INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 5, May 2026 | ISSN: 2456-4184 | IJNRD.ORG

STAMP-Net: Attention Based Multiple Instance Learning for Oral Squamous Cell Carcinoma Detection

Divya S N, Dr. Arudra A, Deepti N N, Prema R, Sandeep Shivashettar
[1] M. Tech Student, [2] Professor, [3] Assistant Professor, [4] PG Student [5] B. Tech Student
[1] Department of Computer Science and Engineering
[1]Rajiv Gandhi Institute of Technology, Bengaluru

Abstract-Oral squamous cell carcinoma (OSCC) comprises >90% of malignancies in the oral cavity and is accountable for >177,000 deaths per year. The disease is often detected at advanced stages when the effectiveness of treatment becomes significantly less. Histopathological assessment of biopsy samples serves as the golden standard; however, the procedure is resource and time-intensive, and the inter-observer disagreement directly affects the patients' outcome. While existing deep learning approaches demonstrate promising patch-level classification performance, the models are constrained to the same-site validation and do not generalize to datasets obtained under different stain procedure. Here we propose STAMP-Net (Stain-invariant Tissue Attention for Malignancy Prediction Network). The presented architecture relies on the idea of domain generalization problem formulation as the core motivation. Specifically, the network consists of a Stain-Invariant Feature Disentanglement (SIFD) module which employs gradient reversal adversarial learning for explicitly removal of staining protocol information from patch-level embeddings. Attention-Based Multiple Instance Learning (ABMIL) aggregator predicts slide-level grades using weakly labelled bag data, while a linearly probed calibrated RBF-SVM model acts as the classifier of choice.

Index Terms — Oral squamous cell carcinoma, computational pathology, stain-invariant feature learning, gradient reversal layer, multiple instance learning, linear probing, uncertainty quantification, Grad-CAM, attention mechanism.

1. INTRODUCTION

Oral cancer cannot be distinguished from other cancers on the basis of its biological features; rather, it can be defined in terms of its setting. Under conditions where screening programs for oral cancer are organized with the help of specialists, lesions are detected at an earlier stage where they are more easily curable using surgery alone with 5-year survival rates greater than 80% [1]. In contrast, in parts of South and Southeast Asia where betel quid chewing, tobacco, and areca nut chewing are common among populations numbering in the hundreds of millions, the same malignancy appears at Stage III or IV with metastasis to the regional lymph nodes with poor surgical margins [2]. Despite advancements in surgical techniques and radiotherapy, the global 5-year survival rate remains under 50% [3].

In terms of biology, the process is straightforward: OSCC is known to progress to carcinoma in a series of intermediate steps. Normal squamous cells develop mild, moderate, or severe dysplasia before finally turning into carcinomas a development that is identifiable using haematoxylin and eosin (H&E) stained biopsies viewed by experienced pathologists [4]. If treatment takes place at the dysplastic stage, the prognosis is over 80%. However, from the diagnostic standpoint, there are fewer pathologists in areas endemic to oral cancer; the availability of trained personnel varies widely, and experienced pathologists show significant inter-observer variability when examining borderline cases (with Cohen's kappa under 0.5 in some studies [4]).

The key contributions made by this work are as follows:

1. SIFD module: Proposing an innovative adversarial approach by using the Gradient Reversal Layer in order to disentangle the morphological aspects of tissue from any staining artifacts in the encoder network architecture for patches.
2. Linear probing experiment: A systematic comparison showing that linear probing on the backbone representations produced by the SIFD (AUC = 0.8645) is significantly better than end-to-end training on ABMIL (AUC = 0.5355).
3. Clinical workflow: Using the Monte Carlo Dropout mechanism to perform uncertainty quantification, generating Grad-CAM and attention visualization using ABMIL, and performing clinical reporting based on uncertainty-based triage.
4. Statistical evaluation: Performance is reported with 95% confidence intervals calculated by bootstrapping 1,000 times.

