A Robust Colon Cancer Detection Model Using Deep-Learning

Vanishka Kadian
Department of Information Technology
Delhi Technological University
New Delhi, India
vanishka.kadian22@gmail.com

Arushi Singh
Department of Information Technology
Delhi Technological University
New Delhi, India
arushi10122000@gmail.com

Kapil Sharma, Senior Member, IEEE
Department of Information Technology
Delhi Technological University
New Delhi, India
kapil@ieee.org

2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC) | 979-8-3503-0071-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSCCC58608.2023.10176535

***Abstract*—Colon cancer is the third most common cancer, with a high mortality rate and a high pathological need for early and accurate diagnosis. Histopathology is a golden standard tool for diagnosing almost all types of cancer, including colon cancer. It can be enhanced using deep learning techniques if the key challenges, namely, clean data annotation and inter-observer variability, are addressed. This highlights a need for a noise-robust algorithm with computational feasibility to work on real-world clinical data. In this paper, the experiment uses different deep learning models– ResNet-34, XCiT, SqueezeNet, MobileNet– individually integrated with a data cleaning architecture in an attempt to (1) achieve unmatched accuracy, (2) draw a comparison amongst the performance of these models and (3) point to the best model as a solution to the given problem. MobileNet outperforms all other models as the backbone for the algorithm, achieving the highest accuracy (84.39%), thus boosting the algorithm performance on the Chaoyang dataset.**

***Index Terms*—Image Classification, Noise handling, Deep-learning, Colon cancer**

I. Introduction

Cancer is a leading cause of death all over the world, accounting for 10 million deaths every year. Colorectal cancer is the third most common cancer, accounting for an estimated 1.8 million mortality cases annually. Although this death rate is higher in Western countries, growing urbanization has also led to a rapid increase in developing countries.

The early pathological diagnosis of colon cancer is crucial for reducing mortality and enhancing treatment efficiency. Studies suggest that post-detection and treatment, the five-year survival rate for early-stage colon cancer is estimated as high as 90%, whereas late-stage colon cancer can be as low as 12%. Histopathology is an essential means for cancer diagnosis, allowing for the microscopic examination of tissue samples to identify cancerous cells and determine the stage and type of cancer. For diagnosis, a sample of suspicious cells and tissues is taken, and the nature of the abnormality is identified. If found malignant, it requires classification of the type of cancer and its grade. However, the microscopic evaluation of a large number of tissue samples can be tedious and time-consuming. Thus, the enormously increasing number of colonoscopy examinations is significantly straining the healthcare system.

Additionally, the accuracy of histopathology is highly dependent on the expertise and experience of pathologists and can lead to conflict in the biopsy results due to interobserver variability. The pathological standards for the diagnosis of colon lesions are also not standardized. Western pathologists diagnose cancer by observing the invasion through the muscularis mucosa into the submucosa. In contrast, Eastern pathologists rely on identifying nuclear and architectural changes in cells. Therefore, there is a requirement for a more unified approach to bringing consensus among specialists.

Developing deep learning methods to assist pathologists in improving the accuracy, consistency, and promptness of cancer diagnosis is a widely accepted solution. Deep learning algorithms extract information from complex images and can predict clinical outcomes directly from histological images of colorectal cancer. A Japanese study recently showcased relatively good accuracy of a classification algorithm compared to human pathologists in categorizing colorectal tumours [1]. Further, a German study demonstrated the application of a deep-learning-based model for the prediction of survival of colorectal cancer patients using histopathology images [1]. The application of image classification, segmentation, anomaly detection, and transfer learning has shown promising results, but several limitations still need to be addressed.

The first limitation is that deep learning algorithms require extensive amounts of annotated data to be trained accurately, but the annotation process can be challenging. In a real medical scenario, collecting a large dataset with clean labels is difficult, as mislabeling is nearly unavoidable in manual annotation. Another limitation is the inter-observer variability that can arise in histopathology, even among experienced pathologists. Samples wherein labelling is conflicted are termed hard samples. This variability can make it difficult to develop deep learning algorithms that are robust and consistent across different data sets and users.

