International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Ensemble Machine Learning based Enhanced Detection of Lymph Node Cancer Through Histopathology Images with Synergistic Fusion of Thepade SBTC and Color Binning Features

Sudeep D. Thepade*[1], Ashish A. Gore[2], Rachana Y. Patil [3]

Submitted: 29/01/2024 **Revised**: 07/03/2024 **Accepted**: 15/03/2024

Abstract: Histopathology cancer detection plays a crucial part in primordial diagnosis and treatment planning. This research introduces a novel approach combining the features formed with color binning and Thepade Sorted Block Truncation Coding (Thepade SBTC) methods for cancer detection from histopathology images using machine learning (ML) algorithms. The proposed feature fusion technique leverages the strengths of Thepade SBTC, enhancing its performance through the integration of color binning. The study evaluates the efficacy of these features across a range of dimensions, from 2-bin to 15-bin in color binning and 2-ary to 15-ary in Thepade SBTC, shedding light on the impact of feature granularity on classification accuracy.

The research compares the accuracy of suggested feature fusion-based cancer detection with the features using Thepade SBTC and color binning individually, for various ML algorithms such as Logistic Regression, XGBoost, K-Nearest Neighbors(KNN), Decision Tree(DT), CatBoost, Gaussian Naive Bayes(NB), Random Forest(RF), AdaBoost, and Neural Network. Results indicate that KNN, Random Forest, and CatBoost consistently outperform other algorithms in both Thepade SBTC and color binning scenarios.

Motivated by these findings, a novel ensemble approach is proposed, utilizing a Voting Classifier with KNN, Random Forest, and CatBoost as base estimators. The ensemble model achieves an impressive 90% accuracy by cascading features extracted from 6-bin color binning and 14-ary Thepade SBTC. This highlights the potential for synergistic combinations of feature extraction methods and ensemble learning in Histopathology cancer detection.

Keywords: *Thepade Sorted Block Truncation Coding (Thepade SBTC), Color Binning, Histopathology Cancer Detection, Machine Learning, Ensemble Learning, Logistic regression, AdaBoost classifier, K-nearest neighbors (KNN), XGBoost, Decision trees, CatBoost, Gaussian Naive Bayes classifier, Random forests, Neural networks*

1. Introduction

Histopathology cancer detection, a pivotal field in medical imaging, has an important contribution in early diagnosis and treatment of various malignancies. Accurate identification of pathological patterns in tissue samples is imperative for ensuring effective patient care and outcomes. In recent years, advancements in image processing and machine learning techniques have propelled the development of innovative methodologies aimed at enhancing the precision and efficiency of cancer detection[1].

This paper introduces a novel approach that combines two distinct yet complementary feature extraction methods, namely Thepade SBTC [2][3][4][5] and color binning[6][7], to elevate the accuracy of Histopathology cancer detection. Thepade SBTC, an extension of the well-established Block Truncation Coding (BTC), introduces a sorting mechanism based on block size before compression, optimizing the representation of smaller blocks and enhancing compression efficiency. The sorting mechanism exploits the inherent characteristics of smaller blocks, which tend to contain less intricate details, leading to improved feature extraction[2].

The integration of color binning, a method that classifies pixel values into bins based on their intensity, provides an additional dimension to feature representation [7]. This hybrid approach seeks to leverage both structural information captured by Thepade SBTC and color-related details identified through binning. By examining the impact of varying feature granularity, ranging from 2-ary to 15-ary for Thepade SBTC and 2-bin to 15-bin for color binning, the research aims to identify optimal feature sets that enhance the discrimination of cancerous tissue in Histopathology images.

The significance of this research lies in its comprehensive evaluation of ML algorithms applied to the features generated from Thepade SBTC and color binning. Nine algorithms, including Logistic Regression, XGBoost, K-Nearest Neighbors(KNN), Decision Tree(DT), CatBoost, Gaussian Naive Bayes(NB), Random Forest(RF), AdaBoost, and Neural Network, are scrutinized to assess their classification performance. The consistent superiority of KNN, Random Forest, and CatBoost emerges,

[1] Pimpri Chinchwad College of Engineering-Pune, India
ORCID ID : 0000-0001-7809-4148
[2] Pimpri Chinchwad College of Engineering-Pune, India
[3] Pimpri Chinchwad College of Engineering-Pune, India
** Corresponding Author Email: sudeepthepade@gmail.com*

underscoring their potential as effective classifiers for Histopathology cancer detection.