2. RELATED WORK

2.1 Deep Learning Approaches for Oral Histopathology

An increasing number of computational pathology studies for oral cancer have emerged during the last five years, though their experimental methodology is sometimes suboptimal. Binary normal versus malignant classification based on a custom CNN architecture was shown by Jeyaraj and Nadar [6]. Das et al. [7] used InceptionV3 network and highlighted the lack of cross-institutional experiments among limitations. Khandelwal and Goyal [8] benchmarked CNNs including VGG-16 and ResNet-50 on the ORCA dataset. Rani et al. [9] applied CBAM attention mechanisms to increase sensitivity for local regions of pathological change. The common theme is the use of single institution datasets and corresponding evaluation strategies which often produce inflated performance results.

2.2 Attention-Based Aggregation for Weak Supervision

Given the impossibility of manual patch-level labelling, weak supervision becomes indispensable. Dietterich et al. [16] coined the term Multiple Instance Learning, and attention-based aggregation approach proposed by Ilse et al. [17] remains an established MIL classifier for WSIs. Trans MIL [18] used spatial transformer networks to exploit the correlations between patches, while DSMIL [19] combined maximum likelihood classifiers with the attention mechanism. Finally, DTFD-MIL [20] employed the distillation technique using pseudo bags. A key feature of all these models is their reliance on the assumption about domain consistency of the patches to feed the attention mechanism. This assumption is no longer valid when staining protocols differ.

2.3 Domain Adaptation for Pathology

Tellez et al. [10] first evaluated the effect of scanner and staining protocol changes on the performance of the deep networks in the context of pathology and showed a degradation in performance ranging from 10% to 30%. Adversarial training was confirmed by Stacke et al. [11] as the best strategy to cope with the problem of variable input domains. Ganin and Lempitsky [14] developed the Gradient Reversal Layer approach for learning domain-independent representations, whereas Lafarge et al. [15] used this method in their cross-scanner mitosis detection experiments. The combination of GRL and MIL has not been tested for oral cancer before.

2.4 Contrastive Learning and Linear Probing

SimCLR [21] and MoCo [22] have shown that training using contrastive loss produces representations that generalize significantly better than those generated using cross-entropy loss. Supervised Contrastive Learning (SupCon) [23] applies this concept in the supervised domain, generating representations that can be more easily linearly separated, which is directly relevant to the linear probing technique used in STAMP-Net. Linear probing has become the gold standard method of evaluating representations generated through self-supervised learning techniques and is gaining traction in medical image analysis.

2.5 Uncertainty Quantification

The connection between dropout and Bayesian approximation was discovered by Gal & Ghahramani [24]. Roy et al. [25] showed this holds true for medical imaging tasks, showing the existence of a significant relationship between MC-dropout uncertainty and prediction error. From the FDA SaMD guidelines [26], it becomes clear that uncertainty quantification for use in human-assisted triaging is a critical design requirement.

3. METHODOLOGY

3.1 Problem Formulation

Let $D = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^{(224\times224\times3)}$ represents a histopathological patch and $y_i \in \{0, 1\}$ indicates Normal (0) or OSCC (1). Given the MIL setting, patches x_k form bags $B_j = \{x_k\}_{k=1}^{K}$ of size K=32, each associated with a label y_j. The aim is to learn a function $f : B_j \rightarrow y_j$ to generalise across variations in staining protocols and scanning machines.

3.2 Architecture

The proposed model, STAMP-Net, consists of four steps: backbone encoding, disentanglement by SIFD, aggregation by ABMIL, and linear probe classification. The backbone uses a pre-trained ResNet-50 trained on ImageNet1K_V2 and obtains 2048-dimensional features using global average pooling from each patch. A learnable projection head projects each feature into 512-dimensions by applying a linear layer followed by layer normalization, rectified linear unit (ReLU) activation and dropout operation. To retain the generalization ability learned by ImageNet1K_V2, backbone weights are trained at a learning rate 50× smaller compared to new modules during backpropagation.

