

# Evaluating the necessity of colour normalisation in deep learning-based histopathological image classification

Natalya Smith<sup>1</sup>

Faculty of Design and Creative Technologies,  
School of Computer and Mathematical Sciences  
Auckland University of Technology, Auckland, New Zealand  
[nnb8479@autuni.ac.nz](mailto:nnb8479@autuni.ac.nz)

**Abstract.** Colour normalisation (CN) is a widely used pre-processing histopathological image analysis (HIA) technique, aimed at standardising staining variations and enhancing model performance. This study investigates the effectiveness of four CN techniques: (1) simple channel-based normalisation, (2) colour deconvolution (CD) with the Haematoxylin component, (3) channel-based normalisation combined with Contrast Limited Adaptive Histogram Equalisation (CLAHE), and (4) CD with the Haematoxylin component combined with CLAHE. These techniques were selected for their popularity and distinct approaches to handling colour variability. Additionally, the performance of models trained with data augmentation (DA) only (0), without CN applied, was also evaluated.

The findings indicate that technique 0 achieved the highest performance, with a ROC-AUC score of 0.997. Technique (1) also showed strong performance with a ROC-AUC score of 0.995. In terms of sensitivity and specificity, technique 0 achieved 98.85% and 96.53%, respectively, while technique (1) achieved 98.42% and 93.62%. However, the combination of CD with the Haematoxylin component and CLAHE (technique 4) exhibited the lowest performance, with a ROC-AUC score of 0.782, sensitivity of 94.07%, and specificity of 30.70%.

The study challenges the necessity of CN in HIA because simple DA is found to be more effective in enhancing model performance. While CN is traditionally believed to improve feature consistency and enhance diagnostic features, our results suggest that CN may introduce variability that affects generalisation. By proposing alternative strategies, such as DA and comprehensive validation, this study helps streamline the pre-processing pipeline, reduce computational costs, and improve the scalability and accessibility of HIA. These findings provide a more efficient approach to developing robust and reliable HIA methods, potentially leading to better diagnostic accuracy and patient outcomes.

**Keywords:** Histopathological Image Analysis · Breast Cancer Binary Classification · Class Imbalance · Convolutional Neural Networks · Colour Normalisation · Cross-Validation with Stratified Split · Medical Imaging

## 1 Introduction

Breast cancer is the most prevalent form of cancer, accounting for 2,308,897 (11.6%) of the 19,964,811 cases reported worldwide in 2022 [12]. Early and accurate diagnosis is critical for improving patient outcomes and survival rates [53]. Histopathological examination of breast tissue biopsies is the gold standard for diagnosing breast cancer. However, it is a labour-intensive and time-consuming process that relies heavily on the expertise of pathologists. The consistency with which pathologists diagnose breast cancer types using histopathologic images (HIs) is not uniform and varies with the condition of the tissue samples and the experience level of the pathologist.

Research indicates that the overall concordance rate among pathologists in cancer diagnosis through HIs stands at approximately 75.3% [31], including instances of both over- and under-interpretation. For instance, while the concordance rate for invasive breast cancer cases is high at 96%, it drops significantly to 48% for atypia cases, underscoring the diagnostic challenges in certain breast cancer types [23]. The variability in tissue appearances and the subtle differences between benign and malignant tissues add to the complexity of the diagnostic process. This variability can lead to either over- or under-diagnosis, significantly affecting patient treatment and outcomes.

To enhance diagnostic accuracy and efficiency, the application of artificial intelligence (AI) and deep learning (DL) presents a promising solution. Recent advancements in digital pathology have created new opportunities for histopathological image analysis (HIA) [54]. Computer vision, in particular, plays a crucial role in extracting quantitative information from HIs [39,56], significantly accelerating analysis and

---

research outcomes [30]. DL models, especially convolutional neural networks (CNNs), have demonstrated promising results in analysing these images and improving diagnostic accuracy.

Crucial for diagnosing and treating diseases such as cancer, HIA relies heavily on the visual features of tissue samples [61]. However, HIs often exhibit significant colour variations due to inconsistent staining processes [95]. These variations can arise from differences in staining protocols, reagent batches, and even the subjective interpretation of staining intensity by pathologists. Colour normalisation (CN) is a common pre-processing step intended to standardise the colour distribution across different images, thereby reducing variability and improving the reliability of subsequent HIA [107]. This process aims to mitigate the impact of staining variability by adjusting the colour profiles of images to match a reference standard [61].

CN is frequently applied in HIA with the assumption that standardising colours across images allows models to focus on structural features. However, this process is computationally expensive and time-consuming. Despite its widespread use, its necessity and impact on advanced DL models for HI classification remain under-explored. Indeed, recent advancements in DL have led to the development of robust models capable of learning invariant features from raw data, potentially reducing the dependency on extensive pre-processing steps like CN [85]. However, the actual benefit of CN in the context of modern DL models, particularly concerning their performance in HI classification, has not been critically evaluated.

This study evaluates the impact of four CN techniques on the performance of the DenseNet201 model in classifying HIs for breast cancer detection. By systematically assessing the effects of these CN methods, the aim is to determine the most effective approach for enhancing model performance. The findings provide valuable insights into the role of CN in medical image analysis, potentially informing best practices and optimising workflow efficiency. Data augmentation (DA) is another critical aspect of HIA, introducing variety during training through techniques like rotations, scaling, and colour perturbations. DA can help DL models learn more robust features and improve their generalisation to unseen data. This study also examines the effectiveness of DA in handling colour variability without relying on CN.

This study aims to provide empirical evidence on the necessity of CN, which could potentially streamline workflows and enhance efficiency. Hence, the primary research question in this study is: *To what extent does CN impact the performance of DL models in classifying HIs?* The objective is to evaluate the necessity of CN for achieving high classification accuracy and to determine whether modern DL models can manage colour variability in HIs without it.

The study utilises the DenseNet201 model, a state-of-the-art CNN, to classify HIs from the BreakHis dataset. The model's performance is evaluated with and without CN using key metrics such as accuracy, sensitivity, specificity, F1-score, and ROC-AUC. The key contributions of the study are:

- Providing empirical evidence on the necessity of CN, potentially leading to more efficient workflows.
- Informing best practices in HIA and model training regarding the handling of colour variability.
- Simplifying the pre-processing pipeline, reducing computational costs, and speeding up analysis if CN is found to be unnecessary.

The findings reveal significant differences in performance metrics across different configurations of CN. Technique 0 (DA only) achieved the highest performance, with a ROC-AUC score of 0.997. Technique (1) also showed strong performance with a ROC-AUC score of 0.995. In terms of sensitivity and specificity, technique 0 achieved 98.85% and 96.53%, respectively, while technique (1) achieved 98.42% and 93.62%. However, the combination of colour deconvolution with the Haematoxylin component and CLAHE (technique 4) exhibited the lowest performance, with a ROC-AUC score of 0.782, sensitivity of 94.07%, and specificity of 30.70%. Interestingly, the model trained with DA only (0), without any CN, outperformed all CN techniques.

