Cross-Domain Knowledge Transfer for Prediction of Chemosensitivity in Ovarian Cancer Patients

Paper ID 33

Asfand Yaar[1], Amina Asif[2], Shan E Ahmed Raza[3], Nasir Rajpoot[3] and Fayyaz Minhas[3]
[1] Department of Computer and Information Sciences, Pakistan Institute of Engineering and Applied Sciences, Pakistan
[2] Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad, Pakistan.
[3] Tissue Image Analytics Lab & PathLAKE CoE, Department of Computer Science, University of Warwick, UK.

*** Corresponding author email**: fayyaz.minhas@warwick.ac.uk

Abstract

In this paper, we report a novel deep neural network framework for prediction of chemo-sensitivity in ovarian cancer patients. The proposed model is based on Multiple Instance Learning (MIL) and a novel variant of Learning using Privileged Information (LUPI). LUPI allows knowledge transfer from highly informative privileged features that are available only at training time to give improved generalization performance on input space features which are available in both training and inference. The proposed model is trained on image patches from Hematoxylin and Eosin (H&E) stained multi-gigapixel whole-slide images (WSIs, the input space) of ovarian cancer tissue sections and their associated gene expression profiles, the privileged feature space. Through cross-domain knowledge transfer with a novel combination of MIL and LUPI, we achieve improved generalization with a limited number of labeled examples in the input space. Informed by the privileged space model output based on relatively expensive and time-consuming gene expression profiles in its training, the proposed LUPI model can generate accurate predictions using routine WSI data alone at the time of inference. The proposed method paves the way for further applications of LUPI in computational pathology and medical image analysis by cross-domain learning especially in cases with a limited number of labeled examples in training.

1. Introduction

Conventional machine learning methods require that all features used in its training are also available at test time. For example, if we use medical imaging data from two different domains (such as radiology and histopathology images) for training a machine learning model, then data from both these domains must be available at test time for prediction as well. Originally proposed by Vapnik and Izmailov [1], [2], Learning using Privileged Information (LUPI) allows overcoming this limitation through cross-domain knowledge transfer in the training phase of machine learning models. LUPI is ideally suited for scenarios in which certain highly informative features, called *privileged space features*, are available during training only, whereas the *input space features* are available for both training and testing examples. For instance, consider the development of an image-based object classification system. If three dimensional (3D) models of different objects are available at training time together with their corresponding 2D images, then these 3D features can be used as privileged space information for improving prediction accuracy of the image based object classification model without requiring 3D feature input at test time. From a machine learning perspective, LUPI allows knowledge from privileged space features to improve the decision boundary in the input feature space [1], [3]. Another way of looking at LUPI is by considering it as a means of knowledge distillation from a "teacher" model trained over privileged space features to a "student" model that operates in the input space only [4]. LUPI has been applied in a number of useful machine learning applications [5]–[9].

In this work, we report a novel algorithm for LUPI based prediction of response to chemotherapy using both whole slide images of tumor biopsies as well as gene expression profiles. Computational analysis of histopathology images is an active area of research due to its importance in the diagnosis of cancer as well as treatment selection for cancer patients [10]–[13]. In addition to tumor grading and profiling, computational pathology also allows development of machine learning models for prediction of response to various cancer treatments such as chemotherapy [14]–[16]. For this purpose, a multi-gigapixel whole slide image (WSI) is first obtained by digitizing the tissue slide from the cancerous tissue specimen. After collecting a training WSI dataset from a number of cancer patients together with information about their response to chemotherapy, a machine learning model can be built for predicting chemosensitivity [17]. In addition to the WSI data, gene expression profiles can also be used for predicting response to chemotherapy [18]–[20]. Typically, gene expression profiling presents a more detailed molecular picture for drug response prediction in comparison to WSI analysis but it is significantly more expensive and time consuming [15]. It is important to note that collection of large cohorts of WSI and genetic profiling

data is very expensive and requires data-efficient machine learning models. In this work, we overcome this challenge through LUPI. Furthermore, due to computational constraints, WSI-based prediction requires that the multi-gigapixel WSI be broken down into patches and weak labels at the WSI-level be used for effective learning at the patch level. In this work, we have achieved this through Multiple Instance Learning. The main contributions of this work are listed below:

- We propose a novel method of knowledge distillation based LUPI for chemosensitivity prediction. It can predict whether an Ovarian Carcinoma (OvCa) patient will respond to adjuvant chemotherapy or not by using both gene expression information as well as WSI data in training while requiring only WSI data at test time. To the best knowledge of the authors, this is the first such cross-domain model in computational pathology.
- Our WSI based pipeline uses multiple instance learning to generate prediction labels at the WSI level using patches drawn from a given WSI.
- In contrast to existing LUPI implementations which are limited to classification tasks only, the LUPI model proposed in this paper can be applied to regression and survival prediction as well as to other cross-domain learning problems.
- We show that the use of LUPI significantly improves the accuracy of WSI-based chemosensitivity prediction over a large 220 patient dataset from The Cancer Genome Atlas (TCGA).

2. Materials and Methods

2.1. Problem formulation

In this work, our objective is to develop a machine learning method for predicting the response of an ovarian cancer patient to adjuvant chemotherapy (sensitive or resistant). For this purpose, we have used a dataset of 220 cancer patients with known chemosensitivity. This dataset is taken from a study aimed at survival prediction for ovarian cancer patients using gene expressions data by Liu *et al*. [15]. For each patient, pre-processed gene expression levels for a total of ~14,000 genes are available together with WSIs of the tissue slides scanned at 20× magnification from TCGA [21]. The average size of the images in this dataset is ~40,000×40,000 pixels. In line with the work in [15], the patients have been labeled according to their response to platinum chemotherapy treatment: a negative label (chemo-resistant) is assigned to a patient if disease symptoms reappear within 6 months of treatment whereas a patient is labeled as chemo-sensitive (positive) if a period of more than 6 months has elapsed since the last chemotherapy treatment and there is no evidence of recurrence within a 6 month follow-up period involving no additional treatment sessions. This dataset contains 154 positive and 66 negative examples. Formally, this dataset can be written as a set $\{(\boldsymbol{x}_i^*, \boldsymbol{x}_i, y_i) | i = 1 \dots N\}$, where, $\boldsymbol{x}_i^*$ and $\boldsymbol{x}_i$ represent the gene expression profile and WSI corresponding to a single patient, respectively, and $y_i \in \{+1, -1\}$ is the associated label.

In their study, Liu *et al*. showed that gene expression analysis can be used for survival prediction with high accuracy. However, acquiring gene expression profiles for a given patient can be expensive and time consuming. WSI based analysis does not require an additional sequencing or profiling facility and can be done in conjunction with tumor grade assessment on diagnostic slides at no extra cost. Hence, an image-based predictor of chemosensitivity is desirable for routine practice. In this work, we propose a novel variant of learning using privileged information (LUPI) model that considers gene expression profile as privileged information and WSI data as input space data for improved accuracy of chemosensitivity prediction using WSI data alone. Before moving to the description of the proposed LUPI model, we first discuss how gene expression profiling and WSI can be independently employed for chemosensitivity prediction.

2.2. Prediction using gene expression data

The gene expression data for each patient in the dataset comprises of expression levels for ~14,000 genes with an averaged activity level of different genes in the sample. In their survival prediction study, Liu *et al*. identified a subset of 227 genes that are predictive of survival of an ovarian cancer patient after being given adjuvant chemotherapy [15]. We used the same subset of 227 genes as input features to develop a classical machine learning model (Radial Basis Function Kernel Support Vector Machine (SVM) to predict chemo-sensitivity. This choice is motivated by the fact that a simple model with low VC dimension can be useful for privileged space predictions [1]. Mathematically, the output of the SVM model for a given gene expression profile feature vector $\boldsymbol{x}_i^*$ can be written as $f_t(\boldsymbol{x}_i^*)$.

2.3. Prediction using histology WSI data

For the input space WSI image data, which is high-resolution (each image of the order of 100K×80K pixels), we developed a Multiple Instance Learning (MIL) based Convolutional Neural Network (CNN) model to predict chemo-sensitivity in ovarian cancer patients. Before training the model, we performed the following preprocessing over the WSIs.