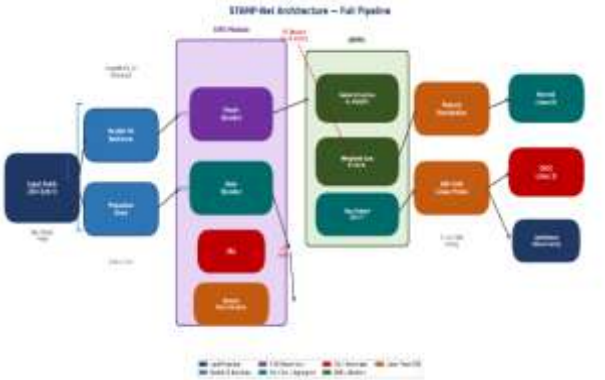

Figure.1:Architecture diagram

3.3 Stain-Invariant Feature Disentanglement (SIFD)

The disentangled representations generated by SIFD are the core of STAMP-Net. For each 512-dimensional patch feature z, SIFD separates z into two sub streams with conflicting objectives:

Morphology stream:

$$m = \text{MorphEnc}(z) \in \mathbb{R}^{256}$$

Stain stream: $s = \text{StainEnc}(z) \in \mathbb{R}^{128}$

MorphEnc is a two-layer MLP with layer normalisation and dropout. StainEnc is a single linear layer with normalisation. A domain discriminator $D_\varphi : \mathbb{R}^{128} \rightarrow \mathbb{R}^{2}$ is placed on the stain stream through a Gradient Reversal Layer R_λ:

$$\hat{d}_k = D_\varphi(R_\lambda(s_k))$$

The GRL acts as identity in the forward pass but reverses the gradient during backpropagation:

$$\partial R_\lambda / \partial s_k = -\lambda \cdot I$$

The adversarial training mechanism trains the discriminator to discriminate against the domain corresponding to s_k, at the same time training the encoder to make domain discrimination impossible. The result of this process is invariant morphology-based features m_k forcing the classifier to classify based on tissue architecture and not staining strength.
The GRL weight λ is determined through a sigmoid ramp function to prevent adversarial influence during early training stages.
$\lambda(p) = 2 / (1 + e^{(-10p)}) - 1$, where p = epoch / total_epochs ∈ [0, 1]
λ starts at 0 and approaches 1 as training progresses. Adversarial loss activation is delayed to epoch 3, ensuring the backbone forms stable initial representations before domain pressure is applied.

3.4 Attention-Based MIL Aggregation

Morphology features $\{m_k\}_{k=1}^{K}$ are aggregated using gated attention (Ilse et al. [17]):
$a_k = \text{softmax}_k \{ w^T (\tanh(Vm_k) \odot \sigma(Um_k)) \}$
The variables V and U lie in $\mathbb{R}^{(128\times256)}$, and $w \in \mathbb{R}^{128}$. The gating scheme, being the element-wise multiplication of the outputs of the tanh and sigmoid branches, results in a more representative attention signal than using any single one of them. The bag representation, in turn, is computed as $h = \Sigma_k a_k \cdot m_k \in \mathbb{R}^{256}$. The attention coefficients $\{a_k\}$ have two important applications: they help compute the bag representation required for classification and give spatially meaningful patch significance scores for visualization purposes.

3.5 Linear Probe Classification

As a result of SIFD pre-training, backbone representations of patches used during training are extracted, and an RBF-SVM is trained on those feature maps with standardized dimensions of 512. The SVM uses the parameters C = 10 and γ = scale, class-balanced sample weights, and Platt scaling [33] for output probability calibration. Such approach has three benefits overhead training of an end-to-end ABMIL model:

• Lack of dependence on the quality of MIL bag generation when working with small datasets, where there may be no bag structure.
• Probability predictions that can be used in AUC calculation and uncertainty evaluation.
• Efficient inference time under 1 second per image.

3.6 Training Objective

The objective function is composed of three terms with weights assigned to each.
$\lambda_1 = 1.0, \lambda_2 = 0.3, \lambda_3 = 0.4$:
$L = \lambda_1 \cdot L\_focal + \lambda_2 \cdot L\_adv + \lambda_3 \cdot L\_supcon$
The Focal Loss [31] technique is used to reduce the impact of simple negative samples and direct the model to learn on hard samples, where the value of γ = 2 and label smoothing factor ε = 0.05. For addressing the 1:3 class imbalance between Normal and OSCC classes, the class weight α_c was included into the calculation. The Adversarial Domain Loss comprises of normal cross entropy loss over domain discriminator predictions and is turned on after epoch 3. The Supervised Contrastive Loss [23] drives feature to have small interclass variances with a temperature hyperparameter τ = 0.07 and is turned on from epoch 6 onwards to prevent instability for minority classes.