While the hard samples enhance the efficiency and effectiveness of training the models, the noisy labels are misleading and result in lower accuracy. Thus, distinguishing the noisy samples from the hard samples proves essential for attaining good model performance.

Proposing a two-phase Hard Sample Aware Noise Robust Learning (HSA-NRL) algorithm, Zhu et al. [2] attempted to involve hard samples for training while simultaneously reducing noise interference. Prediction history for the samples was

used as a key for their differentiation and a self-trained noisy label correction architecture, followed by steps to suppress the noise and enhance the hard samples. The algorithm used the ResNet-34 [3] model as the backbone on the Chaoyang dataset [4].

This paper uses the network pipeline designed by Zhu et al. [2] while replacing the backbone with different models like – Cross-Covariance Image Transformer (XCiT) [5], SqueezeNet [6], and MobileNet [7] on the Chaoyang dataset. The experiment aims to achieve an outperforming test accuracy against that achieved using ResNet-34. A comparative analysis has been conducted between these models based on performance in terms of test accuracy. The work concludes by identifying the best model to be integrated with the noise robust algorithm for real-world pathological image classification without the requirement of cleaning annotations.

This remaining paper is divided into five sections. Section II describes existing works on histopathology image processing, its application in colon cancer detection and the algorithms for handling noisy data. Section III explains the two-phased architecture used and the integrated deep-learning models. Section IV comprises three subsections: dataset description, implementation and parameter settings, and the evaluation criteria. Results are tabulated and discussed in Section V. Finally, the conclusion and future scope are given in Section VI.

II. Related Works

In the field of histopathology, pathologist fatigue, the intricacy of the tissue structure, and numerous other factors negatively impact the accuracy of cancer diagnosis, as suggested by Wang et al. [8]. Lusted [9] was the first to see the potential of computers in medical diagnosis in 1955. Since then, the emergence of deep learning (DL) methods has changed the field altogether. For improving the accuracy and spontaneity in the disciplines of histopathological image processing, convolutional neural network (CNN), one of the DL approaches, has been extensively applied, as presented by Xu et al. [10], and Li et al. [11]. In several publications, DL models are proposed to categorize histopathology images [12]– [15], including colorectal histology images [16]- [18].

Zhu et al. [12] showcased significant harm in model performance due to over-pruning. Caicedo et al. [13] proposed various representations for classifying histopathology images using the bag of features method. However, the performance degraded as the size of the code block increased. Xue et al. [14] built a new conditional generative adversarial network (cGAN) model called HistoGAN for high-fidelity histopathology image synthesis, and a synthetic augmentation approach with quality assurance was suggested. In the work of Coudray et al. [15], a DL model for the automatic analysis of lung tumour slides was developed. However, it did not perform well for other cancers or datasets. Masud et al. [16] proposed a unique supervised learning method based on DL that could classify five distinct cancerous and non-cancerous tissue types in lung and colon cancers. Gessert et al. [17] noted the possibility of identifying colon cancer in conventional light microscopy (CLM) images using CNNs and other transfer learning scenarios. However, he concluded that there was no single transfer method ideal for all CLM classification problems. Rahman et al. [18] proposed feature selection for colon cancer classification using the Best first search method and Artificial Neural Networks (ANNs), showcasing several benefits, including the ability to handle large data, less classification time and a lower probability of missing important information.

Behzadi et al. [19] enabled clinical-grade tissue classification and diagnosis accuracy while enabling computationally efficient analysis of large-scale whole slide imaging (WSI) datasets. However, all these works assumed the availability of large clean annotated datasets for model training. In manual annotation, noisy labels are frequently unavoidable and can adversely affect the performance of the algorithms.

