Classification of prostate histological images using Vision Transformers: an analysis with stain normalization and ensemble learning

Betânia Caroline Silva de Albuquerque
Federal University of São Paulo
Email: betania.caroline@unifesp.br

Hanna Beatriz Couto Monteiro Fernandes de Castro
Federal University of São Paulo
Email: h.castro01@unifesp.br

Leandro Alves Neves
São Paulo State University
Email: nevesleandro@gmail.com

Marcelo Zanchetta do Nascimento
Federal University of Uberlândia
Email: marcelozanchetta@gmail.com

Thaína Aparecida Azevedo Tosta
Federal University of São Paulo
Email: tosta.thaina@gmail.com

***Abstract*—Prostate cancer is the second most common cancer in men worldwide, diagnosed via histopathological evaluation of H&E-stained images. Gleason grading, however, is subjective and prone to inter-observer variability. Deep learning-based computer-aided diagnosis systems offer promising support, but stain color variations pose a challenge, motivating normalization algorithms. This study evaluates color normalization on H&E prostate cancer image classification using a pre-trained ViT and eight classifiers, including a majority voting ensemble. Binary classification on a public dataset compared benign and malignant cases across two normalization methods (SW-CCN and BKSVD) and original images. Results showed original images yielded superior ViT and classifier performance, despite more malignant cases being misclassified as benign. SVM with ViT feature extraction achieved the best overall performance, surpassing both the ensemble and ViT classifier.**

I. Introduction

Cancer is a prevalent health condition, with one in five individuals at risk during their lifetime [1]. Prostate cancer (PCa) is the second most common cancer in men worldwide, after lung cancer [2], [3]. In Brazil, it is the most frequent cancer in men, with 71,730 new cases estimated for 2023–2025 [4].

Diagnosis relies on microscopic examination of H&E-stained biopsies, where hematoxylin stains nuclei purple and eosin stains extracellular components pink [5]. The Gleason grading system is the gold standard for these diagnoses, but subjectivity leads to significant inter- and intra-observer variability [6]. Computer-aided diagnosis (CAD) with AI shows potential in PCa detection and prognosis [7]. Machine learning (ML), a subfield of AI, includes deep learning (DL), which can automatically learn abstract features using neural networks [8], [9]. CNNs and Vision Transformers (ViTs) are promising DL models for feature extraction and classification in medical imaging [10].

Color variations in H&E images challenge CAD systems and can be mitigated with stain normalization [11]. Prior studies on normalized H&E images used mainly CNNs, often showing positive impacts on classification [12]–[21], though some reported weak or negative effects [22]–[24]. Classification studies explored ML algorithms like decision tree (DT), Naive Bayes (NB), K-nearest neighbors (NN) [24], random forest (RF) [21], [23], support vector machine (SVM) [20], [21], [23], and softmax activation [16]. However, no study applied ensemble methods or combined all these classifiers in an analysis.

Also in this context, none of the studies from the correlated literature have analyzed the use of ViT architecture for feature extraction and classification of normalized histopathological images; instead, their application has been mostly limited to CNNs, particularly for prostate cancer images. Therefore, there are still open questions regarding the use of ViTs as well as different classification approaches, including several ML algorithms and ensembles, since most studies analyzed up to two kinds of ML algorithms and none of them studied ensemble methods. This work proposes the analysis of the role of normalization in the training and binary classification of PCa histological images, which has not been performed in classification studies, using a pre-trained ViT and eight classifiers: the ViT classifier, logistic regression (LR), K-NN, NB, Polynomial, RF, SVM, and an ensemble of them, except for the ViT classifier. The main contributions are: (i) analysis of ViT with original and normalized images, (ii) investigation of two normalization techniques on the classification results, and (iii) investigation of different classifiers and their ensemble.

II. Metodology

This work involved three steps. First, input images were normalized using two literature-based techniques. Next, features were extracted with a fine-tuned ViT. Finally, these features were classified using various ML algorithms, including an ensemble. This combination of ViT and normalization has not been previously analyzed. Fig. 1 illustrates the methodology.

Experiments were performed on a desktop with an Intel Ultra 7 265 processor, NVIDIA GeForce RTX 4060, and 32 GB RAM.