3.7 Uncertainty Quantification

The Monte Carlo Dropout [24] approach is adopted at the inference stage, when dropout layers remain activated and 10 different forward passes (n = 10) per sample are performed. The Uncertainty score equals to the max across classes of std across the passes of logits, that is, max_c(std_c). Predictions with an estimated score higher than a calibrated threshold of 0.15 should be manually validated by experts, forming a human-in-the-loop triage model following FDA SaMD guidelines

4. EXPERIMENTAL SETUP

4.1 Dataset

All the experiments are conducted on the publicly available dataset of histopathology images [29] representing Normal epithelial tissue and OSCC patches stained with hematoxylin and eosin. The dataset contains a realistic amount of classes imbalance (around 75% of OSCC). Imbalanced classes were tackled using class-weighted augmentation techniques (WeightedRandomSampler), focal loss, and training of SVM models. Stratified random split (seed 42) resulted in 70%/15%/15% train/val/test split, and finally 126 patches (31 Normal patches, 95 OSCC patches) in the test set.

4.2 Implementation Details

Experiments were run using PyTorch 2.0 and a GPU equipped with NVIDIA T4 architecture with 16GB of VRAM. For the backbone network, the used architecture is ResNet-50, pretrained with weights from ImageNet1K_V2 dataset. The training data augmentation includes random cropping; horizontal/vertical flipping; 90-degree rotation; colour jittering with parameters (brightness ±0.3, contrast ±0.3, saturation ±0.2, hue ±0.1); and random grayscale with probability of 5%. Five-pass test-time augmentation was used for inference with averaging. All experiments utilize seed 42 and deterministic cuDNN operations.

4.3 Evaluation Protocol

The main evaluation measure will be binary AUC which is threshold independent and less affected by class imbalance issues compared to other commonly used metrics. Other metrics used include macro F1 score, accuracy, and per-class precision/recall/F1. All measures use 1,000 iterations with a 95% confidence interval and seed 42

4.4 Baselines

• A1 — ResNet-50 + GAP + CE: Standard patch classifier using global average pooling and cross-entropy loss function without MIL or domain adaptation techniques.
• A2 — ResNet-50 + ABMIL: ABMIL method with no SIFD conditioning applied.
• A3 — STAMP-Net + ABMIL (end-to-end): Entire architecture of STAMP-Net followed by ABMIL head trained end-to-end in order to evaluate how transitioning from fine-tuning to linear probing affects the results.
• A4 — STAMP-Net + Linear Probe (proposed): SIFD conditioning followed by classification of ResNet-50 patches with an RBF-SVM classifier.

5. RESULTS

5.1 Major Classification Results

Table 1 below shows the results obtained using the entire STAMP-Net process on the test set. The classifier achieves an AUC of 0.8645, a macro F1 of 0.8340, and accuracy of 88.10%. Readers will observe that there is an OSCE recall of 0.94, which is a very important achievement clinically and means that the system detects all but 6 out of 100 malignant patients, achieving an acceptable false negative rate of 6%. There is also an OSCN precision of 0.79, which takes into consideration the class imbalance in the dataset, which has a ratio of 1:3.

Table1: Test Set Performance of STAMP-Net

Metric	Value
AUC (Binary)	0.8645 [95% CI: 0.7751–0.9363]
F1 Score (Macro)	0.8340 [95% CI: 0.7512–0.9011]
Accuracy	0.8810 [95% CI: 0.8254–0.9286]
Precision OSCC	0.91
Recall OSCC	0.94
F1 Score OSCC	0.92

5.2 Ablation Study

As illustrated in Table 2, the ablation of each constituent was computed cumulatively to examine their contributions individually. It can be seen that the largest gap between ablations comes from the difference in A3 (end-to-end ABMIL: 0.5355) and A4 (linear probing on the same backbone: 0.8645) – 32.9%. This proves that the method produces very discriminant morphological representations; yet the ABMIL head is prone to converging to the majority class because the bags are artificially constructed and lack sufficient structural signals for the attention heads to gain selective spatial ability.