These results suggest that modern DL models, coupled with robust DA strategies, can effectively handle inherent colour variability in HIs, potentially rendering CN unnecessary. While CN is traditionally believed to improve feature consistency and enhance diagnostic features, our results suggest that CN may introduce variability that affects generalisation. By proposing alternative strategies such as DA and comprehensive validation, this study helps streamline the pre-processing pipeline, reduce computational costs, and improve the scalability and accessibility of HIA. These findings provide a more efficient approach to developing robust and reliable HIA methods, potentially leading to better diagnostic accuracy and patient outcomes.

The paper is organised as follows. Section 2 presents related work. Section 3 describes the methodology used in this study. Section 4 discusses the results, which is followed by the conclusion and future perspectives in Section 5.

---

## 2 Literature Review

Histopathology, defined as the microscopic examination of diseased tissue samples, is crucial in medical diagnostics and plays a vital role in the diagnosis and treatment of breast cancer [33, 65]. It provides invaluable insights into the underlying cellular and structural changes that characterise various pathological conditions, including breast cancer [65]. Histopathological analysis (HA) of breast biopsy specimens is considered the gold standard for confirming the presence of breast cancer [53, 96]. The significance of histopathology in breast cancer management cannot be overstated.

Digital pathology involves the digitisation of tissue samples, allowing for their storage, retrieval, and analysis using specialised software [28, 36, 73]. This technology enhances the accuracy and efficiency of HA by facilitating the detailed analysis required for the accurate classification of diseases, which is essential for determining the most appropriate course of treatment. Traditionally, the interpretation of HIs has relied heavily on the expertise and subjective judgements of pathologists [22, 94]. However, manual interpretation can be time-consuming and subject to inter-observer variability, where different pathologists may arrive at different conclusions based on their individual experiences and interpretations [88, 93].

To address these challenges, advanced analytical techniques, including machine learning (ML) algorithms, have been developed [16, 52, 66]. ML algorithms, trained on large datasets of HIs, assist pathologists in identifying specific cellular and structural patterns, reducing the time and subjectivity associated with manual interpretation [22, 47, 64, 73]. Integrating these technologies improves diagnostic accuracy and consistency in breast cancer diagnosis, ultimately enhancing treatment strategies [17, 71].

Recent research has increasingly focused on applying ML to medical imaging, particularly for breast cancer diagnosis using HIs. Traditional ML models have shown promise in medical image diagnosis, as highlighted by various studies [5, 19, 27, 35, 45, 69, 96]. DL, a subset of ML characterised by neural networks with many layers, has revolutionised numerous fields by identifying patterns and making predictions, often surpassing human performance in specific tasks. Its impact is notably significant in medical imaging, where it enhances diagnostic accuracy and efficiency [106].

CNNs have revolutionised HIA, excelling in processing complex visual data in stained tissue biopsies by autonomously recognising intricate diagnostic patterns, thus eliminating the need for manual feature extraction [6, 41, 58, 87, 103]. CNNs effectively classify diverse tissue types, delineate specific histological structures, such as tumour cells, and quantify areas of interest [41, 63]. Their resilience to variable factors in histological imaging, including staining techniques and sample preparation, ensures uniform analyses [63, 87]. Moreover, CNNs' rapid processing capabilities surpass traditional methods, transforming operational workflows for pathologists [68, 103].

In clinical settings, the precise classification of HIs using CNNs significantly enhances patient care by aiding treatment decisions, minimising subjective interpretations, and improving diagnostic accuracy [40, 41, 103]. The scope of CNN applications has broadened across medical imaging, with early research demonstrating their superior performance in image classification challenges [15, 38, 49, 57, 84]. Studies on breast cancer detection utilising architectures such as EfficientNet, ResNet101, and DenseNet201 underscore their effectiveness in managing complex HIs [48, 86].

Tasks like nuclei segmentation benefit from CNNs' ability to differentiate various nuclear types and background areas effectively [34, 108]. Furthermore, the extraction of deep texture features from CNNs facilitates objective analysis of tumour HIs, enhancing visualisation, retrieval, and supervised learning applications [102]. Overall, CNNs' ability to analyse complex imaging data accurately and efficiently has led to significant strides in diagnostic precision within HIA.

Histopathological examination remains a cornerstone in diagnosing cancer, where pathologists scrutinise tissue samples under microscopes, often aided by Hematoxylin and Eosin (H&E) stains to better view cellular structures and detect anomalies [50, 72, 103]. However, variability in staining techniques can significantly affect automated analysis, making standardised colour normalisation (CN) techniques crucial for ensuring diagnostic precision [11, 44, 55, 78].

One of the major challenges in HIA is the need for CN, which corrects variations in colour appearance between different tissue samples [74]. This process is essential because colour variations can affect the accuracy and reliability of the analysis, leading to misinterpretation of important histopathological features [91]. Without appropriate CN, comparing and interpreting HIs accurately becomes difficult. Developing an effective CN technique for HIA in breast cancer poses several challenges.

One challenge is the inherent variability in staining protocols and slide preparation techniques, leading to variations in colour appearance [50]. These variations can be caused by differences in tissue fixation, processing, staining time, and the intensity of the staining reagents used. Another challenge is the presence

---

of artifacts, such as tissue folds, debris, or air bubbles, which can introduce additional colour variations and complicate the normalisation process [91]. Additionally, different imaging platforms and devices may have their own colour rendering capabilities, making it necessary to account for these variations in the normalisation process [50].

The complexity and heterogeneity of breast cancer tissue samples further complicate CN [7]. The presence of different cell types, tumour sub-types, and morphological variations within breast cancer samples can result in variations in colour appearance that need to be accounted for in the normalisation process [91]. Furthermore, the limited availability of labelled data poses a challenge. Without a large dataset of accurately annotated images for training and validation, it can be difficult to develop and evaluate robust CN algorithms specific to HIA in breast cancer [7].

Several studies have highlighted the challenges and limitations associated with CN in HIAs. Techniques used to standardise colour in HIs for improved visual analysis and computational processing include the structure-preserving CN technique that enhances efficiency in processing whole-slide images [77]; a novel colour-based classifier incorporating a stain colour descriptor to calculate image-specific stain matrices [44]; and histogram specification, which adjusts the colour distribution of an image to match a specified histogram [50].

Comprehensive reviews underscore the necessity of CN for HIA [90], extending its utility to cell segmentation through sophisticated CNN-based image processing [8]. CN consistently augments model performance in cancer detection and classification [3, 32, 79, 101], although robust feature selection may sometimes reduce the need for such normalisation by capturing essential colour-texture information [29].

The impact of CN on the accuracy of nuclei segmentation in H&E-stained images has been demonstrated using advanced CNN architectures like U-Net and Residual U-Net with promising results [104]. Moreover, an enhanced DenseNet architecture has been discussed for its proficiency in breast cancer image classification due to its capability to capture complex tissue patterns [92]. A novel hybrid CN approach, utilising modified non-negative matrix factorisation, surpasses traditional methods in managing staining and illumination variations [80].

