised classifier (Section 2.4). We followed the next steps (Figure 1): 1) we processed the WSIs and tessellate it into patches; 2) we extracted patch features using every feature extractor; 3) these extracted features then served as the input for each weakly-supervised classifier.

3.1. Implementation Details

The DINOv2 foundational model was utilized to obtain the SKCM and CRC2SKCM feature extractors introduced in Section 2.3. WSIs were cropped into patches of 518x518 pixels at 20× magnification, resulting in a dataset exceeding 20 million patches. These patches were subsequently resized and random cropped to 224 by 224 pixels, ensuring compatibility with the input dimensionality of the Vision Transformer model. The training was performed with a batch size of 84 throughout 100,000 iterations.

The dataset was randomly split 10 times into training, validation, and test sets with a ratio of 80:10:10. These splits were performed in a stratified way to maintain the balance of labels throughout the splits. We preprocess the WSIs and segmented tissue regions as explained in Section 2.2. From these regions, 256x256 patches were extracted at 10X magnification, as shown in Panel B) of Figure 1.

All experiments were conducted using a learning rate of 0.0001 and weight decay of 0.005. For the ACMIL and ACTrans models, the number of branches, mask probability, and number of top-K values were set as proposed in [35] (5, 0.6, and 10 respectively).

3.2. Results

The experimental outcomes, detailed in Tables 2 and 3, highlight the different performances of bag aggregators across the BRAC, SKCM, and CRC2SKCM feature sets. We employ the following classification metrics: F1 score, Precision, Recall, ROC AUC, and Balanced Accuracy.

When predicting BRAF positivity, the models generally exhibit improved performance with the SKCM and CRC2SKCM feature sets, demonstrating the efficacy of domain adaptation in the context of skin cancer. However, a performance decrement was observed for ACMIL and ACTrans with SKCM, underscoring the variability in response to different feature sets. When adapting domains from colorectal to skin cancer, models with self-attention mechanisms outperformed others, indicating the potential benefit of these architectures in complex image-based diagnostics. Although ACMIL-based models display higher recall compared to the rest of the weakly-supervised classifiers, they also show a relatively lower precision.

The SLN positivity predictions in Table 3 present a different trend, simpler models like the mean aggregator exhibit superior performance compared to their more complex counterparts. This suggests that the association between image features and SLN labels may not be as strong, a finding consistent with the literature that indicates a challenging predictive task where models struggle to significantly outperform random chance [4, 21]. The modest performance improvement over random chance by these models underscores the need for ongoing research to identify more effective feature representations or alternative approaches that can more accurately capture the complexities associated with SLN positivity in melanoma diagnostics.

3.3. Visualization of Heatmaps

Figure 2 presents heatmaps illustrating the assigned patch-level importance given by the model for the task of BRAF mutation positivity. These were obtained using the ACMIL model for each of the feature extractors mentioned in Section 2.3. Patches that significantly influenced the model's prediction are highlighted in warmer hues (reds and oranges), indicating higher attention scores. Conversely, areas depicted in cooler tones (blues) correspond to patches with lesser influence on the predictive decision.

The BRCA feature extractor's heatmap reveals focused areas of attention, suggesting an emphasis on regions with stark colour contrasts. Conversely, the heatmaps for the other two feature extractors display a more distributed attention pattern, aligning with diverse image characteristics.

Aggregator	F1 score	Precision	Recall	ROC AUC	Balanced accuracy
BRAC Feature Extractor					
Mean	0.601 ± 0.077	0.614 ± 0.086	0.646 ± 0.104	**0.594 ± 0.096**	0.563 ± 0.097
GA	0.603 ± 0.076	**0.621 ± 0.089**	0.652 ± 0.106	0.593 ± 0.093	**0.568 ± 0.098**
Transformer	0.596 ± 0.075	0.607 ± 0.090	0.636 ± 0.104	0.582 ± 0.109	0.546 ± 0.101
ACMIL	0.605 ± 0.065	0.534 ± 0.061	0.718 ± 0.155	0.523 ± 0.072	0.523 ± 0.072
ACTrans	**0.633 ± 0.082**	0.563 ± 0.060	**0.773 ± 0.196**	0.555 ± 0.060	0.555 ± 0.060
SKCM Feature Extractor					
Mean	0.609 ± 0.072	0.630 ± 0.085	0.658 ± 0.103	**0.606 ± 0.098**	0.575 ± 0.093
GA	0.612 ± 0.074	**0.637 ± 0.083**	0.671 ± 0.115	0.600 ± 0.095	**0.584 ± 0.087**
Transformer	**0.616 ± 0.073**	0.635 ± 0.096	0.668 ± 0.120	0.599 ± 0.110	0.578 ± 0.104
ACMIL	0.600 ± 0.048	0.530 ± 0.037	**0.705 ± 0.130**	0.521 ± 0.042	0.521 ± 0.042
ACTrans	0.597 ± 0.078	0.538 ± 0.057	0.700 ± 0.185	0.529 ± 0.065	0.529 ± 0.065
CRC2SKCM Feature Extractor					
Mean	0.620 ± 0.061	0.634 ± 0.087	0.656 ± 0.102	0.611 ± 0.094	0.571 ± 0.102
GA	0.648 ± 0.063	**0.660 ± 0.095**	0.689 ± 0.126	0.621 ± 0.092	0.584 ± 0.090
Transformer	0.631 ± 0.069	0.652 ± 0.098	0.678 ± 0.135	**0.644 ± 0.099**	**0.591 ± 0.113**
ACMIL	**0.655 ± 0.046**	0.530 ± 0.029	**0.868 ± 0.129**	0.529 ± 0.042	0.529 ± 0.042
ACTrans	0.637 ± 0.073	0.545 ± 0.052	0.796 ± 0.194	0.543 ± 0.065	0.543 ± 0.065