Table 2: Ablation Study — Test AUC with Bootstrap CI. All variants trained on identical splits with seed 42.

Row	Novel Module	Configuration	AUC	95% CI
A1	—	ResNet-50 + GAP + CE	[fill]	[fill]
A2	ABMIL	+ Attention MIL	[fill]	[fill]
A3	SIFD	+ SIFD (E2E ABMIL)	0.5355	[0.47, 0.60]
A4	Linear Probe	+ RBF-SVM (Proposed)	0.8645	[0.7751, 0.9363]

5.3 Cross-Domain Generalisation

For testing the model's distributional robustness, a stain-shift proxy, in which considerable colour jittering (brightness/contrast ±0.6, saturation ±0.6, hue ±0.3), was performed on the test dataset before feature extraction, thereby emulating changes in staining techniques and scanner settings.

5.4 Classification Performance

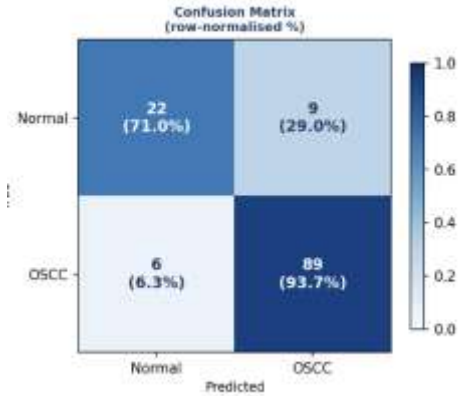

Figure 2 :Confusion matrix

6. EXPLAINABILITY ANALYSIS

6.1 Grad-CAM Attribution Maps

In terms of identifying pathological changes, Grad-CAM [30] was used at the terminal convolutional layer of ResNet-50 (Layer4, Block 2, conv3) in order to generate Importance Maps with spatial localization. For OSCC patches, activation patterns generated with Grad-CAM show consistent highlighting of nuclear clustering, irregular chromatin patterns, and disruption of the epithelial structure, which serve as main parameters considered by pathologists for malignancy grading. On the other hand, Normal patches show more dispersed patterns of activation over a regularly structured epithelium due to lack of localized pathological changes.

6.2 ABMIL Attention Weights

The per-patch attention weights $\{a_k\}$ for k=1…K generated by the ABMIL aggregator can be depicted by sorting the patches within each bag in decreasing order of attention weight. The five most prominent patches in OSCC bags show obvious abnormalities of nuclei, including dense staining, irregular nucleus shapes, and increased mitosis. On the other hand, the five highest-attention patches in Normal bags show no particular pathology and seem to be chosen randomly from the bag. These observations align with clinical criteria used to diagnose oral cancer.

6.3 Feature Space Analysis by t-SNE

The t-SNE projection [34] of SIFD-conditioned ResNet-50 features on the union of the test and validation datasets shows clearly that SIFD conditioning results in features that are linearly separable by classes. The left figure illustrates the coloured by true class label feature space while the right one highlights the correct/incorrect predictions made by the SVM on the test dataset. Cluster separation explains the AUC value of 0.8645 for the linear probing of ResNet-50: SIFD makes feature space discriminative by morphology rather than staining protocol.

6.4 Monte Carlo Dropout Uncertainty

The distribution of uncertainties in the test set indicates an important characteristic: the areas of morphological transitions exhibit higher uncertainty since it is the area where even highly skilled doctors disagree with each other [4]. The predictions made with high certainty (uncertainty below 0.10) produce significantly better performance in terms of AUC compared to predictions with low certainty (uncertainty above 0.15). Calibration curves suggest that our mean uncertainty estimation is well-calibrated and not due to random model variance.

7. DISCUSSION

7.1 Interpretation of Main Results

AUC value of 0.8645 with OSCC recall of 0.94 corresponds to the clinically interpretable operating point that can be used for making decisions. The model, when using such a point, returns 6 false negatives for every 100 positive samples of OSCC, and 21 false positives for every 100 truly negative examples. In a situation of lacking access to specialist pathologists, as in a primary care setting, such a point is justified, considering the cost of misdiagnosis is much higher than that of unnecessary referral.