Motivated by these promising results, the paper proposes a novel ensemble learning approach, employing a Voting Classifier with KNN, Random Forest, and CatBoost as base estimators. This ensemble model capitalizes on the complementary strengths of individual classifiers, achieving a remarkable 90% accuracy. The cascading of features extracted from 6-bin color binning and 14-ary Thepade SBTC further exemplifies the potential for synergistic combinations in elevating cancer detection accuracy[8].

This paper introduces a novel approach combining Thepade SBTC and color binning to improve Histopathology cancer detection accuracy, evaluating nine ml algorithms and proposing an ensemble learning approach for enhanced classification performance.

The contributions of the work presented are listed as.

- Feature fusion of Thepade SBTC and color binning for improved Lymph node cancer detection through histopathology images.
- Performance analysis of proposed feature fusion using PatchCamelyon dataset with Ensemble ML algorithms.

In section 2. Literature Review along with the pivotal role of histopathology cancer detection, recent advancements in image processing and machine learning techniques, and existing methodologies' limitations are reviewed. In section 3, the Proposed Method is introduced with a novel approach combining Thepade SBTC and color binning for enhanced feature extraction and representation in histopathology cancer detection. In section 4, Results evaluated with nine ML algorithms are shown with the identification of KNN, Random Forest, and CatBoost as the superior classifiers, and proposed an ensemble learning approach achieving 90% accuracy. In section 5, the Conclusion and Future Work are highlighted.

1.1. Abbreviations and Acronyms

TSBTC=Thepade Sorted Block Truncation Coding

KNN=K-Nearest Neighbors

BTC=Block Truncation Coding

SBTC=Sorted Block Truncation Coding

MLP=Multi-layer Perceptron

2. Literature Review

Histopathology cancer detection has been the subject of extensive research, focusing on refining methodologies for higher accuracy and reliability. This section reviews key contributions and methodologies in the field, emphasizing the relevance and differentiation of the proposed approach.

2.1. Block Truncation Coding (BTC) and Variants:

BTC [9][10] has been widely explored as a lossy image compression technique. BTC splits the image into blocks and utilizes a quantizer to lower the count of gray levels within every block while preserving essential statistical features. Several variants, such as Sorted Block Truncation Coding (SBTC), have been introduced to enhance compression efficiency by sorting blocks based on specific criteria. The presented work builds upon these concepts, particularly Thepade SBTC, which introduces sorting based on block size before applying BTC.

2.2. Color Binning Techniques:

Color binning methods have been extensively employed for feature extraction in various image processing applications[6][7]. These techniques classify pixel values into bins based on their intensity levels, providing a compact representation of color information. While traditional color binning focuses on color quantization, the proposed approach integrates this method with Thepade SBTC to capture both structural and color-related features for improved discrimination in Histopathology images.

2.3. Machine Learning for Histopathology Image Analysis:

Ml algorithms have shown promising success in Histopathology image analysis, particularly in cancer detection tasks. Logistic Regression, XGBoost, K-Nearest Neighbors(KNN), Decision Tree(DT), CatBoost, Gaussian Naive Bayes(NB), Random Forest(RF), AdaBoost, and Neural Network been widely explored[11]. This work aligns with this trend, employing these algorithms to evaluate the effectiveness of Thepade SBTC and color binning features in Histopathology cancer detection.

2.4. Ensemble Learning Approaches:

Ensemble learning has gained prominence for improving classification performance[12] by combining multiple base models. Existing research has explored various ensemble methods, such as Voting Classifiers, Bagging, and Boosting. The presented work introduces a novel ensemble approach by combining KNN, Random Forest, and CatBoost as base estimators, showcasing the potential for synergistic combinations of classifiers in enhancing cancer detection accuracy.

2.5. Feature Integration in Medical Imaging:

The integration of multiple feature extraction techniques for medical image analysis has been explored in the literature. Studies have demonstrated that combining complementary features from different methods can enhance the robustness and discriminative power of classifiers[13][14]. The proposed feature combination method, combining Thepade SBTC and color binning, contributes to this line of research by offering a comprehensive approach to Histopathology cancer detection.

In summary, the related work highlights the evolution of methodologies in Histopathology cancer detection, emphasizing the contributions of the proposed approach. By integrating Thepade SBTC and color binning, leveraging machine learning algorithms, and introducing an ensemble learning paradigm.