Wei et al. in [20] suggested that designing robust algorithms for noisy labelling is crucial. Various approaches can typically be divided into three families of models: estimating the noise transition matrix [21], [22], constructing noise-robust loss functions [23]– [25], and estimating true labels [26]– [28]. Hendrycks et al. [21] and Patrini et al. [22] put forth various strategies for estimating transition matrices, including using trusted samples in a data-efficient manner and using a two-step estimating scheme, respectively. However, these models were also unsuccessful with real-medical datasets.

The second family of models aimed at designing loss functions with greater noise tolerance. Gehlot et al. [23] introduced projection loss to maximize the projection of a sample's activation on the corresponding class vector in addition to imposing orthogonality constraints on the class vectors. This created a dual-branch architecture that boosted performance and handled noise. Wang et al. [24] proposed learning with noisy labels by adopting a Symmetric Cross Entropy (SCE) loss function that outperformed several existing methods in terms of accuracy and stability. However, this method observed poor generalization performance. DivideMix is a framework introduced by Socher et al. [25] that uses semi-supervised learning techniques to enhance the accuracy of a model by learning with noisy labels.

The third category focused on selecting possibly clean labels and correcting noisy labels. Tanaka et al. [26] suggested a joint optimization strategy for predicting true labels and learning parameters. The label correction was done by altering updates to the labels and the parameters during training. Yi et al. [27] published a paper proposing a Probabilistic end-to-end architecture for updating network parameters and label estimations as label distributions. Due to its independence from the backbone network structure and lack of requirements for a clean auxiliary dataset or prior knowledge of noise, this architecture was more robust and simple to use. However, due to their sensitivity to hyperparameters, [26], [27] faced limitations in their performance. The work of Goldberger et al. [28] suggested an expectation-maximization (EM) algorithm to determine the network and noise parameters and estimate the

correct label. Adding a second softmax output layer that linked the correct labels to the noisy ones made the algorithm easily integrable with any deep-learning implementation. In contrast to [27], [28] allowed the dataset to be trained on noisy data.

III. Methodology

This work uses a two-phased architecture: Phase I produces an almost clean dataset by label correction, and Phase II uses this generated dataset to give a robust classification model.

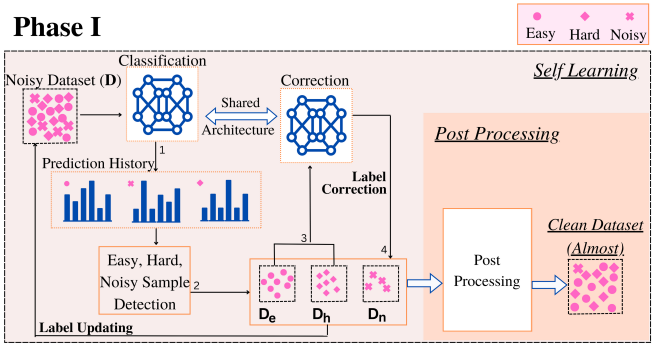

Fig. 1. Phase I of the architecture: produces an almost clean dataset using iterative self learning steps (obtain prediction behavior; generate Easy, Hard, Noisy subsets; train correction model; update labels) and the post processing action.

As depicted in Fig. 1, Phase I of the architecture comprises two steps: Self Learning and Post Processing.

During Self Learning, a Bayesian Classifier is trained on the original noisy dataset, denoted by D. A prediction history for all samples is obtained over all epochs and mean of the prediction history is calculated to detect Easy, Hard, and Noisy samples. Compared to noisy samples, the clean samples have a higher mean prediction value. Hence, by applying a threshold, some clean samples can be extracted. These samples are called 'easy' samples. The samples below the threshold may be noisy or clean. The clean samples below the threshold are called 'Hard' samples and must be further distinguished.

The subset of easy samples, denoted by De is used, and some labels are switched to add noise. The noise ratio of this thus-created subset will be the same as the original dataset. Therefore, the noisy subset is used to train a convolutional neural network (CNN) classification model, generating a new training history according to which the easy samples are discarded. A multilayer perceptron (MLP) classifier is trained on the training history of the remaining (noisy and hard) samples and a corresponding set of labels stating whether they are hard or noisy. Post training, the MLP classifier can distinguish the samples below the threshold in D into 'hard' and 'noisy' subsets, denoted by Dh and Dn respectively.