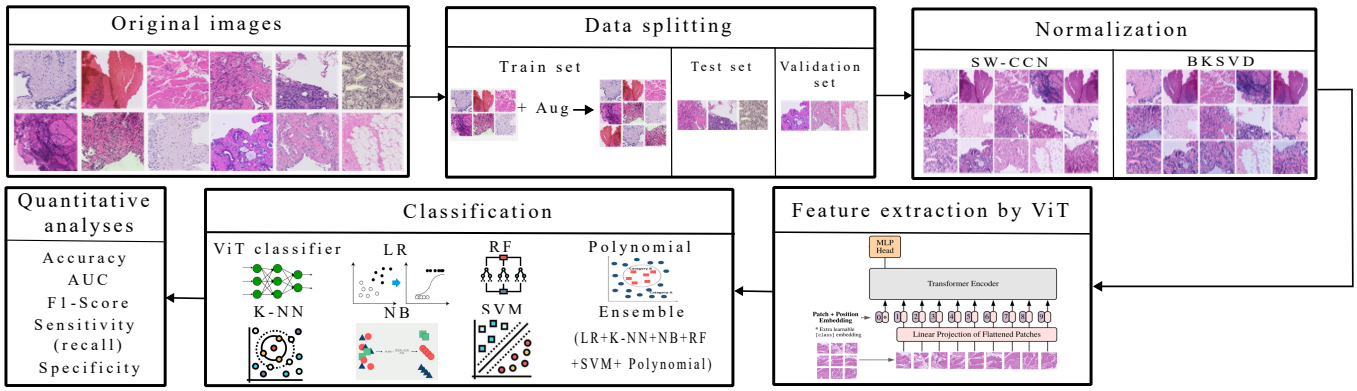

Fig. 1. Flowchart of the proposed methodology for evaluation of normalization applied to classification of PCa images with ensemble learning and ViT. Figure adapted from [25] © 2020 arXiv.

A. Dataset

The images used in this study were obtained from the CrowdGleason dataset, which comprises 19,077 H&E PCa patches from 1,045 whole-slide images with a resolution of 2,048×2,048 pixels and a magnification of 40×, with no overlapping. Pathologists categorized them into four labels: non-cancerous (NC), Gleason Grade 3 (G3), Gleason Grade 4 (G4), or Gleason Grade 5 (G5) [26]. Some examples are shown in Fig. 1.

For balanced analysis, we used binary classification: NC as benign and G3–G5 as malignant. Data sampling followed [26], with 13,824 training, 2,327 validation, and 2,926 testing samples, ensuring no patient overlap. Even after binarization, the training set had 2,124 more benign samples. To balance it, 2,124 malignant samples were randomly selected and augmented following [21]. Test and validation sets remained unbalanced: test with 2,157 benign cases and 769 malignants; validation with 1,329 benign cases and 998 malignants.

B. Stain color normalization

The saturation-weighted complete color normalization (SW-CCN) proposed by Li et al. [27] is a normalization method that addresses the main sources of color variation: illumination inconsistencies and staining variability, separately. It uses a non-negative matrix factorization to estimate stain colors in the RGB color space. In a similar approach, the method in [17] applied a dictionary learning to transfer the colors of a target image to all the PCa images. This method was entitled the Bayesian K-singular value decomposition (BKSVD). The SW-CCN was chosen for reaching the best results in recent segmentation analyses [28]. In contrast, the normalized images by BKSVD were made available with the evaluated dataset, which represents a bias-free dataset for analysis.

C. Vision Transformer

The ViT is a DL architecture for image recognition, derived from the transformer model, originally successful in NLP. Input images are divided into fixed-size patches, flattened into vectors (patch embeddings), and enriched with learnable positional embeddings to preserve spatial information. A learnable classification token (CLS) is prepended as a global representation. This sequence passes through a transformer encoder [25].

ViT can be pre-trained on large datasets and fine-tuned for specific ones [25]. We used a pre-trained ViT-Large (16×16 patch, 224×224 input) from Google Research via Hugging Face, originally trained on ImageNet-21k and fine-tuned on ImageNet-1k. Further fine-tuning on CrowdGleason used a learning rate of 2e-4, weight decay 0.01, batch size 16, and 50 epochs, evaluating validation accuracy for each epoch, from which the best model was used.

For feature extraction, the fine-tuned ViT provided a 1,024-dimensional vector from the final hidden state of the CLS token.