3. Proposed Synergetic Fusion of Features for Cancer Detection in Histopathology Images

The proposed method aims to enhance Histopathology cancer detection by introducing a novel approach that leverages Thepade SBTC in conjunction with color binning. This section outlines the key steps of the proposed methodology, As illustrated in Figure 1, emphasizing the synergy between Thepade SBTC and color binning for comprehensive feature generation.

3.1. Thepade SBTC Feature Extraction:

The initial step of the proposed method involves the application of Thepade SBTC to generate features from Histopathology images. Thepade SBTC extends the traditional Block Truncation Coding (BTC) by introducing a sorting mechanism based on block size before applying BTC. This sorting optimizes the compression efficiency by encoding smaller blocks with fewer bits, taking advantage of the inherent property that smaller blocks tend to contain less intricate details. As shown in Figure 2, The result is a set of features capturing both structural and statistical characteristics of the image. The Thepade SBTC process involves dividing the image into blocks, sorting them based on size, and then applying BTC to encode the blocks. This method provides an effective means of feature extraction, enabling the representation of images with reduced dimensionality while preserving essential information.

3.2. Color Binning Feature Extraction:

Concurrently, color binning is employed to extract features related to pixel intensity levels in the RGB color space. As illustrated in Figure 3, Color binning classifies pixel values into bins, offering insights into the color distribution of the image. This method complements Thepade SBTC by capturing color-related information, providing a holistic representation of the Histopathology images. The binning process involves categorizing pixel values into predefined intensity bins, allowing for a compact representation of color features. The resultant features encapsulate the distribution and intensity of colors within the image, contributing valuable information for discriminating between cancerous and non-cancerous tissue.

3.3 Proposed Feature Combination:

The core innovation of the proposed method lies in the fusion of Thepade SBTC and color binning features. The devised feature set is constructed by concatenating the features extracted from Thepade SBTC with those obtained from color binning. This combination aims to capture both structural intricacies and color-related details, providing a more comprehensive feature representation for subsequent classification. By combining these features, the approach seeks to address the limitations of individual methods, offering a more robust and discriminative feature set that enhances the accuracy of cancer detection in Histopathology images.

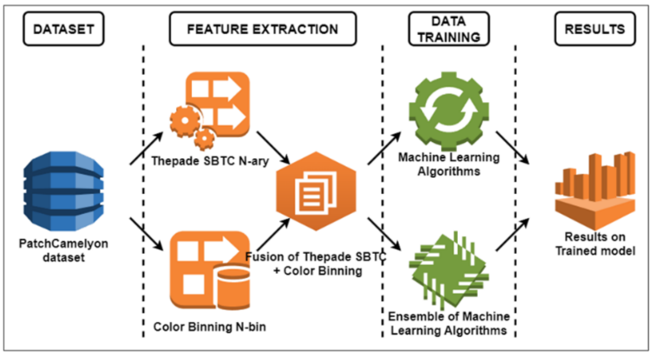

Fig. 1. Process flow chart of proposed Synergistic Fusion of Thepade SBTC and Color Binning Features for Lymph node cancer detection

3.4. Machine Learning Classification:

The extracted features are then subjected to an ML classification pipeline, where various algorithms are employed to evaluate the discriminative power of the features. The method assesses nine machine learning algorithms. This comprehensive evaluation aims to detect the best suiting algorithm for the proposed feature set.

3.5 Ensemble Learning Integration:

Building upon the success of individual classifiers, an ensemble learning approach is explored as shown in Figure 4 to further boost classification accuracy. A Voting Classifier is constructed, utilizing KNN, Random Forest, and CatBoost as base estimators. The ensemble model combines the strengths of these classifiers, leveraging their complementary characteristics to achieve a more robust and accurate Histopathology cancer detection model.