2.3.1 Preprocessing

The IBM pipeline for histopathology image analysis has been employed for segmenting the tissue region from the background based on 'activity' in terms of nuclei pixels

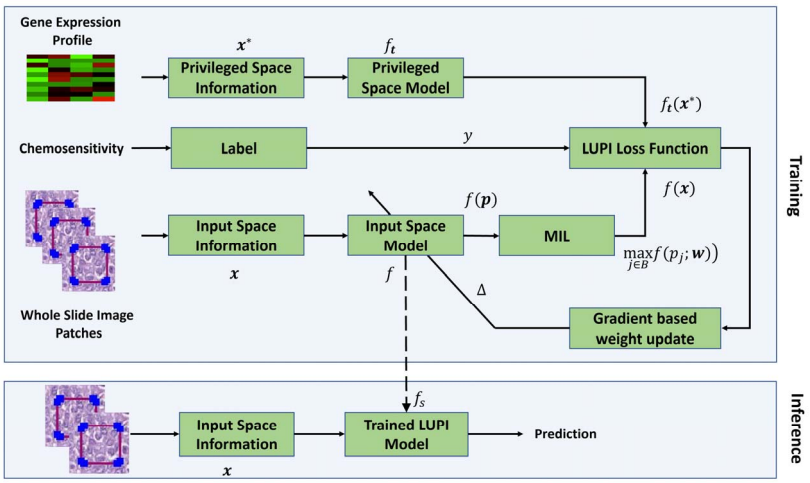

Figure 1- LUPI based chemosensitivity prediction model. In training, gene expression profile and WSI data are used as privileged and input space features, respectively, to train an input space model. In testing, the trained model can then generate predictions using WSI data alone.

[22]. After background removal, the WSI is divided into tiles of size 1,536×2,048 pixels. Each tile is then scored based on its tissue content and top 3 tiles from each image are selected for downstream analysis. Since WSIs are prone to color variations due to staining and scanning conditions of tissue samples, we performed stain normalization over all images using the method proposed by Rakhlin *et al*. [23]. Each tile extracted from the WSIs is then further divided into 12 smaller, non-overlapping patches of 512×512 pixels. Consequently, each WSI is represented by a total of 36 patches which are used for training purposes.

2.3.2 Patch Classification

We have used a Convolutional Neural Network (CNN) to model the problem of chemo-sensitivity prediction from WSI data. Specifically, a standard convolutional neural network consisting of 5 convolutional blocks with a fully connected layer of 4,096 neurons following by a single output neuron is utilized for binary classification at the patch level. Each of the 5 convolution blocks has two 3×3 convolution layers followed by a single 2×2 convolutional layer with batch normalization and rectified linear unit activation in each layer. The 5 convolutional blocks have 16, 32, 64, 128 and 256 convolution filters for each layer in them. The choice for this neural network architecture has been inspired by its effectiveness in the the breast cancer histology image classification study by Nazeri *et al*. [24]. The neural network can generate a prediction score for a single patch of a WSI. The neural network is trained with a multiple instance learning based loss function described in the next section.

2.3.3 Multiple Instance Learning

One of the issues associated with the WSI data is that labels are available for each patient and not for individual patches in the WSI. Thus, we need to aggregate the patch level predictions to a slide level label. We overcome this aggregation problem using Multiple Instance Learning (MIL). Multiple Instance Learning is used to solve machine learning problems where labels are not associated with individual instances but with groups of instances called bags [25]. Several neural network based solutions have been proposed in the literature for modeling MIL [26]–[30].

Here, each WSI is considered as a single bag with each patch as an instance in the bag to model tumor heterogeneity [31]. In order to train the aforementioned CNN with MIL, all patches belonging to a WSI are passed to the CNN one-by-one and the highest scoring patch (based on prediction score) in a WSI is used for computing the loss function and weight updates, i.e., the prediction score of a WSI is computed as the score of the maximum scoring patch in it or $f(\boldsymbol{x_i}) = \max_{j \in B_i} f(\boldsymbol{p_j}; \boldsymbol{w}))$. Mathematically, the MIL training of the CNN can be expressed as the following optimization problem [30]:

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} l(y_i, \max_{j \in B_i} f(p_j; \boldsymbol{w}))$$

where $f(\boldsymbol{p_j}; \boldsymbol{w})$ indicates the output of the CNN with weight parameters $\boldsymbol{w}$ for a patch $\boldsymbol{p_j}$ from bag B_i of 36 patches in WSI represented by $\boldsymbol{x}_i$, $l(y, s) = \max(0, 1 - ys)$ is the hinge loss function and N denotes the number of patients in the training dataset.