D. Classification

For ViT fine-tuning, the classification head consisted of an MLP [25]. Several ML classifiers were also applied. K-NN classifies a sample based on the majority class among its K nearest neighbors [29]. NB uses Bayes' theorem to estimate event probabilities from prior knowledge [30]. LR estimates outcome probabilities from predictors, often used in diagnostic analyses [31]. RF combines decision trees, with leaf nodes providing class probabilities and internal nodes splitting data [32]. SVM separates data with an optimal hyperplane; polynomial refers to SVM with a polynomial kernel, while SVM here uses an RBF kernel [33]. An ensemble combines predictions from multiple classifiers to improve performance, with common approaches including boosting, bagging, and stacking [34]. Here, a majority voting was used, combining K-NN (K=5), LR, NB, polynomial, RF, and SVM, all trained in parallel via scikit-learn.

E. Evaluation Metrics

The metrics used in this work are shown in Fig. 1. Accuracy reflects the proportion of correctly classified instances. The area under the ROC curve (AUC) quantifies a classifier's

ability to discriminate between classes across thresholds. F1-score (a harmonic mean of precision and recall) balances false positives and negatives. Sensitivity (recall) indicates the proportion of true positives correctly identified, while specificity measures true negatives accurately detected. Together, these metrics provide a comprehensive evaluation of a classifier's performance [35].

III. Results and Discussion

This section presents the results of the proposed methodology. To do so, initially, ViT was analyzed for processing the original and normalized images. Thereafter, the investigation of the different classifiers is detailed.

A. Analyses of normalization combined with ViT

Table I presents the classification metrics of ViT fine-tuned on the three PCa datasets: original and normalized versions (SW-CNN [27] and BKSVD [17]). Original data outperformed normalized data across most metrics, except sensitivity, highlighted in bold. Among normalization methods, the maximum metric difference was 0.1224 in sensitivity. Accuracy showed the largest gap, with original data exceeding normalized results by up to 14.26%.

These results indicate that normalization did not improve ViT classification. Notably, normalization increased sensitivity by 0.2368, enhancing the true positive rate. Overall, the F1-score and sensitivity remained low, with maximum values of 0.60 and 0.66, respectively, except for the BKSVD model (0.78 for sensitivity), suggesting that other classifiers could be explored for better performance.

TABLE I
PERFORMANCE COMPARISON OF VIT ACROSS THE ORIGINAL AND NORMALIZED DATASETS THROUGH ACCURACY (ACC), AUC, F1-SCORE, SENSITIVITY (SEN), AND SPECIFICITY (SP).

Dataset	Acc	AUC	F1-score	Sen	Sp
Original	**0.8100**	**0.8361**	**0.6040**	0.5514	**0.9022**
SW-CCN	0.6945	0.7405	0.5339	0.6658	0.7047
BKSVD	0.6941	0.7908	0.5558	**0.7882**	0.6820

Fig. 2 shows ViT confusion matrices: SW-CCN and BKSVD misclassified 637 and 686 benign samples as malignant, respectively, versus 211 for the original data. Conversely, original images had the most malignant samples misclassified as benign (345). Thus, normalized images made benign predictions harder but reduced the misclassification of malignant samples. Since medical applications can have different perspectives for misclassification, the original images imposed more complex and flawed results, identifying more malignant cases as benign.

B. Classification of ViT features through different methods

Table II shows ML classifiers' performance on the datasets after feature extraction with fine-tuned ViT. The dataset normalized with SW-CCN had the lowest results across classifiers, with maximum accuracy, F1-score, sensitivity, and specificity of 0.68, 0.63, 0.71, and 0.69, respectively. Original images achieved the best overall performance, though BKSVD was similar. For LR, RF, and SVM, the maximum metric difference was 0.04; other classifiers showed larger differences, with K-NN having the highest specificity gap (0.17).

Among all classifiers, SVM obtained the best metrics, except for sensitivity, reaching 0.77 accuracy, 0.64 F1-score, and 0.77 specificity. The K-NN achieved the highest sensitivity (0.80) for original and BKSVD images. Normalization did not improve results; in fact, performance was equal or lower for normalized datasets compared to the original, except for the BKSVD method, wich improved RF accuracy and specificity. Overall, ViT classifier outperformed ML classifiers on original images. For normalized datasets, BKSVD with LR, NB, RF, SVM, and polynomial classifiers achieved up to 8% higher accuracy than ViT. These classifiers also increased overall F1-score and sensitivity, notably SVM, which reached 0.64 F1-score and 0.77 sensitivity for the original images, compared to 0.60 and 0.55 for ViT on the same data.