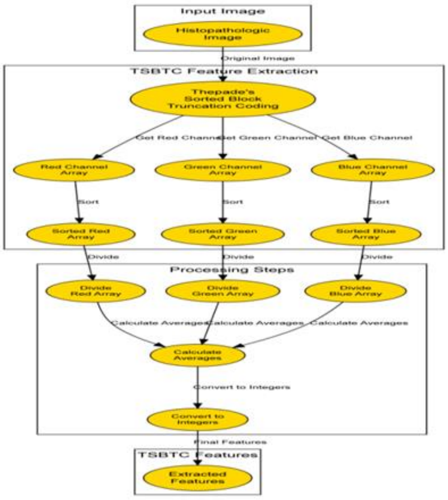

Fig. 2. Flowchart of Thepade SBTC Feature Extraction Process

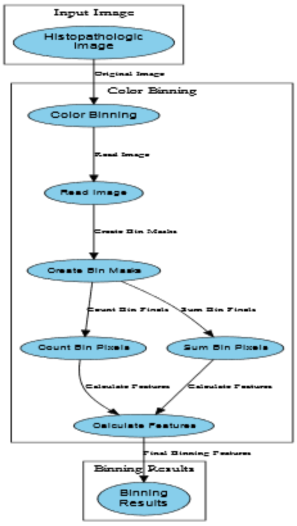

Fig. 3. Flowchart of Color Binning method

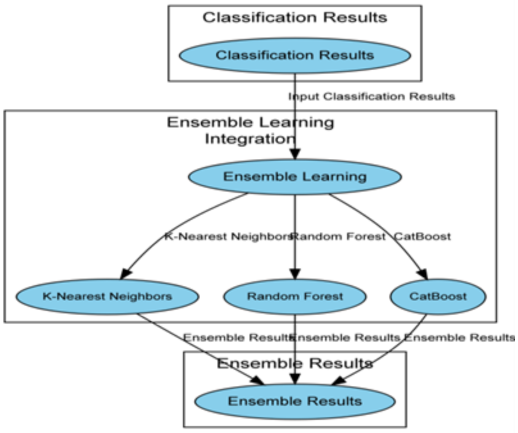

Fig. 4. Flowchart of Ensemble Learning Integration Process

In summary, the proposed method presents a novel approach by combining Thepade SBTC and color binning for feature extraction, followed by a comprehensive evaluation using machine learning and ensemble learning techniques. This synergistic combination aims to enhance the accuracy and reliability of Histopathology cancer detection, contributing to the advancement of medical image analysis methodologies.

4. Results and Discussion

A comprehensive review of the experimental result observed from the application of the Thepade SBTC method and the proposed Feature Combination for Histopathology cancer detection is drafted in this section.

4.1. Dataset Details

Figure 5 Indicates the dataset used for exploration is the 'PatchCamelyon' dataset [19], consisting of 220025 color images with dimensions of 96 x 96 pixels. These images were acquired using Histopathology scans of lymph node sections, and every image is annotated for the absence or presence of metastatic tissue. Specifically, a positive label reflects that the central 32x32 pixel region contains tumor tissue, while tumor tissue in the outer region does not influence the label.

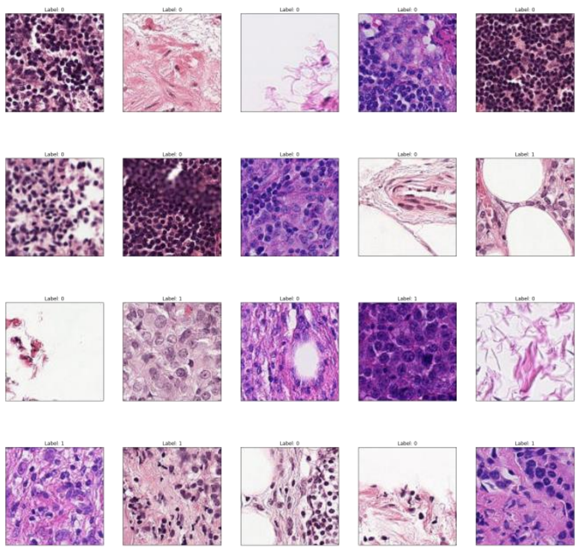

Fig. 5. Images in the PatchCamelyon dataset

4.2. Thepade SBTC Feature Extraction Method

It involves the Feature Extraction and Classification Results Process.

4.2.1. Features Extraction

The Thepade SBTC method was applied to extract features from the Histopathology images, ranging from 2-ary to 15-ary representations. For each representation, the average values of the lower and upper bounds for the Red, Green, and Blue channels were calculated. The features are used as inputs for subsequent ml classification.

4.2.2. Classification Results

The Thepade SBTC-derived features were subjected to nine ml algorithms. The accuracy of each algorithm was measured for each ary level, ranging from 2-ary to 15-ary, As illustrated in Figure 6.