The next part of Self Learning accounts for label correction. The Correction model is trained using De and Dh and some noisy labels are corrected. From the model output, the most probable class is chosen as the pseudo label to replace the labels of hard and noisy samples. The labels are updated and the entire self-learning process is repeated to purify the dataset.

The next step of Phase I comprises the Post Training Action, which aims to discard the samples that could not be corrected and thus remained noisy even post the Self Learning steps. These are the samples from the Dn for which the labels were not changed and the samples from Dh for which the labels were changed. Discarding these samples generates an almost clean dataset.

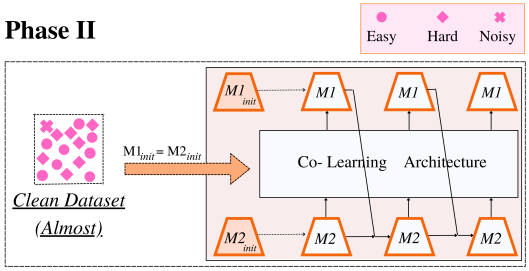

Fig. 2. Phase II of the architecture: co-trains two models using the resulting dataset, updating their parameters by back-propagation and momentum coefficient value.

Phase II of the architecture enhances hard samples and suppresses noise, as shown in Fig. 2. The hard samples are proven to improve the efficiency of training of the model, while the noisy samples are misleading and result in poor accuracy. This phase uses a co-learning architecture to avoid confirmation bias. Two models, M1 and M2 are initialized with the same parameters(M1_init= M2_init) but adopt different parameter updating schemes. This further suppresses the interference of noisy samples. The parameter for the first model is updated by backpropagation, while the second model uses momentum coefficient value. M2 selects training data for M1 at each epoch by discarding the samples with lower prediction probabilities. Further, the focal loss is used to emphasize the hard sample subset. The focal loss was calculated using the predicted probability of the actual class and a hyper-parameter.

The following deep-learning models are individually used and replaced as the backbones for the algorithm:

1) ResNet-34 developed by K. He et al. [3] is a convolutional neural network with 34 layers. It utilizes the residuals from each layer in the subsequent connected layers, hence training deep networks and minimizing the percentage of errors.
2) XciT, proposed by Nouby et al. [5], operates across feature channels instead of tokens. It leverages the cross-covariances matrix between keys and queries for interactions.
3) SqueezeNet, given by Iandola et al. [6], reduces the number of parameters by applying design techniques, particularly the usage of fire modules that "squeeze" parameters using 1x1 convolutions.
4) MobileNet by Howard et al. [7] introduces the use of two global hyper-parameters. These parameters enable an efficient balance between latency and accuracy while incorporating a streamlined architecture that em-

ploys depth-wise separable convolutions for developing lightweight deep neural networks.

IV. Experiment Implementation

A. Dataset

The Chaoyang dataset comprises Colon slide samples for training and testing with the following domination: 1111 normal, 842 serrated, 1404 adenocarcinoma, 664 adenomas, and 705 normal, 321 serrated, 840 adenocarcinomas, 273 adenomas, respectively. This noisy dataset was created by collecting data from Chaoyang hospital, China. These are colon slides with a 512 × 512 patch size. Three expert pathologists labeled the patches. Conflicted labeling was observed for about 40% of the samples. We used the portion of labeled patches with the three pathologists' consensus as the testing set. The remaining samples, with inconsistent labeling, were used as the training set with a random selection of one of the suggested labels. Fig. 3 shows the sample patches of the Chaoyang dataset.