TABLE II
PERFORMANCE OF CLASSIFIERS ACROSS THE DIFFERENT DATASETS.

Classifier	Dataset	Accuracy	F1-score	Sensitivity	Specificity
K-NN	Original	**0.71**	**0.59**	**0.80**	**0.69**
	SW-CCN	0.67	0.51	0.65	0.67
	BKSVD	0.60	0.51	**0.80**	0.52
LR	Original	**0.75**	**0.62**	**0.78**	**0.73**
	SW-CCN	0.68	0.53	0.71	0.66
	BKSVD	0.74	0.61	0.77	**0.73**
NB	Original	**0.75**	**0.61**	**0.75**	**0.75**
	SW-CCN	0.67	0.51	0.63	0.69
	BKSVD	0.72	0.56	0.67	0.74
RF	Original	0.75	**0.61**	**0.76**	0.74
	SW-CCN	0.66	0.52	0.71	0.64
	BKSVD	**0.76**	**0.61**	0.74	**0.76**
SVM	Original	**0.77**	**0.64**	**0.77**	**0.77**
	SW-CCN	0.68	0.63	0.70	0.67
	BKSVD	0.75	0.60	0.74	0.75
Polynomial	Original	**0.75**	**0.63**	**0.79**	**0.74**
	SW-CCN	0.67	0.51	0.69	0.67
	BKSVD	0.73	0.59	0.74	0.72

Fig. 3 shows the AUC variations. The original images achieved the highest AUCs for all classifiers, exceeding 0.775. SW-CCN had the lowest AUCs, with the best around 0.750, while BKSVD showed more variation, ranging from nearly 0.675 to 0.825.

Analysis of the classifiers highlights the high performance of LR, RF, SVM, and polynomial, all above the mean indicated by the dotted lines. K-NN and NB consistently had lower AUCs across all datasets. Overall, despite the AUC variations, classifier performance was similar between original and normalized datasets.

C. Ensemble performance analyses

Table III presents the ensemble results. Original dataset achieved the best results, confirming that normalization did not improve classification. Ensemble outperformed some individual ML algorithms, e.g. K-NN (up to 8% lower accuracy). For F1-score and specificity, the ensemble surpassed nearly all ML algorithms except SVM, which had a higher F1-score (0.64 original, 0.63 SW-CCN, 0.60 BKSVD). Sensitivity was up to

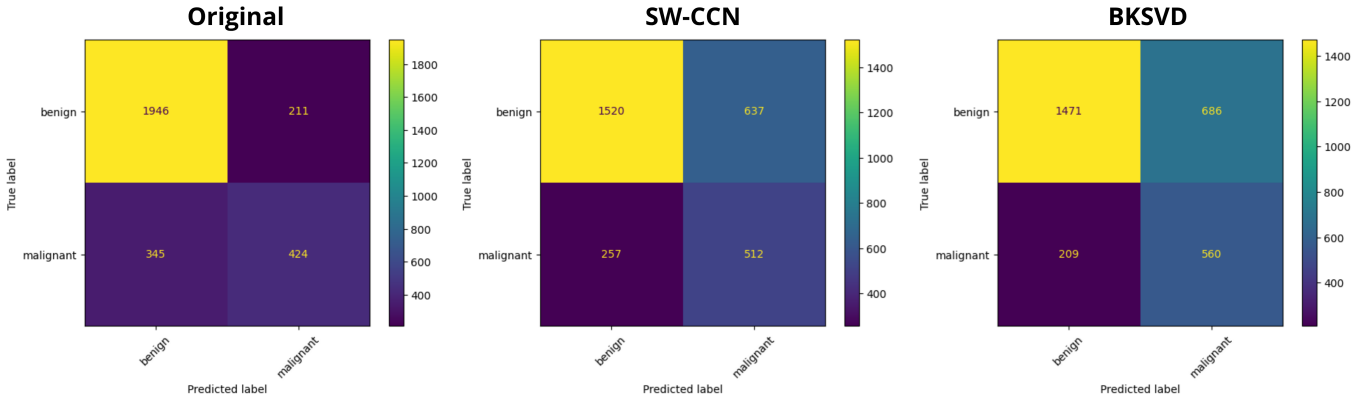

Fig. 2. Confusion matrices for ViT classification on test sets from each dataset (original and normalized by SW-CNN and BKSVD).