The results revealed notable variations in accuracy across ary levels, with certain algorithms exhibiting superior performance for specific ary representations. Particularly, KNN, Random Forest, and CatBoost consistently demonstrated high accuracy across various ary levels.

4.3. Color Binning Method

It involves the Feature Extraction and Classification Results Process.

4.3.1. Features Extraction

The color Binning method was employed to extract features in the form of color-based bins. Specifically, a 2-bin representation was utilized, where pixel values below 127 were classified into one bin, and values above were classified into another. Similar to Thepade SBTC, these features were averaged for each channel.

4.3.2. Classification Results

Similar to the Thepade SBTC analysis, the color Binning features were subjected to the same set of nine machine learning algorithms. The classification accuracy was measured for each bin level, ranging from 2-bin to 15-bin, As illustrated in Figure 7.

The analysis highlighted the impact of binning on classification performance, with certain algorithms demonstrating proficiency in interpreting binned features. KNN, Random Forest, and CatBoost once again emerged as the top-performing algorithms across various bin levels.

4.4. Ensemble Learning Integration

It involves the Feature Combination and Ensemble Classification Process.

4.4.1. Proposed Feature Combination

A Feature Combination was proposed to leverage the strengths of both Thepade SBTC and Color Binning. The features obtained from Thepade SBTC (14-ary) and Color Binning (6-bin), which exhibited the highest accuracies individually, were cascaded together.

4.4.2. Ensemble Classification

The combined features are explored with an ensemble learning approach. KNN, Random Forest, and CatBoost were employed as base estimators in a Voting Classifier. The results demonstrated a significant enhancement in classification accuracy, reaching an impressive 90%.

Fig. 6. Performance comparison of ML algorithms for respective Thepade SBTC Ary sizes for feature extraction of histopathology images in Cancer Detection

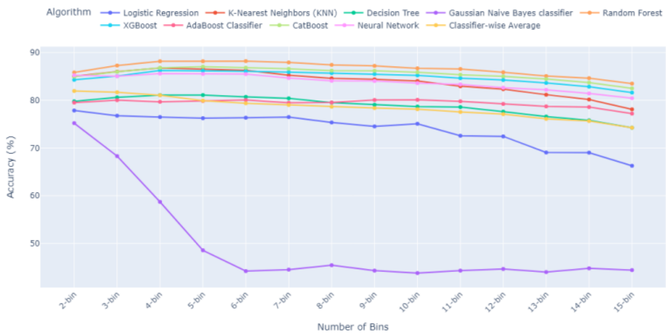

Fig. 7. Performance comparison of ML algorithms for different Bin sizes for feature extraction of histopathology images in Cancer Detection

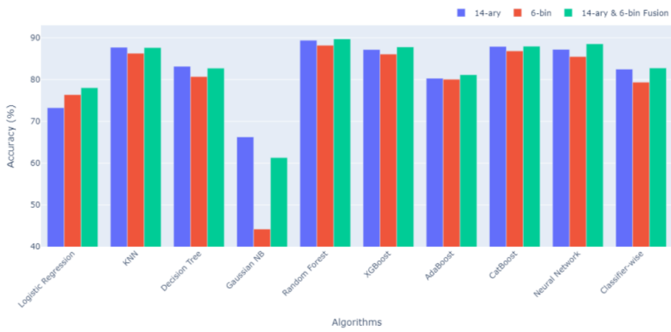

Fig. 8. Comparative analysis of Thepade SBTC 14-ary, Binning with 6-bins, and proposed fusion of TThepade SBTc with Binning in cancer detection for respective ML algorithm

Table 1. Comparative analysis of the proposed method with the relevant similar existing methods

References	*Technique*	*Dataset Used*	*Accuracy (%)*
Eric Bonnet [15], 2021	Shelf deep learning framework	PCam (Patch CAMELYON)	89
Wang et al.[16], 2022	Hybrid DL model - CNN-GRU	Kaggle Histopathologic Cancer Detection	86.21
Abhijeet Patil et	Attention-based multiple instance	BreakHIS and BACH	84.43

al.[17], 2019	learning (A-MIL)		
Bryan He et al.[18], 2020	ST-NET	Histopathology images of breast cancer	66
Proposed Method	Color Binning Method	PCam (Patch CAMELYON)	88.20
Proposed Method	Thepade SBTC	PCam (Patch CAMELYON)	89.40
Proposed Method	Fusion of Color binning and Thepade SBTC	PCam (Patch CAMELYON)	90.08

4.5. Comparative Analysis

This analysis involves a performance comparison of 14-ary, 6-bin, and 14-ary & 6-bin Fusion.