Researchers have explored various CN techniques. Simple channel-wise normalisation, for example, aims at standardising the colour information across different images [7]. This method can show effectiveness in scenarios where colour consistency helps DL models to differentiate between healthy and cancerous tissues. Moreover, this technique can particularly benefit processes where variations due to staining protocols or imaging equipment settings introduce unwanted discrepancies in the data. Colour deconvolution (CD) is an advanced technique used in histopathology to separate and analyse the constituent stain components of stained tissues (H&E) [51, 60].

CD works by mathematically decomposing the colour information into separate channels that represent the pure stain distributions [7, 51]. This separation leverages the specific light absorption characteristics of each stain, allowing the algorithm to standardise staining effects across different images [9, 87, 105]. By isolating these stain channels, CD enhances the comparability and analysis of tissue structures across various histopathological samples, regardless of staining intensity variations [7, 60].

This technique is particularly effective in cancer classification tasks, providing accurate representations of histological features crucial for distinguishing between benign and malignant structures [18, 75, 98]. CD improves subsequent analytical steps, such as segmentation, feature extraction, and classification by machine learning models, by ensuring a more uniform and precise dataset [98]. Specifically, CD with Hematoxylin focuses on nucleic acids, offering detailed insights into cell nuclei and other tissue structures, aiding in the detection and analysis of diseases like cancer [10, 14].

Channel-based normalisation combined with CLAHE involves a more advanced form of histogram equalisation that operates on small regions in an image, called tiles, rather than the entire image. CLAHE is designed to improve the contrast of the images while overcoming the issue of noise amplification seen with standard histogram equalisation [4, 21]. By doing so, it can enhance the visibility of important features in medical images that can be critical for accurate cancer classification. Unlike standard histogram equalisation that uniformly enhances contrast across the entire image, CLAHE enhances local contrast and limits noise amplification [70, 83, 97]. It is particularly useful in improving the visibility and definition of important features within histopathology slides without over-saturating the image.

The effectiveness of CN techniques and CLAHE can be highly dependent on the type of cancer, the quality of medical images, and the specific characteristics of the dataset being used. CLAHE might be more effective in certain contexts, particularly where contrast enhancement is crucial for detecting subtle differences in tissue appearance that are indicative of disease [20]. Ultimately, the choice between different CN methods and CLAHE - or whether to combine them - should be made based on experimental results showing the impact of these techniques on the performance of the classification algorithms used. Hence,

---

it is advisable to conduct a comparative analysis under different conditions to determine the relative effectiveness of these pre-processing methods [2].

DA is crucial in DL models, particularly in medical imaging, where it addresses the scarcity and high cost of obtaining labelled datasets. By introducing realistic variations that mimic potential occurrences during data acquisition, DA helps prevent model over-fitting and ensures adaptability to diverse imaging conditions [1, 25, 37]. In the medical field, DA significantly enhances the effective size of available datasets, facilitating the training of more accurate and generalisable models [13, 25, 67]. It ensures the model's invariance to discrepancies caused by differences in imaging equipment, patient positioning, and other factors [26, 37]. The impact of DA on model generalisation and robustness is profound [100].

DA introduces transformations such as rotation, scaling, flipping, cropping, and altering brightness or contrast. This variety helps the model generalise better to new, unseen data, preventing over-fitting [62]. By exposing the model to a wider range of variations, DA reduces the likelihood of confusion from minor changes or distortions common in real-world scenarios [42, 100]. DA helps address class imbalance by creating more examples of under-represented classes, improving performance in classifying these classes [37]. When applying a model to a new but related problem (i.e., transfer learning), DA can incorporate domain-specific variations, tailoring the model to the new task [42, 46].

Several studies have explored the use of DA to address the limited availability and variability of medical datasets, tackling challenges such as class imbalance and over-fitting. Comprehensive surveys by [42, 81], among many others, have reviewed numerous DA techniques and their applications, particularly in medical imaging. Research highlights the effectiveness of geometric transformations and colour space augmentations, noting that techniques like rotation and flipping significantly enhance model generalisation.

For example, [24] used extensive DA, including rotations, translations, and colour adjustments, to train a CNN for skin cancer classification, achieving dermatologist-level accuracy. Similarly, [13] systematically examined various DA methods in medical imaging, demonstrating consistent improvements in model accuracy and robustness, with geometric transformations being the most effective. [76] explored the combination of transfer learning with DA techniques in medical image classification tasks, showing significant performance improvements, especially in scenarios with limited data. [89] further validated this approach by integrating transfer learning with a DA strategy, enhancing model performance across various datasets.

Implementing DA techniques helps synthetically expand datasets, providing the model with diverse situations to learn from. This reduces the risk of over-fitting by preventing the model from memorising specific patterns in the training data, thereby improving its ability to generalise to new, unseen data. Consequently, DA is essential for creating robust models that perform well across diverse datasets and in real-world applications.

### 3 Methodology

A series of experiments were conducted to quantify the impact in performance of the four CN techniques - simple channel-based normalisation, CD with Hematoxylin component, channel-based normalisation combined with CLAHE, and CD with CLAHE - in classifying HIs for breast cancer detection. We argue that while CN is widely assumed to enhance model performance by standardising staining variations, its necessity and effectiveness need thorough evaluation in the context of advanced DL models. By systematically comparing the chosen CN techniques, we aimed to identify the most effective approach for enhancing model performance and reliability, contributing valuable insights to the ongoing research in HIA.

*Dataset* For the experiments, a well-established public Breast Cancer Histopathological Database (BreakHis) was utilised (for a description, see [86])<sup>1</sup>. Although widely used in the development and evaluation of computer-aided diagnosis (CAD) systems for breast cancer, it presents a challenging benchmark for CAD system development due to the inherent variability in tissue appearances.

BreakHis consists of 7,909 microscopy images of breast tissue biopsies obtained from 82 patients, with 2,480 being benign and 5,429 malignant. With 5429 samples in the majority class and 2480 in the minority class (representing about 31% of the dataset), there is a noticeable class imbalance. Although is not extreme, such imbalance can impact model performance, since more complex models (e.g., deep neural networks) tend to be more sensitive to imbalance.

---

<sup>1</sup> Available here.

---

*Imbalance strategy* To address the inherent challenge of class distinction within the BreakHis dataset and mitigate bias towards the majority class, the model built in this study employs a custom weighted binary cross-entropy loss function. This approach compensates for the under-representation of the minority class by assigning different weights to the classes, enhancing the model's ability to learn distinguishing features accurately. The weighting strategy is depicted in Equation 1:

$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^N [w_p y_i \log(p_i) + w_n (1 - y_i) \log(1 - p_i)], \quad (1)$$

where  $w_p$  and  $w_n$  represent the weights for the positive and negative classes, respectively.

*The Split* The full dataset - comprising paired image data and labels - was divided into training and testing sets using a stratified split, allocating 20% of the data to testing. This ensured that each subset maintains the same proportion of class labels as the original dataset, which is particularly important for handling imbalanced classes in cancer detection. Such an approach allows for an unbiased evaluation of the model's performance on unseen data, providing a robust assessment of its generalisability. To further refine the model and prevent over-fitting, the training data was further subdivided, with 20% set aside as a validation set, also using stratified sampling. This validation set is crucial for tuning the model's hyperparameters and making iterative adjustments based on performance metrics that are not directly influenced by the test data. This strategy ensures that the model is both accurate and reliable.