2.4. Learning using Privileged Information (LUPI)

In this work, we model chemosensitivity prediction as a learning using privileged information problem. In LUPI, privileged space features are assumed to be more informative, however, they are available during training only. The input space features are available during both

training and testing phases. A teacher or privileged space model $f_t(\boldsymbol{x_i^*}) \in \mathcal{F}_t$ is trained using privileged features $\boldsymbol{x_i^*}$ and then knowledge from the teacher is distilled to a student model $f_s(\boldsymbol{x_i}) \in \mathcal{F}_s$. The student model uses input space features $\boldsymbol{x_i}$ only to generate predictions for inference. We have developed a LUPI model that performs distillation or transfer of knowledge from a trained privileged space model to an input space model. This is achieved by training the input space student model by minimizes the following objective based on a custom LUPI meta-loss function:

$$f_s = \underset{f \in \mathcal{F}_s}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \big(1 - \exp(-Tl(f_t(\boldsymbol{x_i^*}), y_i)\big) l(f(\boldsymbol{x_i}), y_i) + \exp(-Tl(f_t(\boldsymbol{x_i^*}), y_i)) l\big(f(\boldsymbol{x_i}), f_t(\boldsymbol{x_i^*})\big)$$

Here, $l(f_t(\boldsymbol{x_i^*}), y_i)$ represents the loss value of the privileged space model for the i^{th} example, $l(f_s(\boldsymbol{x_i}), y_i)$ is the loss between the input space model score and the actual label and $l\big(f(\boldsymbol{x_i}), f_t(\boldsymbol{x_i^*})\big)$ is the loss between the predictions of the input space and the privileged space model. The non-negative hyperparameter T controls the extent of knowledge transfer: smaller values of T encourage the input space model to mimic the privileged space model, whereas, for larger values of T or the cases where the privileged space model has high error, the input space model tries to learn from the true labels instead. In contrast to existing LUPI implementations which are restricted to classification, the proposed LUPI model can be used for regression and other machine learning tasks as well depending upon the choice of the loss function. For further details on the formulation of the above loss function and the role of the knowledge transfer control parameter T, the interested reader is referred to our manuscript [32].

The learning scheme of LUPI used in this work is illustrated in Figure 1. As discussed earlier, we treat gene expression profile-based features as privileged information and the corresponding RBF-SVM model as the privileged space model. The MIL CNN model for WSI-based chemosensitivity prediction is used as the input space model. The input (or WSI) space model generates a prediction score for a given WSI based on the score of the maximum scoring patch in the given image, i.e., $f_s(\boldsymbol{x_i}) = \max_{j \in B_i} f_s(p_j)$. The same MIL CNN model architecture is then trained as a student model using the meta-loss function described above with knowledge distillation from the privileged space model. This results in a LUPI model that uses both input and privileged space information in its training while still being able to generate predictions with the input features alone at test time.

2.5. Implementation & Performance Evaluation

In order to analyze the performance of the input space (WSI-based), privileged space (gene expression profile based) and LUPI models, we use 5-fold stratified cross-validation with the same division of examples into folds for the three models. The average and standard deviation of the area under the receiver operating characteristic curve (AUC-ROC) is used as the performance metric. The hyperparameters for the RBF-SVM and the LUPI CNN models are selected based on a small validation set of 10 cases. The code for the proposed model is available at the URL: https://github.com/asfandasfo/LUPI

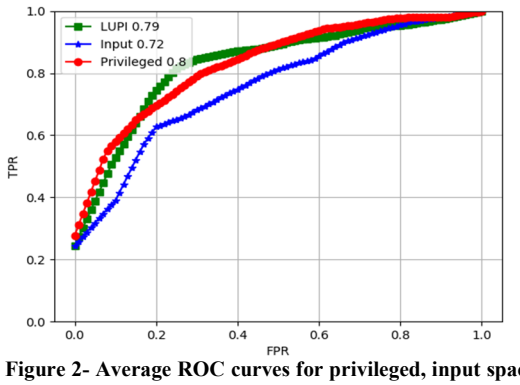

Figure 2- Average ROC curves for privileged, input space and LUPI models.