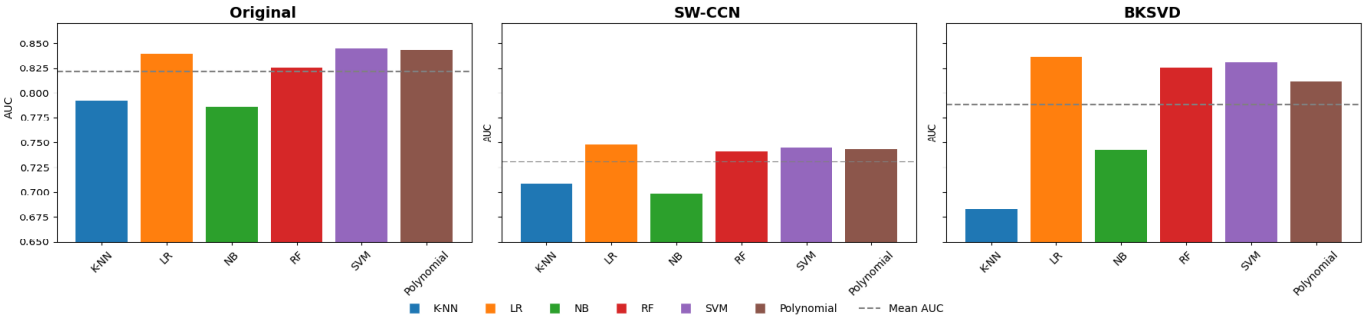

Fig. 3. Comparison of AUC results of different classifiers on the three analyzed datasets.

4% better in individual classifiers. Therefore, the ensemble had higher F1-score and specificity, but lower sensitivity.

TABLE III
PERFORMANCE COMPARISON OF ENSEMBLE ACROSS THE ORIGINAL AND NORMALIZED DATASETS THROUGH ACCURACY (ACC), AUC, F1-SCORE, SENSITIVITY (SEN), AND SPECIFICITY (SP).

Dataset	Acc	AUC	F1-score	Sen	Sp
Original	**0.7717**	**0.8419**	**0.6381**	**0.7659**	**0.7738**
SW-CCN	0.6887	0.7416	0.5340	0.6788	0.6922
BKSVD	0.7447	0.8108	0.5990	0.7256	0.7515

The original dataset had the best ensemble performance (AUC 0.8419), followed by BKSVD (0.8108) and SW-CCN lowest (0.7416), which was expected since the individual ML classifiers also achieved this result. Despite normalization decreasing performance, all results exceeded a random baseline (AUC 0.50). SVM AUCs were 0.8450, 0.7449, 0.8307 (original, SW-CCN, BKSVD), showing similar but better performance than the ensemble. Therefore, it is interesting to note that even by combining different classifiers, we have not yet obtained better AUC results. This is an important insight because the original images did not benefit from the ensemble learning, and the limitations imposed by the normalized images' features were not surpassed by this approach.

An important observation is that the PCa dataset used in this work has not yet been used in any classification study, presenting a challenge regarding images that have not been analyzed by other authors.

IV. CONCLUSION

In this work, we investigated the impact of two normalization methods (SW-CCN and BKSVD) on a pre-trained ViT for classifying H&E-stained histological PCa images, a dataset not previously studied for classification. We also analyzed various machine learning algorithms: NB, K-NN, LR, polynomial, RF, SVM, and a majority-voting ensemble. Our results revealed that these normalization methods can reduce the performance of both the ViT and classifiers using ViT-extracted features. This can be a consequence of challenges imposed by the normalization on benign cases' features. Notably, SVM consistently achieved the best metrics among the classifiers, indicating that ViT feature extraction from original images with SVM alone is sufficient for this classification. For further research, the present work provides a foundation for exploring other ViT models, especially hybrid architectures, as well as for conducting studies on different databases.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the National Council for Scientific and Technological Development - CNPq (Grants #313643/2021-0, #311404/2021-9 and #307318/2022-2) and the State of Minas Gerais Research

Foundation - FAPEMIG (Grant #APQ-00727-24). This study was financed, in part, by the São Paulo Research Foundation (FAPESP), Brazil. Processes Numbers #2022/03020-1, #2023/15834-6, and #2025/05775-8.