*Model Architecture and Training* The core of the built model is the DenseNet201 architecture. It is recognised for its dense connectivity pattern between layers, which significantly enhances feature propagation and utilisation, making it highly effective for medical image analysis including the detection of cancerous tissues in HIs [82], [92]. More specifically,

- *DenseNet201 backbone*: utilised dense connectivity to ensure maximum information flow between layers;
- *Additional convolutional layers*: with ReLU activation refined the features extracted by DenseNet201 for the binary cancer classification;
- *Batch normalisation*: was applied after each convolution to stabilise learning and reduce the number of epochs needed for convergence;
- *MaxPooling and Dropout*: were used as the former reduces spatial dimensions to decrease computational load, while the latter mitigates over-fitting by randomly omitting units during training.

The model employed a binary classification output using a *Sigmoid Activation Function*.

*Training Strategy* was conducted over 25 epochs, using a batch size of 32. The Adam optimiser facilitated the optimisation process, it was set with an initial learning rate of  $1 \times 10^{-4}$  and leveraging a momentum of 0.9. This learning rate configuration aided in a smooth and effective convergence during the training phase. To enhance training efficiency and prevent over-fitting, the following strategies were employed:

- *Early stopping*: training was monitored for validation loss and was halted if no improvement was seen for five consecutive epochs to prevent over-fitting;
- *ReduceLROnPlateau*: the learning rate was reduced by a factor of 0.4 if no improvement was observed in the training loss over two epochs, with the minimum learning rate set to  $1 \times 10^{-7}$ .

To ensure the robustness and generalisability of the findings, a k-fold cross-validation approach was employed. Specifically, the dataset was divided into 5 folds, and the model was trained and validated five times, with each fold serving as the validation set once and the remaining four folds used for training. This method provides several advantages:

- *Variance reduction*: averaging the results across multiple folds helps mitigating the impact of variability in the dataset, leading to more stable and reliable performance metrics;
- *Comprehensive evaluation*: each data point is used for both training and validation, ensuring that the model is tested on all available data, which enhances the thoroughness of our evaluation;
- *Bias reduction*: cross-validation helps in reducing selection bias by ensuring that the model's performance is not overly dependent on any specific subset of the data.

*DA* (applied to all configurations) was employed to enhance model robustness and generalisation and included:

- *Zooming* (20% range): helps the model learn from varying scales and proportions of image features, crucial in histopathology where pathological feature sizes differ;
- *Flipping* (horizontal and vertical): increases data variability and promotes a model invariant to tissue and cell orientation;
- *Shearing* (20% range): simulates changes in viewing angle by displacing parts of the image, mimicking variations in microscope slide placement;
- *Rotation* (20-degree limit): handles random orientation of tissues on pathology slides;
- ‘*Nearest fill*’ mode: helps maintain HI integrity by filling missing pixels with nearby values.

These transformations were carefully chosen to produce realistic images while preserving essential diagnostic features. It is important to note that, in the development of this model, various DA approaches with differing degrees of aggressiveness were tested. These included varying the extent of rotations, zoom levels, and the introduction of additional types of transformations. However, it was observed that more aggressive augmentations tended to reduce the model’s accuracy. This was possibly due to the introduction of features into the images that do not accurately represent true histopathological scenarios [99], [43], potentially confusing the model and causing it to learn incorrect patterns. For example, extremely high zoom levels can distort cellular structures beyond realistic proportions, and excessive rotation can present tissues in orientations that are not encountered in actual diagnostic settings.

Figure 1 demonstrates the effectiveness of the augmentation strategy, creating diverse training samples without introducing misleading artifacts.

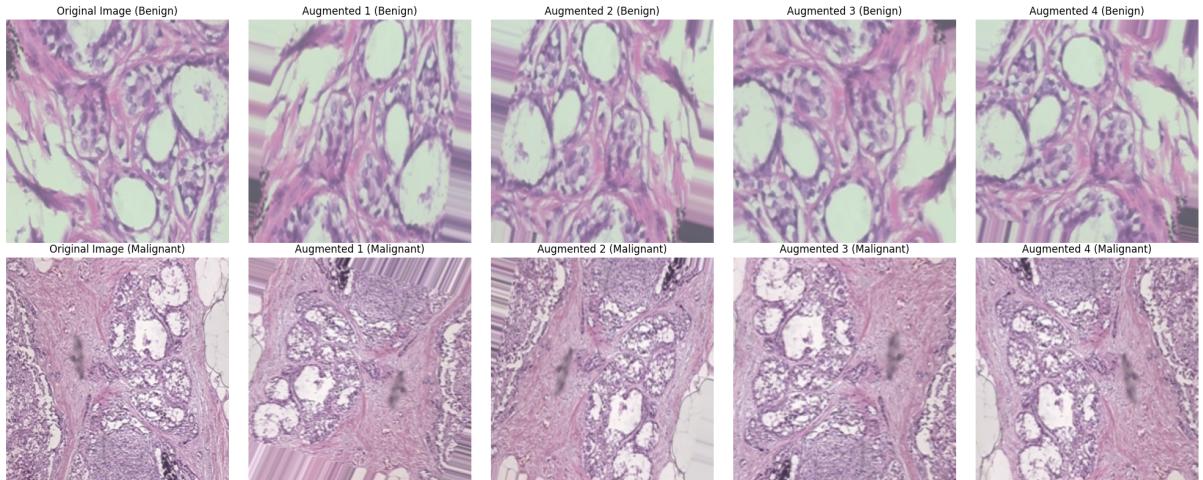


Fig. 1: Examples of augmented images from the training set.

*CN* The following configurations were utilised:

- *No colour normalisation (0)*.

(0) represents the baseline configuration where no CN was applied. It relies on the model’s inherent ability to handle colour variations present in the raw HIs. It serves as a control to evaluate the necessity and impact of applying CN techniques.

- *Simple channel-based normalisation (1)*.

(1) employs basic channel normalisation, scaling each channel of the RGB (red, green, blue) image independently to a range of [0, 1]: it adjusts each pixel value such that the minimum intensity is mapped to 0 and the maximum to 1, effectively mitigating variations in illumination. This ensures that pixel intensities are scaled to a consistent range, enhancing the uniformity across different imaging conditions as shown in the below equation:

$$I_{\text{normalised}} = \frac{I - \min(I)}{\max(I) - \min(I)} \quad (2)$$

---

where  $I$  represents the pixel values of an image. As a result, the images retain their original colours but with balanced intensity across channels. This method ensures consistent colour representation, which can improve the model's ability to generalise. However, it may not address specific staining variability issues in HIs.