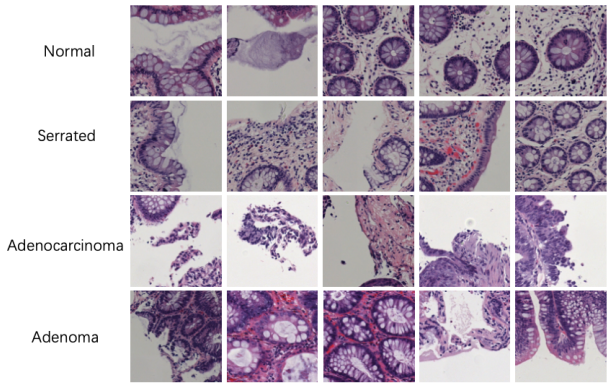

Fig. 3. The Chaoyang Dataset

B. Implementation and Parameter settings

We employed the four models - ResNet-34, XCiT, MobileNet, and SqueezeNet on the Chaoyang dataset and trained it using Adam with a momentum of 0.9 and a batch size of 16. During both, Phase I and Phase II, the networks were trained for 80 epochs. Furthermore, after 30 epochs, the learning rates were linearly reduced, the initial value being 0.001.

For the Chaoyang dataset, the noise cross-validation approach given by Chen et al. [29] was used to compute the dataset noise ratio, ρ. The easy sample threshold was set to 0.1 for an 80% noise ratio. The hyper-parameter used to calculate the focal loss was set to 2, and the discarding ratio was set to $0.1 * \rho$.

C. Evaluation Criteria

For each model, the Test Accuracy (ACC), Precision, Recall, and F1 Score (F1), as defined below, have been used as the evaluation criteria:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

V. Results And Discussions

In the network pipeline of the HSA-NRL algorithm for histopathology image classification, we introduced new backbones to analyze the performance of the latest deep learning models on replacing ResNet-34 in the algorithm. The four models—ResNet-34, XCiT, Squeezenet, and Mobilenet—were initially trained using the same dataset (The Chaoyang dataset). The models were trained on 4021 sample images, followed by processing and cleaning, which generated 4019 clean images. Out of these, 2139 almost clean images were used for testing.

TABLE I
ACC, Precision, Recall, F1 Score of the models used

Model	*ACC*	*Precision*	*Recall*	*F1*
ResNet-34	83.40	78.33	75.45	76.54
XCiT	80.83	75.82	75.01	75.41
SqueezeNet	78.07	74.42	73.01	73.71
MobileNet	84.39	81.79	79.67	80.73

Table I presents a detailed comparison of the model performances, listing respective test Accuracy, Precision, Recall, and F1 Scores for our experiment on the Chaoyang dataset. The network with MobileNet as the backbone consistently outperforms the other models on all these metrics.

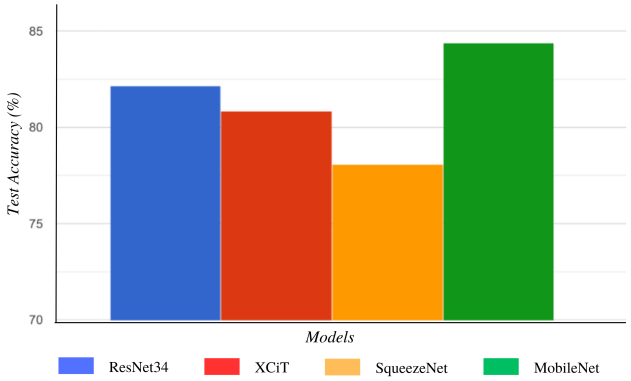

Fig. 4. Bar graph depicting the test accuracies on the Chaoyang dataset for ResNet-34, XCiT, SqueezeNet, and MobileNet models.

The performance of the algorithm when integrated with MobileNet exceeds those when integrated with ResNet-34, SqueezeNet, and XCiT with a test accuracy of 84.39%, as depicted in Fig. 4.MobileNet beats the previous state-of-the-art performance of ResNet-34 by 0.99% It is also observed

that on using MobileNet, in Phase II of the architecture, both M1 and M2 of the co-learning architecture give maximum accuracy and negligible loss.