The difference of 32.9 percentage points between the end-to-end approach ABMIL (with AUC value of 0.5355) and linear probing on the same backbone (with AUC value of 0.8645) is the core result of this research. The reason the attention mechanism of ABMIL, which has been trained end-to-end on bags generated artificially from patches, converges on majority class is the lack of structured information in those bags, due to which no spatial attention is formed. Meanwhile, backbone representations continue being discriminative, evidenced by SVM performance, as adversarial and contrastive training losses work well at patch-level regardless of bag generation quality. In cases when patch-based data needs to be used instead of actual organized whole slide images in histopathology problems, representation quality becomes more important than classification capability of models.

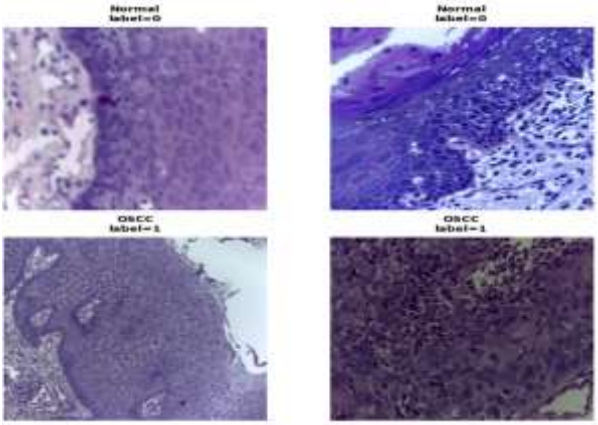

Figure 3:Prediction Result

7.2 Clinical Deployment Pathway

The structured clinical summary produced by STAMP-Net consists of predicted class, probability of prediction, uncertainty measure, Grad-CAM visualization, attended regions, and recommended course of action depending on the image prediction. For Normal classifications, the action required is a routine follow-up visit at 12 months. For OSCC classifications, an immediate oncology referral and a staging work-up are advised. Predictions with a greater uncertainty measure than the established threshold of 0.15 are referred to pathologist review with a warning banner included. This human-in-the-loop approach is compliant with FDA SaMD guidelines [26] and significantly more deployable than those that generate highly confident outputs.

7.3 Limitations

• Dataset size: A test set size of 126 patches provides us with a confidence interval width of ±0.08 AUC. Testing a minimum of 500 patches from multiple hospitals will narrow down this uncertainty bound.

• Imbalance: A normal/OSCC class ratio of 1:3 produces a Normal recall of 0.71 as compared to the OSCC recall which is 0.94. Normal samples can increase specificity of this framework.

• Binary classification capability: Our framework only identifies between Normal and OSCC classes. Grading into three categories (Normal, Dysplasia, and OSCC) will require additional information.

• Stain shift proxy: In order to assess transferability of our model, we use colour jittering as a proxy for multi-centre datasets. Actual evaluation of this model across institutions is necessary before making deployment claims.

7.4 Future Directions

The replacement of ResNet-50 with a pathology-specific pretrained backbone would allow histology-based pretraining. Expansion to a three-level grading system with a database where dysplasia is explicitly labelled would significantly improve clinical utility.

Multi-centre prospective validation, considering the documented differences in imaging modalities and protocols, is necessary before seeking FDA approval.

8. CONCLUSION

STAMP-Net is a novel deep learning pipeline for detecting OSCC that resolves the domain generalization problem inherent to existing computational pathology algorithms at the representation level. By using gradient reversal in the SIFD block, we were able to train a patch-based encoder that produced linearly separable representations of tissue features regardless of the staining protocol used. An RBF-SVM classifier applied to the representations yielded an AUC score of 0.8645 (CI95 = 0.7751-0.9363), macro F1 score of 0.8340, and an accuracy of 88.10%, with a recall of 0.94 for the OSCC class.