- *CD with Hematoxylin component (2).*

(2) advances the pre-processing by implementing CD to isolate specific stain components using pre-defined H&E vectors. This technique enhances key diagnostic features by focusing on the Hematoxylin component, which provides a more consistent representation of cellular structures. The operation to separate pixel intensities into individual stain contributions as described in:

$$C = A^{-1} \times I \quad (3)$$

where  $C$  denotes the stain concentrations,  $I$  is the intensity matrix, and  $A^{-1}$  represents the inverse of the stain vector matrix  $A$ . This equation assumes a linear combination of stains represented by  $A$ , effectively isolating individual stain contributions for clearer analysis. As a result, the images appear with a focus on the Hematoxylin channel, which highlights the nuclei. Emphasising the Hematoxylin stain can improve the model's focus on nuclei, crucial for cancer detection. However, this focus might reduce contrast for other important features.

- *Channel-based normalisation with CLAHE (3).*

(3) incorporates contrast enhancement through CLAHE after basic-channel normalisation. CLAHE is applied specifically to the luminance channel in the LAB colour space (defined by the International Commission on Illumination, consisting of Lightness (L), and chromaticity layers A (green to red) and B (blue to yellow)), improving the contrast in critical areas and aiding both visual inspection and automated analysis. This method enhances image contrast by adjusting luminance levels, optimising visibility of critical features in medical imaging.

The process is described by the following equation:

$$I_{\text{CLAHE}} = \text{applyCLAHE}(I_{\text{LAB}}) \quad (4)$$

where  $I_{\text{CLAHE}}$  represents the image after applying the CLAHE algorithm, and  $I_{\text{LAB}}$  denotes the image converted into LAB colour space. Basic normalisation with CLAHE applied to enhance contrast results in the images that are more vibrant with enhanced local contrast, making features like cell boundaries more distinct. CLAHE can help the model by making important features more prominent, potentially improving accuracy. However, it might also introduce noise, affecting performance.

- *CD normalisation with CLAHE (4).*

(4) combines (2) with CLAHE, applying the latter to enhance local contrast throughout the image. This method aims to improve overall visual clarity and contrast by focusing on the Hematoxylin channel and enhancing it through CLAHE. The process is described by the following equation that represents the conversion of a normalised image to LAB colour space, enhancement using CLAHE, and re-conversion to RGB format, enhancing image contrast and visual clarity effectively:

$$I_{\text{final}} = \text{toRGB}(\text{CLAHE}(\text{toLAB}(I_{\text{normalised}}))) \quad (5)$$

where  $I_{\text{normalised}}$  is the image after initial normalisation. (4) for contrast enhancement is similar to (2) but with enhanced contrast. Combining Hematoxylin normalisation with CLAHE can enhance critical features while maintaining focus on nuclei. This balance might yield the best performance, improving both sensitivity and specificity.

Therefore, the choice of the pre-processing approach is based on highlighting the relevant features for the model while maintaining important structural and colour information:

- *Feature visibility:* the pre-processing should enhance the features that are crucial for cancer detection, such as cell nuclei and tissue structures.
- *Consistency:* the method should provide consistent results across all images to avoid introducing bias or variability that could affect model performance.
- *Colour information:* while grayscale images can work, preserving some colour information may help the model learn more effectively from the data.

CN techniques that focus on the Hematoxylin channel, i.e., (2) and (4), are likely to emphasise nuclei, which are crucial for identifying cancerous cells. This emphasis helps the model learn relevant features more effectively. Applying CLAHE, as in (3) and (4), enhances local contrast, making features more distinct, which can aid in better feature extraction by the model, though it might introduce noise if over-applied. Config. (1) ensures consistent colour representation but might not address all variability in staining protocols. Combining (1) with CLAHE in (3) can improve feature visibility. Balancing between normalisation and enhancement is key, making (4) potentially the best configuration as it enhances critical features while maintaining focus on nuclei.

Figure 2 shows how each pre-processing method refines HIs, potentially enhancing diagnostic accuracy. Each configuration varies in how it adjusts contrast and emphasises features but preserves essential tissue properties.

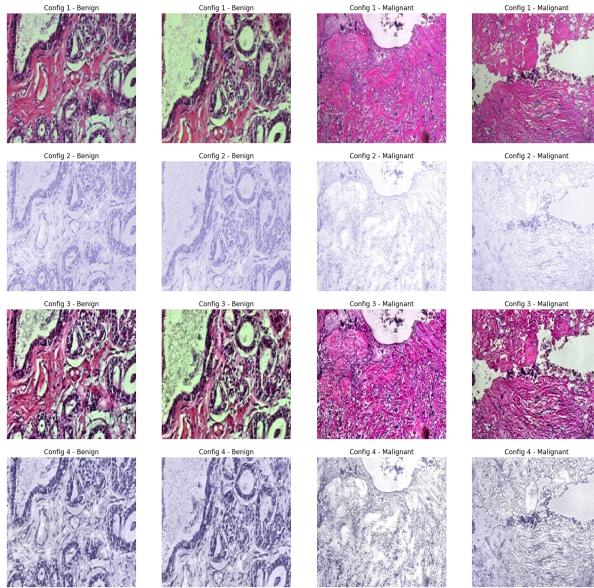


Fig. 2: Four CN techniques applied to HIs from BreakHis. Each configuration shows two benign and two malignant images.

While each configuration varies in contrast adjustment and feature emphasis, they maintain essential tissue properties, critical for diagnostic accuracy. The differences highlight the need for a pre-processing strategy that optimally balances feature enhancement with clinical detail preservation.

## 4 Results

Training curves in Figure 3 provide a detailed understanding of each configuration's learning process. (1) and (0) offer the best performance, while configurations involving CD, i.e., (2) and (4) show instability and potential over-fitting.

Figure 4 presents the ROC-AUC curves for the configurations tested. (1) and (0) achieve the highest performances with AUCs of 0.995 and 0.997 respectively, closely followed by (3) with an AUC of 0.977. These methods enhance the model's ability to differentiate between benign and malignant HIs effectively. On the other hand, (4) shows the lowest performance, highlighting potential issues with this approach.

According to Figure 5, (0), (1), and (3) demonstrate their effectiveness in accurately identifying malignant cases while maintaining low false positive rates. (2), while still performing well, shows a slight trade-off between precision and recall. (4), however, shows a notable decline in performance, suggesting that this combined approach may not be as effective.

The performance of the proposed configurations is summarised in Table 1. The high accuracy of (1) of 96.93% and F1-score of 97.79% indicate a well-balanced model with good precision (97.18%) and recall (98.42%) (Figure 5). The model exhibits strong performance in correctly identifying both positive and negative cases, as reflected by its high specificity of 93.62%.