Table 2. BRAF positivity classification results showing $\mu \pm \sigma$ for the 10 random splits, across all the MIL strategies using the three different feature sets.

Aggregator	F1 score	Precision	Recall	ROC AUC	Balanced accuracy
BRAC Feature Extractor					
Mean	**0.655 ± 0.068**	0.575 ± 0.070	**0.785 ± 0.143**	**0.592 ± 0.069**	0.566 ± 0.058
GA	0.625 ± 0.100	**0.579 ± 0.082**	0.732 ± 0.204	0.579 ± 0.083	0.568 ± 0.059
Transformer	0.633 ± 0.065	0.563 ± 0.072	0.750 ± 0.145	0.582 ± 0.078	0.546 ± 0.050
ACMIL	0.568 ± 0.050	0.533 ± 0.057	0.630 ± 0.142	0.573 ± 0.036	0.573 ± 0.036
ACTrans	0.582 ± 0.131	0.529 ± 0.112	0.699 ± 0.258	0.575 ± 0.107	**0.575 ± 0.107**
SKCM Feature Extractor					
Mean	**0.661 ± 0.070**	**0.580 ± 0.065**	**0.790 ± 0.138**	**0.591 ± 0.058**	0.572 ± 0.066
GA	0.633 ± 0.080	0.576 ± 0.069	0.740 ± 0.173	0.571 ± 0.080	0.564 ± 0.053
Transformer	0.628 ± 0.060	0.571 ± 0.064	0.724 ± 0.141	0.590 ± 0.077	0.555 ± 0.052
ACMIL	0.581 ± 0.096	0.510 ± 0.103	0.713 ± 0.173	0.551 ± 0.092	0.551 ± 0.092
ACTrans	0.601 ± 0.044	0.549 ± 0.093	0.702 ± 0.149	0.588 ± 0.064	**0.588 ± 0.064**
CRC2SKCM Feature Extractor					
Mean	**0.652 ± 0.090**	**0.586 ± 0.080**	**0.762 ± 0.164**	**0.599 ± 0.056**	**0.581 ± 0.060**
GA	0.623 ± 0.101	0.578 ± 0.080	0.735 ± 0.224	0.594 ± 0.090	0.569 ± 0.058
Transformer	0.612 ± 0.070	0.567 ± 0.057	0.697 ± 0.174	0.572 ± 0.068	0.550 ± 0.051
ACMIL	0.589 ± 0.096	0.534 ± 0.112	0.674 ± 0.113	0.572 ± 0.110	0.572 ± 0.110
ACTrans	0.598 ± 0.077	0.552 ± 0.132	0.715 ± 0.205	0.577 ± 0.089	0.577 ± 0.089

Table 3. SLN positivity classification results showing $\mu \pm \sigma$ for the 10 random splits, across all the MIL strategies using the three different feature sets.