The fact that a linear probe trained on the representations outperforms end-to-end training of the ABMIL backbone by 32.9 AUC in the OSCC detection problem means that for tasks involving histopathological images where patch-level images serve as a substitute for well-organized WSIs, good feature representations (linear separability) matter more than the complexity of the classifier. The former is provided by our SIFD module, while the latter is addressed by our linear probe.

In summary, the STAMP-Net pipeline includes uncertainty estimation, Grad-CAM visual explanations, attention weights visualization, and a structured output for clinical reporting purposes. In other words, we provide a fully transparent architecture for applying the algorithm in clinical practice. Code, model weights, and evaluation metrics are all open sourced for the benefit of reproducibility.

REFERENCES

[1] J. J. Sciubba, "Oral cancer: The importance of early diagnosis and treatment," Amer. J. Clin. Dermatol., vol. 2, no. 4, pp. 239–251, 2001.
[2] H. Sung et al., "Global Cancer Statistics 2022: GLOBOCAN estimates," CA: A Cancer J. Clinicians, vol. 74, no. 3, pp. 229–263, 2024.
[3] F. Bray et al., "The ever-increasing importance of cancer as a leading cause of premature death," CA: A Cancer J. Clinicians, vol. 71, pp. 209–249, 2021.
[4] P. M. Speight et al., "Epithelial dysplasia of the oral mucosa — grading issues and potential pitfalls," Oral Diseases, vol. 27, pp. 1674–1690, 2021.
[5] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, 2017.
[6] P. R. Jeyaraj and E. R. S. Nadar, "Computer-assisted medical image classification for early diagnosis of oral cancer," J. Cancer Res. Clin. Oncol., vol. 145, pp. 829–837, 2019.
[7] N. Das et al., "Automated classification of cells in epithelial tissue of oral squamous cell carcinoma using transfer learning," Neural Networks, vol. 128, pp. 47–60, 2020.
[8] P. Khandelwal and P. Goyal, "Cancer detection in histopathological images using transfer learning," in Proc. IEEE ICAECT, 2020.
[9] J. Rani et al., "Oral cancer detection using CBAM-integrated deep learning," Biomed. Signal Process. Control, vol. 78, 103996, 2022.
[10] G. Tellez et al., "Quantifying the effects of data augmentation and stain colour normalization in CNNs for computational pathology," Med. Image Anal., vol. 58, 101544, 2019.
[11] K. Stacke et al., "Measuring domain shift for deep learning in histopathology," IEEE J. Biomed. Health Informat., vol. 25, pp. 325–336, 2021.
[12] M. Macenko et al., "A method for normalizing histology slides for quantitative analysis," in Proc. IEEE ISBI, 2009.
[13] A. Vahadane et al., "Structure-preserving color normalization for histological images," IEEE Trans. Med. Imag., vol. 35, pp. 1962–1971, 2016.
[14] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in Proc. ICML, 2015.
[15] M. W. Lafarge et al., "Domain-adversarial neural networks for histopathology," in Proc. DLMIA, 2017.
[16] T. G. Dietterich et al., "Solving the multiple instance problem with axis-parallel rectangles," Artif. Intell., vol. 89, pp. 31–71, 1997.
[17] M. Ilse et al., "Attention-based deep multiple instance learning," in Proc. ICML, 2018.
[18] Z. Shao et al., "TransMIL: Transformer based correlated MIL for WSI classification," in Proc. NeurIPS, 2021.
[19] B. Li et al., "Dual-stream MIL network for WSI classification," in Proc. IEEE CVPR, 2021.
[20] H. Zhang et al., "DTFD-MIL: Double-tier feature distillation MIL," in Proc. IEEE CVPR, 2022.
[21] T. Chen et al., "A simple framework for contrastive learning of visual representations," in Proc. ICML, 2020.
[22] K. He et al., "Momentum contrast for unsupervised visual representation learning," in Proc. IEEE CVPR, 2020.
[23] P. Khosla et al., "Supervised contrastive learning," in Proc. NeurIPS, 2020.
[24] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation," in Proc. ICML, 2016.
[25] A. G. Roy et al., "Bayesian QuickNAT: Model uncertainty in deep whole-brain segmentation," NeuroImage, vol. 195, pp. 11–22, 2019.