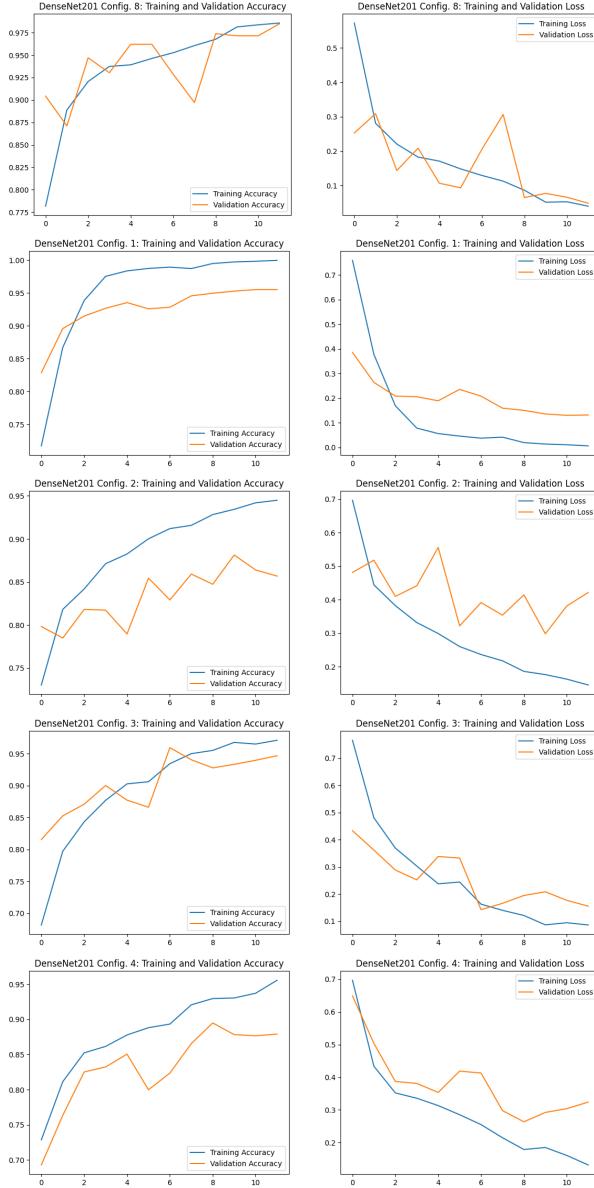


Fig. 3: Training performance metrics for configurations 0 through 4. Each row displays the training and validation accuracy on the left and the corresponding training and validation loss on the right.

(2) performs well but with a noticeable drop in specificity (84.96%) compared to (1). The accuracy of 92.13% and F1-score of 94.36% suggest a slightly less balanced model than (1), particularly in its ability to correctly identify true negatives, leading to a higher rate of false positives. (3) shows improvements over (2) with a higher accuracy of 93.25% and F1-score of 95.18%. The precision (94.05%) and recall (96.38%) in Figure 5 are also better, but the specificity remains lower than (1) at 86.28%. This indicates a balanced performance with good detection capabilities but it still struggles slightly with false positives.

(4) exhibits a significant drop in overall performance. The accuracy of 74.43% and F1-score of 83.58% are lower, and the specificity of 30.70% is particularly poor. While the recall is high at 94.07%, indicating good detection of true positives, the low specificity indicates a high rate of false positives. (0), however, shows the best overall performance with the highest metrics: the accuracy of 98.14% and F1-score of 98.65% indicate exceptional model performance. The precision (98.45%) and recall (98.85%) are also the highest among all configurations, and the specificity of 96.53% reflects a well-balanced model. The ROC-AUC of 99.67% further confirms its superior discrimination ability.

Figure 6 offer a granular look at each configuration's predictive performance. It shows (0), (1), and (3) perform well in terms of true positives and true negatives, indicating their effectiveness in accurately classifying HIs. (2), while still performing well, shows a slight trade-off with increased false positives

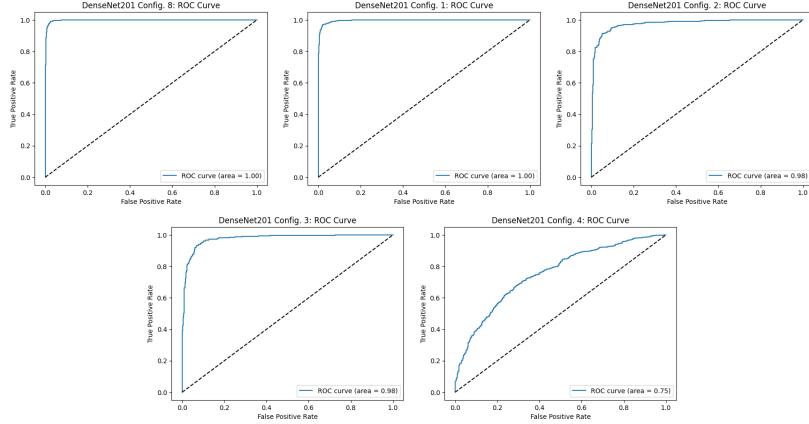


Fig. 4: ROC-AUC curves.

Table 1: Performance metrics for different configurations.

Config.	Acc.	Prec.	Recall	Spec.	F1	ROC-AUC
0	0.9814	0.9845	0.9885	0.9653	0.9865	0.9967
1	0.9693	0.9718	0.9842	0.9362	0.9779	0.9945
2	0.9213	0.9339	0.9535	0.8496	0.9436	0.9697
3	0.9325	0.9405	0.9638	0.8628	0.9518	0.9773
4	0.7443	0.7529	0.9407	0.3070	0.8358	0.7824

and negatives. (4), however, shows a notable decline in performance, suggesting that such a combined approach may not be effective.

In addition to the configurations presented in this study, we also experimented with two other well-known CN techniques, i.e., Reinhard [78] and Macenko [59] methods. However, these methods yielded poorer results compared to the techniques we ultimately focused on. Reinhard’s normalisation results were highly problematic. With a perfect recall at 100.00%, the model identified all malignant cases, but the specificity of nearly zero at 0.20%, indicated it classifies nearly all samples as positive, leading to a very high false positive rate. The accuracy (69.07%) and F1-score (81.70%) were also notably low. Macenko’s normalisation showed slightly improved metrics compared to Reinhard’s but still exhibits poor specificity (11.42%). The accuracy of 70.67% and F1-score of 82.09% indicate an imbalanced model with a high false positive rate. The Macenko’s normalisation results mirrored Reinhard’s in terms of recall (100.00%) and specificity (0.00%), confirming that without effective CN, the model struggled significantly with specificity and overall performance.

*Note on Limitations and the Impact of Stratified Split* During our extensive cross-validation analysis, several limitations of CN techniques became apparent. While CN methods are intended to standardise the colour distribution across HIs, they did not uniformly enhance model performance. One significant limitation observed was the introduction of artefacts during the normalisation process, particularly with more advanced techniques like CD combined with CLAHE (as in (4)). Indeed, such artefacts can obscure critical histological features, leading to decreased model performance as evidenced by lower specificity and ROC-AUC scores.

Another limitation is the potential for over-fitting to normalised features. Techniques that aggressively normalise colour, such as CD, can cause the model to focus too narrowly on specific colour-related patterns, which may not generalise well across different subsets of the data. This was particularly evident in (4), where despite enhanced contrast and feature visibility, the model’s overall performance suffered due to over-reliance on these artificially enhanced features.