REFERENCES

[1] J. Ferlay, M. Colombet, I. Soerjomataram, D. M. Parkin, M. Piñeros, A. Znaor, and F. Bray, "Cancer statistics for the year 2020: An overview," *International journal of cancer*, vol. 149, no. 4, pp. 778–789, 2021.

[2] J. Ferlay, M. Ervik, F. Lam, M. Colombet, L. Mery, M. Piñeros, A. Znaor, I. Soerjomataram, and F. Bray, "Global cancer observatory: cancer today," *Lyon: International agency for research on cancer*, vol. 20182020, 2020.

[3] P. Rawla, "Epidemiology of prostate cancer," *World journal of oncology*, vol. 10, no. 2, p. 63, 2019.

[4] Instituto Nacional de Câncer (Brasil), *Estimativa 2023: incidência de câncer no Brasil*. Rio de Janeiro: INCA, 2022, il. color.

[5] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE reviews in biomedical engineering*, vol. 2, pp. 147–171, 2009.

[6] W. C. Allsbrook Jr, K. A. Mangold, M. H. Johnson, R. B. Lane, C. G. Lane, and J. I. Epstein, "Interobserver reproducibility of gleason grading of prostatic carcinoma: general pathologist," *Human pathology*, vol. 32, no. 1, pp. 81–88, 2001.

[7] S. M. Ayyad, N. B. Abdel-Hamid, H. A. Ali, and L. M. Labib, "Multimodality imaging in prostate cancer diagnosis using artificial intelligence: basic concepts and current state-of-the-art," *Multimedia Tools and Applications*, pp. 1–30, 2025.

[8] J. J. Twilt, K. G. van Leeuwen, H. J. Huisman, J. J. Fütterer, and M. de Rooij, "Artificial intelligence based algorithms for prostate cancer classification and detection on magnetic resonance imaging: a narrative review," *Diagnostics*, vol. 11, no. 6, p. 959, 2021.

[9] S. Wang, K. Burtt, B. Turkbey, P. Choyke, and R. M. Summers, "Computer aided-diagnosis of prostate cancer on multiparametric mri: a technical review of current research," *BioMed research international*, vol. 2014, no. 1, p. 789561, 2014.

[10] X. Jiang, Z. Hu, S. Wang, and Y. Zhang, "Deep learning for medical image-based cancer diagnosis. cancers, 15 (14), 3608," 2023.

[11] R. A. Hoffman, S. Kothari, and M. D. Wang, "Comparison of normalization algorithms for cross-batch color segmentation of histopathological images," in *2014 36th annual international conference of the IEEE engineering in medicine and biology society*. IEEE, 2014, pp. 194–197.

[12] A. Sethi, L. Sha, A. R. Vahadane, R. J. Deaton, N. Kumar, V. Macias, and P. H. Gann, "Empirical comparison of color normalization methods for epithelial-stromal classification in h and e images," *Journal of pathology informatics*, vol. 7, no. 1, p. 17, 2016.

[13] F. Ciompi, O. Geessink, B. E. Bejnordi, G. S. De Souza, A. Baidoshvili, G. Litjens, B. Van Ginneken, I. Nagtegaal, and J. Van Der Laak, "The importance of stain normalization in colorectal tissue classification with convolutional networks," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 160–163.

[14] C. Hu, X. Sun, Z. Yuan, and Y. Wu, "Classification of breast cancer histopathological image with deep residual learning," *International Journal of Imaging Systems and Technology*, vol. 31, no. 3, pp. 1583–1594, 2021.

[15] M. Salvi, F. Molinari, U. R. Acharya, L. Molinaro, and K. M. Meiburger, "Impact of stain normalization and patch selection on the performance of convolutional neural networks in histological breast and prostate cancer classification," *Computer Methods and Programs in Biomedicine Update*, vol. 1, p. 100004, 2021.

[16] J. Boschman, H. Farahani, A. Darbandsari, P. Ahmadvand, A. Van Spankeren, D. Farnell, A. B. Levine, J. R. Naso, A. Churg, S. J. Jones *et al.*, "The utility of color normalization for ai-based diagnosis of hematoxylin and eosin-stained pathology images," *The Journal of Pathology*, vol. 256, no. 1, pp. 15–24, 2022.

[17] F. Pérez-Bueno, J. G. Serra, M. Vega, J. Mateos, R. Molina, and A. K. Katsaggelos, "Bayesian k-svd for h and e blind color deconvolution. applications to stain normalization, data augmentation and cancer classification," *Computerized Medical Imaging and Graphics*, vol. 97, p. 102048, 2022.