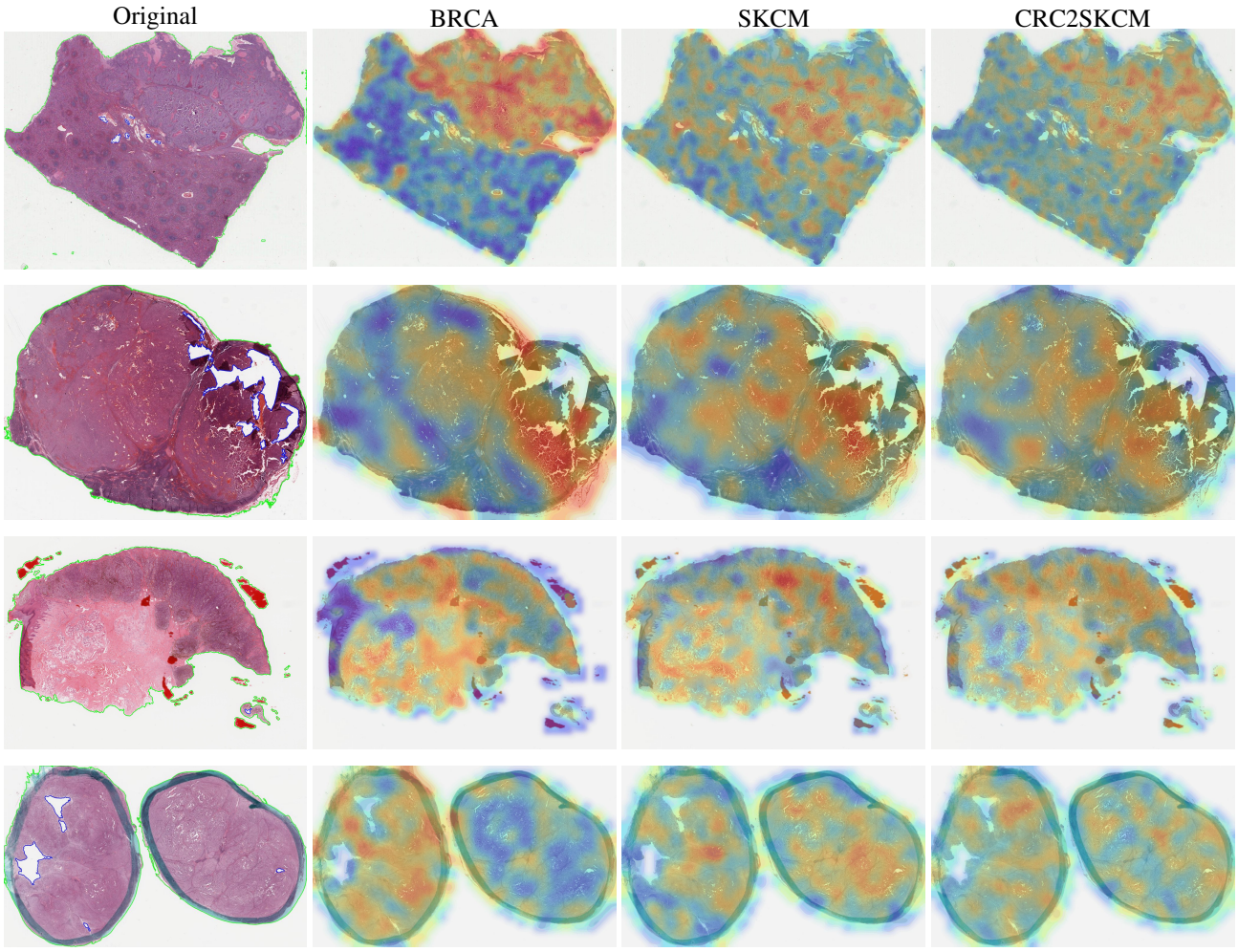

Figure 2. Comparative heatmap visualizations for melanoma WSI analysis. The first column shows the original histopathological images of skin cutaneous melanoma. The subsequent columns display the corresponding heatmaps generated using the ACMIL approach and three distinct feature extractors: BRCA, SKCM, and CRC2SKCM. These heatmaps represent the model's focus areas, with warmer colours indicating regions of higher relevance to the model's predictions and cooler colours representing areas of lesser importance.

4. Conclusion

In this paper, we have shown a methodological approach to harness the capabilities of domain-adapted transformer-based models to predict relevant melanoma biomarkers, and is enhanced through self-supervised and weakly supervised learning techniques. Our findings suggest that the application of domain adaptation holds promise, particularly for the prediction of BRAF mutation status from WSIs of melanoma patients. The domain adaptation of foundational models pre-trained on other WSI datasets yielded performance gains, underscoring the efficacy of domain adaptation in this context.

However, the predictive outcomes for SLN positivity were less conclusive, showing the inherent challenge of inferring metastatic spread to lymph nodes, a condition not directly depicted in WSIs. This underscores the current limitation of models that rely solely on local features from WSIs to make predictions about distal pathologies.