The use of a stratified split was crucial in maintaining class balance during our experiments. By ensuring that each fold in the cross-validation process had a proportional representation of both malignant and benign samples, we aimed to provide a robust evaluation of the model’s performance. However, the stratified split alone could not eliminate variability in colour distribution across the different subsets of data. Variations in staining protocols and slide preparation techniques still introduced significant colour differences, which the CN techniques struggled to standardise completely. This variability led to

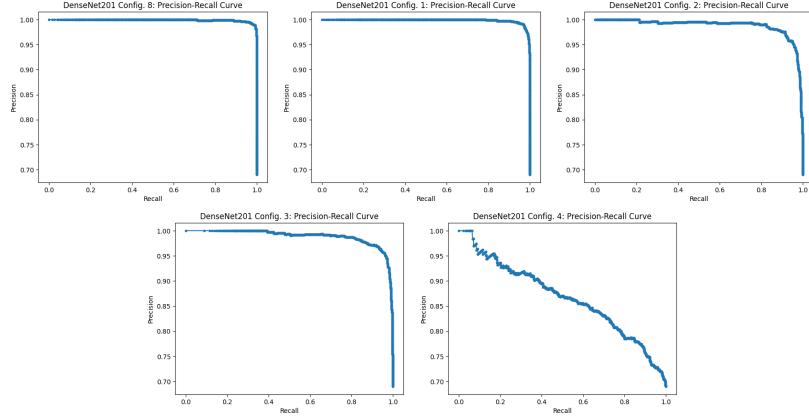


Fig. 5: Precision-Recall curves.

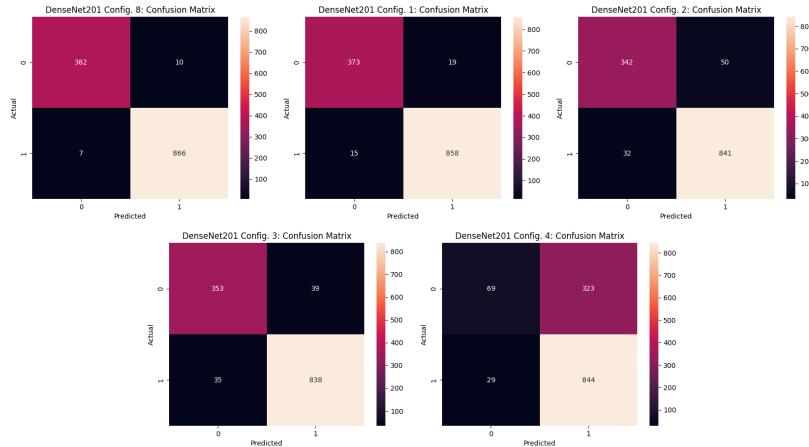


Fig. 6: Confusion matrices.

inconsistent model performance across different folds, highlighting the limitations of CN methods in fully addressing colour heterogeneity in HIs.

*Data Augmentation as a Mitigation Strategy* To mitigate the issues observed with CN techniques, DA emerged as a promising strategy. By introducing a variety of transformations to the training images, including rotations, shifts, and colour adjustments, DA helped the model become more robust to variations in colour and other image attributes. This robustness is crucial for improving the generalisability of the model. Our findings demonstrated that models trained with DA only, i.e., (0) outperformed all CN techniques, achieving the highest metrics across the board. This suggests that DA alone can effectively simulate the diversity of real-world data, helping the model to learn more invariant features that are not overly dependent on specific colour distributions. The ability to handle diverse staining variations without explicit CN makes DA a valuable tool in HIA.

## 5 Conclusion, Limitations and Future Directions

The cross-validation analysis of four CN techniques and DA provided several key insights into their impact on the performance of the DenseNet201 model in HI classification. The findings challenge the conventional reliance on CN, demonstrating that modern DL models, coupled with robust DA, can effectively manage colour variability without extensive pre-processing. This study provides empirical evidence that DA alone can enhance model performance, streamline pre-processing pipelines, reduce computational costs, and improve scalability. More specifically:

1. Simple channel-based normalisation achieved high accuracy and robust performance across all metrics, indicating its effectiveness in standardising the colour distribution without introducing significant variability. The model achieved an accuracy of 96.93%, precision of 97.18%, recall of 98.42%,

- 
- specificity of 93.62%, F1-score of 97.79%, and ROC-AUC of 0.995. These results suggest that simple channel-based normalisation effectively balances sensitivity and specificity.
2. CD with the Hematoxylin component resulted in lower performance, with an accuracy of 92.13% and specificity of 84.96%. Although it enhanced key diagnostic features by isolating the Hematoxylin component, it introduced variability that affects the model's ability to generalise, resulting in a higher false positive rate.
  3. Channel-based normalisation combined with CLAHE provided a slight improvement over CD alone, achieving an accuracy of 93.25%, precision of 94.05%, recall of 96.38%, specificity of 86.28%, F1-score of 95.18%, and ROC-AUC of 0.977. despite being able to enhance image contrast and visibility of critical features, it still falls short compared to simple channel-based normalisation.
  4. CD with the Hematoxylin component combined with CLAHE showed the lowest performance, with an accuracy of 74.43% and specificity of 30.70%. Despite enhancing certain features, it introduced significant variability, adversely impacting the model's specificity and leading to a high false positive rate.
  5. The model trained with DA alone, without any CN, outperformed all CN techniques, achieving an accuracy of 98.14%, precision of 98.45%, recall of 98.85%, specificity of 96.53%, F1-score of 98.65%, and ROC-AUC of 0.997.

By documenting these attempts and their outcomes, the study provides a comprehensive view of the experimental process and the rationale behind the final methodology selection. This thorough approach contributes to a deeper understanding of the challenges and complexities involved in HIA, guiding future research in the field.

The key message here is that, while various CN techniques have their merits, they should not be seen as a panacea for the challenges posed by colour variability in HIs. Moreover, the stratified split ensures class balance but does not fully address colour distribution variability. DA, by introducing diverse transformations, can enhance model robustness and generalisability, offering a viable alternative to traditional CN methods. Hence, this study underscores the importance of a balanced approach, combining different pre-processing strategies to optimise model performance and reliability in HIA.

Despite these insights, several limitations warrant consideration. The study utilised the BreakHis dataset, which may not capture the full spectrum of variability found in other HIs. The findings are based on the DenseNet201 model; however, different architectures may respond differently to CN techniques. The computational cost and training time were substantial, potentially limiting the breadth of configurations explored. Cross-validation mitigated over-fitting, but the potential remains, particularly with small validation subsets. A further more detailed analysis of specific DA strategies and an external validation phase with independent datasets would further validate these findings. Lastly, the clinical applicability of these techniques needs further exploration to assess their impact on diagnostic decision-making.

Such limitations underscore the need for continued research. The presented research should be considered but a foundational experiment that opens numerous avenues for further exploration and refinement. Future studies should expand the scope of techniques evaluated, include multiple model architectures, and ensure thorough external validation. By addressing these limitations, the field can move towards more robust and clinically applicable solutions for HIA.

Indeed, the systematic cross-validation of CN techniques to robustly compare their effectiveness across multiple data splits need to be prioritised. Exploring dynamic adjustments of CN methods and extending evaluations to broader histopathology datasets are crucial next steps. These advancements will further refine CNN applications in digital pathology, enhancing diagnostics and potentially improving patient outcomes.

## 6 Disclosure of Interests

The author has no competing interests to declare.