4.5.1. Performance of Machine Learning Algorithms with 14-ary and 6-bin Features:

First, the performance of all ml algorithms is examined when applied to histopathology cancer detection using features extracted through both Thepade SBTC with 14-ary and color binning with 6-bin methods. The algorithms considered for evaluation include 9 ml algorithm, and a classifier-wise average.

As illustrated in Figure 8, the results demonstrate varying levels of accuracy across the different algorithms and feature extraction techniques. Notably, Random Forest consistently achieves high accuracy with both 14-ary and 6-bin features, indicating its robustness in capturing complex patterns within the dataset. Conversely, Gaussian Naive Bayes exhibits relatively lower accuracy, particularly with 6-bin features, suggesting its limitations in handling the inherent complexity of histopathology images.

4.5.2. Ensemble Learning Integration:

To further enhance the predictive performance, an ensemble learning technique is employed, specifically a Voting Classifier, to integrate the predictions from the Random Forest with 14-ary features and the best-performing algorithm with 6-bin features. The ensemble approach aims to leverage the complementary strengths of individual classifiers to achieve superior accuracy.

The experimentation with ensemble learning yielded promising results, achieving a remarkable accuracy of 90.08%. This substantial improvement over individual classifiers underscores the effectiveness of ensemble methods in harnessing the diversity of classifiers and mitigating the limitations associated with individual algorithms.

5. Conclusion and Future Work

The proposed work presented an exploration of feature extraction methods and their integration for the task of Histopathology cancer detection. Leveraging the PatchCamelyon benchmark dataset, the study evaluated the efficacy of Thepade SBTC, Color Binning, and a novel Hybrid Feature Combination approach. The application of Thepade SBTC across varying ary levels (2-ary to 15-ary) unveiled the significance of block sorting in enhancing compression efficiency. By exploiting the intrinsic characteristics of smaller and larger blocks, Thepade SBTC demonstrated its ability to extract meaningful features for cancer detection. The analysis of nine machine learning algorithms revealed nuanced performance variations across ary levels, with KNN, Random Forest, and CatBoost consistently exhibiting superior accuracy. The Color Binning method, utilizing a 2-bin to 15-bin representation, showcased the impact of pixel classification on feature extraction. The experimental results emphasized the adaptability of certain algorithms, particularly KNN, Random Forest, and CatBoost, in interpreting binned features. This underlines the importance of considering the nature of the dataset and feature characteristics when selecting an appropriate feature extraction method.

Motivated by the strengths observed in Thepade SBTC (14-ary) and Color Binning (6-bin), The Feature Combination approach is proposed. The ensemble learning integration, incorporating KNN, Random Forest, and CatBoost, demonstrated a remarkable accuracy improvement, reaching 90%. This underscores the potential of combining complementary features for enhanced cancer detection.

While Thepade SBTC and Color Binning individually provided valuable insights, their integration surpassed individual capabilities. The Feature Combination not only showcased superior accuracy but also demonstrated the potential for combining diverse feature extraction techniques to achieve more robust results.

In summary, the proposed work contributes valuable insights into feature extraction methodologies for Histopathology cancer detection. The amalgamation of Thepade SBTC, Color Binning, and Feature Combination offers a holistic perspective on feature representation, paving the way for improved diagnostic tools in the realm of medical image analysis.

Author contributions

Sudeep D. Thepade: Conceptualization, Methodology, Experimentation, Result analysis, Conclusion

Ashish A. Gore: Experimentation, Validation, Draft writing,

Rachana Patil: Draft review, Conclusion and editing

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Krithiga, R., Geetha, P. "Breast Cancer Detection, Segmentation and Classification on Histopathology Images Analysis: A Systematic Review." Arch Computat Methods Eng 28, 2607–2619 (2021). https://doi.org/10.1007/s11831-020-09470-w

[2] Thepade, Sudeep D., and Akash P. Bhalerao. "An Integrated Approach for Land Use Identification of Aerial Images using ResNet50 DCNN Model and Thepade SBTC." International Journal of Computing and Digital Systems 14 (2023): 1-xx.