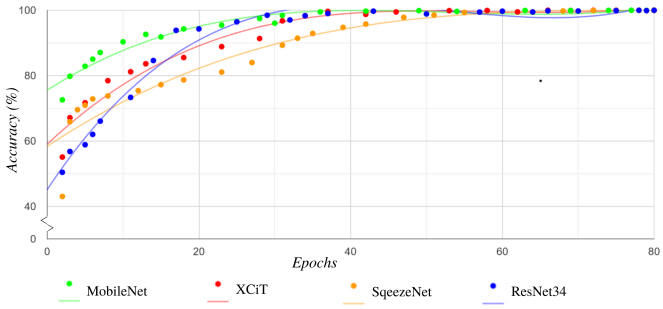

Fig. 5. Plot of accuracy v/s epochs for ResNet-34, XCiT, SqueezeNet, and MobileNet models.

A plot has been curated, Fig. 5, depicting a comparison of accuracy trends among models. Here, the vertical axis represents the accuracy, and the horizontal axis represents the number of epochs. MobileNet outperforms other models in terms of all the evaluation metrics.

VI. Conclusion

This paper presents an analysis of the performance of different deep-learning models integrated with the HSA-NRL algorithm on colon slide samples. The three models analyzed are XCiT, MobileNet, and SqueezeNet and are compared with the ResNet-34, the originally used model in the algorithm. The paper presents a comprehensive performance evaluation based on test accuracy, precision, recall, and F1 score. Model XCiT as the backbone of the network performs reasonably well and achieves comparable values to ResNet-34. The analysis shows that Mobilenet when used in the architecture outperforms other models and has significantly boosted the overall performance from an accuracy of 83.40% to a newly achieved 84.39%.

Additionally, Mobilenet is a lightweight and fast model that uses less computational power and time. This work hence proposes an improved noise robust architecture for colon cancer detection. The obtained results are only based on the models' performances on the Chaoyang dataset, which can be treated as a limitation of our study. Thus, the necessity to apply these models to other medical areas is highlighted for future works.

References

[1] N. Thakur, H. Yoon, and Y. Chong, "Current trends of artificial intelligence for colorectal cancer pathology image analysis: A systematic review," Cancers, vol. 12, no. 7, 2020. [Online]. Available: https://www.mdpi.com/2072-6694/12/7/1884

[2] C. Zhu, W. Chen, T. Peng, Y. Wang, and M. Jin, "Hard sample aware noise robust learning for histopathology image classification," IEEE Transactions on Medical Imaging, vol. 41, no. 4, pp. 881–894, 2021.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[4] Chaoyang University of Technology, "Chaoyang dataset," 2021. [Online]. Available: https://bupt-ai-cz.github.io/HSA-NRL/

[5] A. Ali, H. Touvron, M. Caron, P. Bojanowski, M. Douze, A. Joulin, I. Laptev, N. Neverova, G. Synnaeve, J. Verbeek et al., "Xcit: Cross-covariance image transformers," Advances in neural information processing systems, vol. 34, pp. 20 014–20 027, 2021.

[6] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size," arXiv preprint arXiv:1602.07360, 2016.

[7] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[8] Z. Wang, M. Li, H. Wang, H. Jiang, Y. Yao, H. Zhang, and J. Xin, "Breast cancer detection using extreme learning machine based on feature fusion with cnn deep features," IEEE Access, vol. 7, pp. 105 146– 105 158, 2019.

[9] L. B. Lusted, "Medical electronics," New England Journal of Medicine, vol. 252, no. 14, pp. 580–585, 1955.

[10] Y. Xu, Z. Jia, L.-B. Wang, Y. Ai, F. Zhang, M. Lai, E. I. Chang et al., "Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features," BMC bioinformatics, vol. 18, no. 1, pp. 1–17, 2017.

[11] J. Li, W. Li, A. Sisk, H. Ye, W. D. Wallace, W. Speier, and C. W. Arnold, "A multi-resolution model for histopathology image classification and localization with multiple instance learning," Computers in biology and medicine, vol. 131, p. 104253, 2021.