Despite these challenges, the advancements in foundational models for feature extractor fine-tuning and the ensuing a level of interpretability, as evidenced by our heatmap visualizations, mark a significant step towards more accurate diagnostic tools in the field of dermatopathology. Our work contributes to the expanding field of explainable AI in medicine, indicating a path for future research to build upon these findings and explore additional modalities that may augment the predictive capabilities for SLN positivity.

Acknowledgments

This research was supported by the Spanish Research Agency (AEI) under project PID2020-116907RB-I00 of the call MCIN/ AEI /10.13039/501100011033 and the project

718/C/2019 with id 201923-30 and 201923-31 funded by Fundació la Marato de TV3. Also by the FI-AGAUR (2022 FI_B 00634) grant funded by Direcció General de Recerca (DGR) of Departament de Recerca i Universitats (REU) of the Generalitat de Catalunya and the European Social Fund.

References

[1] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22):2199–2210, 2017. 3

[2] Nadia Brancati, Anna Maria Anniciello, Pushpak Pati, Daniel Riccio, Giosuè Scognamiglio, Guillaume Jaume, Giuseppe De Pietro, Maurizio Di Bonito, Antonio Foncubierta, Gerardo Botti, et al. Bracs: A dataset for breast carcinoma subtyping in h&e histology images. *Database*, 2022: baac093, 2022. 3

[3] Daciana Elena Brănişteanu, Mihai Cozmin, Elena Porumb-Andrese, Daniel Brănişteanu, Mihaela Paula Toader, Diana Iosep, Diana Sinigur, Cătălina Ioana Brănişteanu, George Brănişteanu, Vlad Porumb, et al. Sentinel lymph node biopsy in cutaneous melanoma, a clinical point of view. *Medicina*, 58(11):1589, 2022. 1

[4] Titus J Brinker, Lennard Kiehl, Max Schmitt, Tanja B Jutzi, Eva I Krieghoff-Henning, Dieter Krahl, Heinz Kutzner, Patrick Gholam, Sebastian Haferkamp, Joachim Klode, et al. Deep learning approach to predict sentinel lymph node status directly from routine histology of primary melanoma tumours. *European Journal of Cancer*, 154:227–234, 2021. 5

[5] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021. 2, 3

[6] Giorgia Castellani, Mariachiara Buccarelli, Maria Beatrice Arasi, Stefania Rossi, Maria Elena Pisanu, Maria Bellenghi, Carla Lintas, and Claudio Tabolacci. Braf mutations in melanoma: Biological aspects, therapeutic implications, and circulating biomarkers. *Cancers*, 15(16):4026, 2023. 1, 2

[7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. 2

[8] Liang Cheng, Antonio Lopez-Beltran, Francesco Massari, Gregory T MacLennan, and Rodolfo Montironi. Molecular testing for braf mutations to inform melanoma treatment decisions: a move toward precision medicine. *Modern Pathology*, 31(1):24–38, 2018. 1

[9] Philip Chikontwe, Meejeong Kim, Soo Jeong Nam, Heounjeong Go, and Sang Hyun Park. Multiple instance learning with center embeddings for histopathology classification. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23*, pages 519–528. Springer, 2020. 2

[10] Ozan Ciga, Tony Xu, and Anne Louise Martel. Self supervised contrastive learning for digital histopathology. *Machine Learning with Applications*, 7:100198, 2022. 2

[11] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 2

[12] Claus Garbe, Ulrike Keim, Teresa Amaral, Carola Berking, Thomas K Eigentler, Lukas Flatz, Anja Gesierich, Ulrike Leiter, Rudolf Stadler, Cord Sunderkötter, et al. Prognosis of patients with primary melanoma stage i and ii according to american joint committee on cancer version 8 validated in two independent cohorts: implications for adjuvant treatment. *Journal of Clinical Oncology*, 40(32):3741, 2022. 1

[13] Jeffrey E Gershenwald, Richard A Scolyer, Kenneth R Hess, Vernon K Sondak, Georgina V Long, Merrick I Ross, Alexander J Lazar, Mark B Faries, John M Kirkwood, Grant A McArthur, et al. Melanoma staging: evidence-based changes in the american joint committee on cancer eighth edition cancer staging manual. *CA: a cancer journal for clinicians*, 67(6):472–492, 2017. 1

[14] Jianping Gou, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6):1789–1819, 2021. 3

[15] Hao Guan and Mingxia Liu. Domain adaptation for medical image analysis: a survey. *IEEE Transactions on Biomedical Engineering*, 69(3):1173–1185, 2021. 2

[16] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136. PMLR, 2018. 2, 4

[17] Mingu Kang, Heon Song, Seonwook Park, Donggeun Yoo, and Sérgio Pereira. Benchmarking self-supervised learning on diverse pathology datasets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3344–3354, 2023. 2

[18] Narmin Ghaffari Laleh, Hannah Sophie Muti, Chiara Maria Lavinia Loeffler, Amelie Echle, Oliver Lester Saldanha, Faisal Mahmood, Ming Y Lu, Christian Trautwein, Rupert Langer, Bastian Dislich, et al. Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology. *Medical image analysis*, 79:102474, 2022. 2

[19] Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570, 2021. 2, 3

[20] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al.

Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 2, 3

[21] Carlos Hernandez Perez, Marc Combalia Escudero, Susana Puig, Josep Malvehy, and Veronica Vilaplana Besler. Contrastive and attention-based multiple instance learning for the prediction of sentinel lymph node status from histopathologies of primary melanoma tumours. In *MICCAI Workshop on Cancer Prevention through Early Detection*, pages 57–66. Springer, 2022. 2, 3, 5

[22] Benedikt Roth, Valentin Koch, Sophia J Wagner, Julia A Schnabel, Carsten Marr, and Tingying Peng. Low-resource finetuning of foundation models beats state-of-the-art in histopathology. *arXiv preprint arXiv:2401.04720*, 2024. 2, 3

[23] Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in neural information processing systems*, 34:2136–2147, 2021. 2

[24] Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. 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, 2021. 1

[25] Paul Tourniaire, Marius Ilie, Paul Hofman, Nicholas Ayache, and Herve Delingette. Attention-based multiple instance learning with mixed supervision on the camelyon16 dataset. In *MICCAI Workshop on Computational Pathology*, pages 216–226. PMLR, 2021. 2

[26] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 4

[27] Rachel K Voss, Tessa N Woods, Kate D Cromwell, Kelly C Nelson, and Janice N Cormier. Improving outcomes in patients with melanoma: strategies to ensure an early diagnosis. *Patient related outcome measures*, pages 229–242, 2015. 1

[28] Sophia J Wagner, Daniel Reisenbüchler, Nicholas P West, Jan Moritz Niehues, Jiefu Zhu, Sebastian Foersch, Gregory Patrick Veldhuizen, Philip Quirke, Heike I Grabsch, Piet A van den Brandt, et al. Transformer-based biomarker prediction from colorectal cancer histology: A large-scale multicentric study. *Cancer Cell*, 41(9):1650–1661, 2023. 2, 3, 4

[29] Pin Wang, Pufei Li, Yongming Li, Jin Xu, and Mingfeng Jiang. Classification of histopathological whole slide images based on multiple weighted semi-supervised domain adaptation. *Biomedical Signal Processing and Control*, 73:103400, 2022. 2

[30] Xiyue Wang, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Wei Yang, Junzhou Huang, and Xiao Han. Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical image analysis*, 81:102559, 2022. 2

[31] Frederik Wessels, Max Schmitt, Eva Krieghoff-Henning, Malin Nientiedt, Frank Waldbillig, Manuel Neuberger, Maximilian C Kriegmair, Karl-Friedrich Kowalewski, Thomas S Worst, Matthias Steeg, et al. A self-supervised vision transformer to predict survival from histopathology in renal cell carcinoma. *World Journal of Urology*, 41(8):2233–2241, 2023. 2, 3

[32] Sandra L Wong, Mark B Faries, Erin B Kennedy, Sanjiv S Agarwala, Timothy J Akhurst, Charlotte Ariyan, Charles M Balch, Barry S Berman, Alistair Cochran, Keith A Delman, et al. Sentinel lymph node biopsy and management of regional lymph nodes in melanoma: American society of clinical oncology and society of surgical oncology clinical practice guideline update. *Annals of surgical oncology*, 25:356–377, 2018. 1, 3

[33] Jiawen Yao, Xinliang Zhu, Jitendra Jonnagaddala, Nicholas Hawkins, and Junzhou Huang. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. *Medical Image Analysis*, 65: 101789, 2020. 2

[34] Jade N Young, Kelly Griffith-Bauer, Emma Hill, Emile Latour, Ravikant Samatham, and Sancy Leachman. The benefit of early-stage diagnosis: A registry-based survey evaluating the quality of life in patients with melanoma. *Skin Health and Disease*, 3(4):e237, 2023. 1

[35] Yunlong Zhang, Honglin Li, Yuxuan Sun, Sunyi Zheng, Chenglu Zhu, and Lin Yang. Attention-challenging multiple instance learning for whole slide image classification. *arXiv preprint arXiv:2311.07125*, 2023. 2, 3, 4, 5