[3] Thepade, S. D., & Adrakatti, S. S. (2023). Fusing Deep Convolution Neural Network features of MobileNet with Thepade SBTC for land usage identification. International Journal of Computing and Digital Systems, 14(1), 1-xx.

[4] Thepade, S. D., & Sayyad, A. I. (2023). Melanoma Skin Cancer Detection from Dermatoscopy Images Using Xception and Thepade SBTC Model. International Journal of Computing and Digital Systems, 14(1), 1-xx.

[5] S. D. Thepade, D. Abin and K. H. Chauhan, "Improved Rice Leaf Disease Detection using Fusion of Otsu Thresholding and Thepade SBTC Features," 2022 International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC), Hyderabad, India, 2022, pp. 1-6, doi: 10.1109/ICMACC54824.2022.10093433.

[6] Wee Kheng Leow, Rui Li, The analysis and applications of adaptive-binning color histograms, Computer Vision and Image Understanding, Volume 94, Issues 1–3, 2004, Pages 67-91, ISSN 1077-3142, https://doi.org/10.1016/j.cviu.2003.10.010.

[7] Jin, X., Hirakawa, K. Analysis and processing of pixel binning for color image sensor. EURASIP J. Adv. Signal Process. 2012, 125 (2012). https://doi.org/10.1186/1687-6180-2012-125

[8] Thepade, Sudeep D., Piyush R. Chaudhari, and Rik Das. "Identifying land usage from aerial image using feature fusion of thepade's sorted n-ary block truncation coding and Bernsen thresholding with ensemble methods." International Journal of Engineering and Advanced Technology (IJEAT) 9.3 (2020): 2612-2621.

[9] E. Delp and O. Mitchell, "Image Compression Using Block Truncation Coding," in IEEE Transactions on Communications, vol. 27, no. 9, pp. 1335-1342, September 1979, doi: 10.1109/TCOM.1979.1094560.

[10] Bibhas Chandra Dhara, Bhabatosh Chanda, Block truncation coding using pattern fitting, Pattern Recognition, Volume 37, Issue 11, 2004, Pages 2131-2139, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2004.02.008.

[11] Anant Madabhushi, George Lee, Image analysis and machine learning in digital pathology: Challenges and opportunities, Medical Image Analysis, Volume 33, 2016, Pages 170-175, ISSN 1361-8415, https://doi.org/10.1016/j.media.2016.06.037.

[12] Zheng, Yuchao, et al. "Application of transfer learning and ensemble learning in image-level classification for breast histopathology." Intelligent Medicine 3.02 (2023): 115-128.

[13] Madhu, and Raman Kumar. "A hybrid feature extraction technique for content based medical image retrieval using segmentation and clustering techniques." Multimedia Tools and Applications 81.6 (2022): 8871-8904.

[14] K. Karthik and S. Sowmya Kamath, "A Hybrid Feature Modeling Approach for Content-Based Medical Image Retrieval," 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS), Rupnagar, India, 2018, pp. 7-12, doi: 10.1109/ICIINFS.2018.8721432.

[15] Bonnet E. Using convolutional neural networks for the classification of breast cancer images. arXiv. 20212108.13661

[16] Wang X., Ahmad I., Javeed D., Zaidi S.A., Alotaibi F.M., Ghoneim M.E., Daradkeh Y.I., Asghar J., Eldin E.T. Intelligent Hybrid Deep Learning Model for Breast Cancer Detection. Electronics. 2022;11:2767. doi: 10.3390/electronics11172767

[17] Patil A., Tamboli D., Meena S., Anand D., Sethi A. Breast Cancer histopathology image classification and localization using multiple instance learning; Proceedings of the 2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE); Bengaluru, India. 15–16 November 2019; pp. 1–4.

[18] He B., Bergenstråhle L., Stenbeck L., Abid A., Andersson A., Borg Å., Maaskola J., Lundeberg J., Zou J. Integrating spatial gene expression and breast tumour morphology via deep learning. Nat. Biomed. Eng. 2020;4:827–834. doi: 10.1038/s41551-020-0578-x.

[19] Bejnordi, B.E.; Veta, M.; Van Diest, P.J.; Van Ginneken, B.; Karssemeijer, N.; Litjens, G.; Van Der Laak, J.A.W.M. "Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women with Breast Cancer." JAMA 2017, 318, 2199–2210.