3. Results and Discussion

Averaged ROCs for the three models across 5-fold cross-validation are shown in the Figure 2. The privileged space model trained over gene expression profiles shows better performance with average AUC-ROC of 0.8 (standard deviation or SD: 0.03) as compared to 0.72 (SD: 0.10) for the WSI or input model trained. The LUPI based model that uses WSIs as input with knowledge transferred from the privileged space model produces an average AUC of 0.79 (SD: 0.07). The performance of LUPI model is not only consistently and substantially better than the simple input space model across each fold, it is also comparable to the performance for the privileged space model, demonstrating effectiveness of the proposed LUPI model for chemosensitivity prediction using only WSI data at test time.

4. Conclusions and Future Work

We have shown that learning using privileged information can be effectively applied to improve the generalization performance for prediction of chemosensitivity in OvCa patients by cross-domain knowledge distillation from gene expression profiling to whole slide imaging. We have also shown that LUPI allows efficient use of data in the input space by cross-domain learning. This work can form the basis of further applications of LUPI knowledge distillation in computational pathology and cross-domain medical image analysis. Future work can be focused on identifying associations between gene expression profiles and WSI features and validation of the proposed method on larger datasets.

Acknowledgements:

SAR, NR and FM are partially supported by the PathLAKE digital pathology consortium, which is funded from the Data to Early Diagnosis and Precision Medicine strand of the government's Industrial Strategy Challenge Fund, managed and delivered by UK Research and Innovation (UKRI). URL: https://www.warwick.ac.uk/PathLAKE.

References

[1] V. Vapnik and A. Vashist, "A new learning paradigm: Learning using privileged information," *Neural Netw.*, vol. 22, no. 5, pp. 544–557, 2009.

[2] V. Vapnik and R. Izmailov, "Learning Using Privileged Information: Similarity Control and Knowledge Transfer," *J. Mach. Learn. Res.*, vol. 16, pp. 2023–2049, 2015.

[3] V. Vapnik and R. Izmailov, "Knowledge transfer in SVM and neural networks," *Ann. Math. Artif. Intell.*, vol. 81, no. 1–2, pp. 3–19, 2017.

[4] D. Lopez-Paz, L. Bottou, B. Schölkopf, and V. Vapnik, "Unifying distillation and privileged information," in *International Conference on Learning Representations (ICLR) May 2–4, 2016*, San Juan, Puerto Rico.

[5] X. Xu, W. Li, and D. Xu, "Distance metric learning using privileged information for face verification and person re-identification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3150–3162, 2015.

[6] H.-T. Shiao and V. Cherkassky, "Learning using privileged information (LUPI) for modeling survival data," in *Neural Networks (IJCNN), 2014 International Joint Conference on*, 2014, pp. 1042–1049.

[7] Y. Yan, F. Nie, W. Li, C. Gao, Y. Yang, and D. Xu, "Image classification by cross-media active learning with privileged information," *IEEE Trans. Multimed.*, vol. 18, no. 12, pp. 2494–2502, 2016.

[8] J. Feyereisl, S. Kwak, J. Son, and B. Han, "Object localization based on structural SVM using privileged information," in *Advances in Neural Information Processing Systems*, 2014, pp. 208–216.

[9] W. Li, L. Niu, and D. Xu, "Exploiting privileged information from web data for image categorization," in *European Conference on Computer Vision*, 2014, pp. 437–452.

[10] D. Bošnački, N. van Riel, and M. Veta, "Deep Learning with Convolutional Neural Networks for Histopathology Image Analysis," in *Automated Reasoning for Systems Biology and Medicine*, Springer, 2019, pp. 453–469.

[11] Y. Xu *et al.*, "Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features," *BMC Bioinformatics*, vol. 18, no. 1, p. 281, May 2017, doi: 10.1186/s12859-017-1685-x.

[12] J. Arevalo, A. Cruz-Roa, and others, "Histopathology image representation for automatic analysis: A state-of-the-art review," *Rev. Med*, vol. 22, no. 2, pp. 79–91, 2014.

[13] M. Veta, J. P. Pluim, P. J. Van Diest, and M. A. Viergever, "Breast cancer histopathology image analysis: A review," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1400–1411, 2014.