[12] C. Zhu, F. Song, Y. Wang, H. Dong, Y. Guo, and J. Liu, "Breast cancer histopathology image classification through assembling multiple compact cnns," BMC medical informatics and decision making, vol. 19, no. 1, pp. 1–17, 2019.

[13] J. C. Caicedo, A. Cruz, and F. A. Gonzalez, "Histopathology image classification using bag of features and kernel functions," in Artificial Intelligence in Medicine: 12th Conference on Artificial Intelligence in Medicine, AIME 2009, Verona, Italy, July 18-22, 2009. Proceedings 12. Springer, 2009, pp. 126–135.

[14] Y. Xue, J. Ye, Q. Zhou, L. R. Long, S. Antani, Z. Xue, C. Cornwell, R. Zaino, K. C. Cheng, and X. Huang, "Selective synthetic augmentation with histogan for improved histopathology image classification," Medical image analysis, vol. 67, p. 101816, 2021.

[15] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyo, A. L. Moreira, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," Nature medicine, vol. 24, no. 10, pp. 1559–1567, 2018.

[16] M. Masud, N. Sikder, A.-A. Nahid, A. K. Bairagi, and M. A. AlZain, "A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework," Sensors, vol. 21, no. 3, p. 748, 2021.

[17] M. A. Rahman and R. C. Muniyandi, "Feature selection from colon cancer dataset for cancer classification using artificial neural network," International Journal on Advanced Science, Engineering and Information Technology, vol. 8, no. 4-2, pp. 1387–1393, 2018.

[18] N. Gessert, M. Bengs, L. Wittig, D. Dromann, T. Keck, A. Schlaefer, and D. B. Ellebrecht, "Deep transfer learning methods for colon cancer classification in confocal laser microscopy images," International journal of computer assisted radiology and surgery, vol. 14, pp. 1837–1845, 2019.

[19] M. M. Behzadi, M. Madani, H. Wang, J. Bai, A. Bhardwaj, A. Tarakanova, H. Yamase, G. H. Nam, and S. Nabavi, "Weakly-supervised deep learning model for prostate cancer diagnosis and gleason grading of histopathology images," arXiv preprint arXiv:2212.12844, 2022.

[20] Y. Wei, C. Gong, S. Chen, T. Liu, J. Yang, and D. Tao, "Harnessing side information for classification under label noise," IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 9, pp. 3178–3192, 2019.

[21] Z. Hameed, S. Zahia, B. Garcia-Zapirain, J. Javier Aguirre, and A. Maria Vanegas, "Breast cancer histopathology image classification using an ensemble of deep learning models," Sensors, vol. 20, no. 16, p. 4373, 2020.

[22] G. Patrini, A. Rozza, A. Krishna Menon, R. Nock, and L. Qu, "Making deep neural networks robust to label noise: A loss correction approach," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1944–1952.

[23] S. Gehlot, A. Gupta, and R. Gupta, "A cnn-based unified framework utilizing projection loss in unison with label noise handling for multiple

myeloma cancer diagnosis," Medical Image Analysis, vol. 72, p. 102099, 2021.
[24] Y. Wang, X. Ma, Z. Chen, Y. Luo, J. Yi, and J. Bailey, "Symmetric cross entropy for robust learning with noisy labels," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 322–330.
[25] J. Li, R. Socher, and S. C. Hoi, "Dividemix: Learning with noisy labels as semi-supervised learning," arXiv preprint arXiv:2002.07394, 2020.
[26] D. Tanaka, D. Ikami, T. Yamasaki, and K. Aizawa, "Joint optimization framework for learning with noisy labels," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 5552–5560.
[27] K. Yi and J. Wu, "Probabilistic end-to-end noise correction for learning with noisy labels," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7017–7025.
[28] J. Goldberger and E. Ben-Reuven, "Training deep neural-networks using a noise adaptation layer," in International conference on learning representations, 2017.
[29] P. Chen, B. B. Liao, G. Chen, and S. Zhang, "Understanding and utilizing deep neural networks trained with noisy labels," in International Conference on Machine Learning. PMLR, 2019, pp. 1062–1070