[18] R. Bazargani, W. Chen, S. Sadeghian, M. Asadi, J. Boschman, A. Darbandsari, A. Bashashati, and S. Salcudean, "A novel h&e color augmentation for domain invariance classification of unannotated histopathology prostate cancer images," in *Medical Imaging 2023: Digital and Computational Pathology*, vol. 12471. SPIE, 2023, pp. 224–229.

[19] T. A. A. Tosta, A. D. Freitas, P. R. de Faria, L. A. Neves, A. S. Martins, and M. Z. do Nascimento, "A stain color normalization with robust dictionary learning for breast cancer histological images processing," *Biomedical Signal Processing and Control*, vol. 85, p. 104978, 2023.

[20] N. Altini, T. M. Marvulli, F. A. Zito, M. Caputo, S. Tommasi, A. Azzariti, A. Brunetti, B. Prencipe, E. Mattioli, S. De Summa *et al.*, "The role of unpaired image-to-image translation for stain color normalization in colorectal cancer histology classification," *Computer Methods and Programs in Biomedicine*, vol. 234, p. 107511, 2023.

[21] V. B. Fernandes, A. B. Silva, D. C. Pereira, S. V. Cardoso, P. R. de Faria, A. M. Loyola, T. A. Tosta, L. A. Neves, and M. Z. do Nascimento, "Investigation of deep neural network compression based on tucker decomposition for the classification of lesions in cavity oral," *Proceedings Copyright*, vol. 516, p. 523, 2024.

[22] G. Lee, M. Bajger, and K. Clark, "Deep learning and color variability in breast cancer histopathological images: a preliminary study," in *14th International Workshop on Breast Imaging (IWBI 2018)*, vol. 10718. SPIE, 2018, pp. 370–375.

[23] A. Kumar, S. K. Singh, S. Saxena, K. Lakshmanan, A. K. Sangaiah, H. Chauhan, S. Shrivastava, and R. K. Singh, "Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer," *Information Sciences*, vol. 508, pp. 405–421, 2020.

[24] F. Bianconi, J. N. Kather, and C. C. Reyes-Aldasoro, "Experimental assessment of color deconvolution and color normalization for automated classification of histology images stained with hematoxylin and eosin," *Cancers*, vol. 12, no. 11, p. 3337, 2020.

[25] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[26] M. López-Pérez, A. Morquecho, A. Schmidt, F. Pérez-Bueno, A. Martín-Castro, J. Mateos, and R. Molina, "The crowdgleason dataset: Learning the gleason grade from crowds and experts," *Computer Methods and Programs in Biomedicine*, vol. 257, p. 108472, 2024.

[27] X. Li and K. N. Plataniotis, "A complete color normalization approach to histopathology images using color cues computed from saturation-weighted statistics," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7, pp. 1862–1873, 2015.

[28] A. F. Q. da Silva, A. D. Freitas, P. R. de Faria, L. A. Neves, M. Z. Do Nascimento, and T. A. A. Tosta, "Color normalization by dictionary learning with nuclear segmentation evaluation in he histological images," in *38th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2025.

[29] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," in *2019 international conference on intelligent computing and control systems (ICCS)*. IEEE, 2019, pp. 1255–1260.

[30] T. M. Mitchell and T. M. Mitchell, *Machine learning*. McGraw-hill New York, 1997, vol. 1, no. 9.

[31] T. G. Nick and K. M. Campbell, "Logistic regression," *Topics in biostatistics*, pp. 273–301, 2007.

[32] W. Chen, X. Xie, J. Wang, B. Pradhan, H. Hong, D. T. Bui, Z. Duan, and J. Ma, "A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility," *Catena*, vol. 151, pp. 147–160, 2017.

[33] E. Y. Boateng, J. Otoo, and D. A. Abaye, "Basic tenets of classification algorithms k-nearest-neighbor, support vector machine, random forest and neural network: A review," *Journal of Data Analysis and Information Processing*, vol. 8, no. 4, pp. 341–357, 2020.

[34] Z.-H. Zhou and Z.-H. Zhou, *Ensemble learning*. Springer, 2021.

[35] T. Hastie, R. Tibshirani, J. Friedman *et al.*, "The elements of statistical learning," 2009.