[14] D. Komura and S. Ishikawa, "Machine Learning Methods for Histopathological Image Analysis," *Comput. Struct. Biotechnol. J.*, vol. 16, pp. 34–42, Feb. 2018, doi: 10.1016/j.csbj.2018.01.001.

[15] Y. Liu *et al.*, "Integrated Analysis of Gene Expression and Tumor Nuclear Image Profiles Associated with Chemotherapy Response in Serous Ovarian Carcinoma," *PLOS ONE*, vol. 7, no. 5, p. e36383, May 2012, doi: 10.1371/journal.pone.0036383.

[16] U. Djuric, G. Zadeh, K. Aldape, and P. Diamandis, "Precision histology: how deep learning is poised to revitalize histomorphology for personalized cancer care," *Npj Precis. Oncol.*, vol. 1, no. 1, pp. 1–5, Jun. 2017, doi: 10.1038/s41698-017-0022-1.

[17] H. R. Ali *et al.*, "Computational pathology of pre-treatment biopsies identifies lymphocyte density as a predictor of response to neoadjuvant chemotherapy in breast cancer," *Breast Cancer Res.*, vol. 18, no. 1, p. 21, Feb. 2016, doi: 10.1186/s13058-016-0682-8.

[18] J. D. Wells and T. W. Miller, *Development of pan-cancer transcriptional signatures that predict chemosensitivity*. AACR, 2019.

[19] X. Zhang, I.-H. Cha, and K.-Y. Kim, "Use of a Combined Gene Expression Profile in Implementing a Drug Sensitivity Predictive Model for Breast Cancer," *Cancer Res. Treat. Off. J. Korean Cancer Assoc.*, vol. 49, no. 1, pp. 116–128, Jan. 2017, doi: 10.4143/crt.2016.085.

[20] X. Lu, J. Lu, B. Liao, X. Li, X. Qian, and K. Li, "Driver pattern identification over the gene co-expression of drug response in ovarian cancer by integrating high throughput genomics data," *Sci. Rep.*, vol. 7, no. 1, pp. 1–17, Nov. 2017, doi: 10.1038/s41598-017-16286-5.

[21] K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge," *Contemp. Oncol.*, vol. 19, no. 1A, p. A68, 2015.

[22] *CODAIT/deep-histopath*. IBM CODAIT - Center for Open-source Data & AI Technologies, 2020.

[23] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis," *ArXiv Prepr. ArXiv180200752*, 2018.

[24] K. Nazeri, A. Aminpour, and M. Ebrahimi, "Two-Stage Convolutional Neural Network for Breast Cancer Histology Image Classification," *ArXiv180304054 Cs*, vol. 10882, pp. 717–726, 2018, doi: 10.1007/978-3-319-93000-8_81.

[25] B. Babenko, "Multiple Instance Learning : Algorithms and Applications," 2008.

[26] J. Wu, Y. Yu, C. Huang, and K. Yu, "Deep multiple instance learning for image classification and auto-annotation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3460–3469.

[27] O. Z. Kraus, J. L. Ba, and B. J. Frey, "Classifying and segmenting microscopy images with deep multiple instance learning," *Bioinformatics*, vol. 32, no. 12, pp. i52–i59, Jun. 2016, doi: 10.1093/bioinformatics/btw252.

[28] Y. Xu *et al.*, "Deep learning of feature representation with multiple instance learning for medical image analysis," in *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2014, pp. 1626–1630.

[29] X. Liu *et al.*, "Deep multiple instance learning-based spatial–spectral classification for PAN and MS imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 461–473, 2018.

[30] A. Asif and F. ul A. A. Minhas, "An embarrassingly simple approach to neural multiple instance classification," *Pattern*

Recognit. Lett., Oct. 2019, doi: 10.1016/j.patrec.2019.10.022.

[31] B. Rybinski and K. Yun, "Addressing intra-tumoral heterogeneity and therapy resistance," *Oncotarget*, vol. 7, no. 44, pp. 72322–72342, Sep. 2016, doi: 10.18632/oncotarget.11875.

[32] A. Asif, M. Dawood, and others, "A generalized meta-loss function for distillation and learning using privileged information for classification and regression," *ArXiv Prepr. ArXiv181106885*, 2018.