## References

1. Abbas, A., Abdelsamea, M.M., Gaber, M.M.: 4s-dt: self-supervised super sample decomposition for transfer learning with application to covid-19 detection. *IEEE Transactions on Neural Networks and Learning Systems* **32**(7), 2798–2808 (2021)

- 
2. Alshdaifat, E., Doaa, A., Alsarhan, A., Hussein, F., El-Salhi, S.: The effect of pre-processing techniques, applied to numeric features, on classification algorithms' performance (Jan 2021)
  3. Anglada-Rotger, D., Marques, F., Pardas, M.: Color deconvolution applied to domain adaptation in her2 histopathological images. In: IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW). pp. 1–5. IEEE (2023)
  4. Aosiman, A., Abulijiang, A., Abulkem, A., Abdulkirim, T., Maimaiti, M.: Medical image enhancement algorithm based on histogram equalisation and dyadic wavelet transform (May 2020)
  5. Araujo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polonia, A., Campilho, A.: Classification of breast cancer histology images using convolutional neural networks. *PLoS one* **12**(6), e0177544 (2017)
  6. Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polónia, A., Campilho, A.: Classification of breast cancer histology images using convolutional neural networks. *PLoS ONE* **12**(6), e0177544 (2017)
  7. Bai, B., Yang, X., Li, Y., Zhang, Y., Pillar, N., Ozcan, A.: Deep learning-enabled virtual histological staining of biological samples (Mar 2023)
  8. Bao, M., Wu, R., Yao, Z., Lu, X., Zhang, H., Zhao, W., Ma, Y.: Erunet: cell segmentation based on multi-scale features and mobile inverted bottleneck convolution. *SSRN* 4756742
  9. BenTaieb, A., Hamarneh, G.: Adversarial stain transfer for histopathology image analysis. *IEEE Transactions on Medical Imaging* **37**(3), 792–802 (2018)
  10. Bhat, H., Kanakatte, A., Nayak, R., Gubbi, J.: A hybrid approach for nucleus stain separation in histopathological images. In: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 1218–1221. IEEE (2017)
  11. Boschman, J., Farahani, H., Darbandsari, A., Ahmadvand, P., Van Spankeren, A., Farnell, D., Levine, A.B., Naso, J.R., Churg, A., Jones, S.J., Kobel, M., Huntsman, D.G., Gilks, C.B., Bashashati, A.: The utility of colour normalisation for ai-based diagnosis of hematoxylin and eosin-stained pathology images. *Journal of Pathology* **256**, 15–24 (2022)
  12. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I.: Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer Journal for Clinicians* (2024)
  13. Chlap, P., Min, H., Vandenbergh, N., Dowling, J., Holloway, L., Haworth, A.: A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology* **65**(5), 545–563 (2021)
  14. Chrobociński, K.: Nuclei detection in images of hematoxylin and eosin-stained tissues using normalisation of value channel in hsv colour space. In: Innovations and Developments of Technologies in Medicine, Biology and Healthcare: IEEE EMBS International Student Conference (ISC). pp. 8–17. Springer (2022)
  15. Cireşan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Mitosis detection in breast cancer histology images with deep neural networks. In: 16th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 411–418. Springer (2013)
  16. Cui, M., Zhang, D.Y.: Artificial intelligence and computational pathology. *NPJ Precision Oncology* **5**(1), 1–15 (2021)
  17. Dan, Q., Xu, Z., Burrows, H., Bissram, J., Stringer, J.S., Li, Y.: Diagnostic performance of deep learning in ultrasound diagnosis of breast cancer: a systematic review. *NPJ Precision Oncology* **8**(1), 21 (2024)
  18. Dang Vu, Q., Jewsbury, R., Graham, S., Jahanifar, M., Raza, S., Minhas, F., Bhalerao, A., Rajpoot, N.: Nuclear segmentation and classification: on colour & compression generalisation. *arXiv e-prints* pp. arXiv–2301 (2023)
  19. Das, Saini, S., Kataria, P., Dipanshu, D.: Breast cancer detection from histopathological images using machine learning models. *International Journal of Health Sciences* **6**(S3), 9542–9553 (2022)
  20. Dawood, F.: The importance of contrast enhancement in medical images analysis and diagnosis (Dec 2018)
  21. Eilers, P., Ruckebusch, C.: Fast and simple super-resolution with single images (Jul 2022)
  22. Elazab, N., Soliman, H., El-Sappagh, S., Islam, S.M., Elmogy, M.: Objective diagnosis for histopathological images based on machine learning techniques: classical approaches and new trends. *Mathematics* **8**(11), 1863 (2020)
  23. Elmore, J.G., Longton, G.M., Carney, P.A., Geller, B.M., Onega, T., Tosteson, A.N., Nelson, H.D., Pepe, M.S., Allison, K.H., Schnitt, S.J., et al.: Diagnostic concordance among pathologists interpreting breast biopsy specimens. *Jama* **313**(11), 1122–1132 (2015)
  24. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**(7639), 115–118 (2017)
  25. Fabian, Z., Heckel, R., Soltanolkotabi, M.: Data augmentation for deep learning based accelerated mri reconstruction with limited data. In: International Conference on Machine Learning. pp. 3057–3067. PMLR (2021)
  26. Garcea, F., Serra, A., Lamberti, F., Morra, L.: Data augmentation for medical imaging: a systematic literature review. *Computers in Biology and Medicine* **152**, 106391 (2023)
  27. Ghaffar, N.N., Kaplanoglu, E., Nasab, A.: Evaluation of artificial intelligence techniques in disease diagnosis and prediction. *Discover Artificial Intelligence* **3**(1), 5 (2023)
  28. Go, H.: Digital pathology and artificial intelligence applications in pathology (Apr 2022)

- 
29. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: International Conference on Machine Learning. pp. 1321–1330. PMLR (2017)
30. Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., Yener, B.: Histopathological image analysis: a review. *IEEE Reviews in Biomedical Engineering* **2**, 147–171 (2009)
31. Haines, S., Eaton, E., Ali, M.L.: Machine learning models for histopathological breast cancer image classification. In: IEEE World AI IoT Congress (AIIoT). pp. 0036–0041. IEEE (2023)
32. Hamidinekoo, A., Zwiggelaar, R.: Stain colour normalisation to improve mitosis detection on breast histology images. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 3rd International Workshop, DLMIA, and 7th International Workshop, ML-CDS, Held in Conjunction with MICCAI. pp. 213–221. Springer (2017)
33. Hanahan, D.: Hallmarks of cancer: new dimensions. *Cancer Discovery* **12**(1), 31–46 (2022)
34. Herdiantoputri, R.R., Komura, D., Fujisaka, K., Ikeda, T., Ishikawa, S.: Deep texture representation analysis for histopathological images. *STAR Protocols* **4**(2), 102161 (2023)
35. Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L., Sun, Q.: Deep learning for image-based cancer detection and diagnosis - a survey. *Pattern Recognition* **83**, 134–149 (2018)
36. Huss, R., Coupland, S.E.: Software-assisted decision support in digital histopathology. *Journal of Pathology* **250**(5), 685–692 (2020)
37. Islam, T., Hafiz, M.S., Jim, J.R., Kabir, M.M., Mridha, M.: A systematic review of deep learning data augmentation in medical imaging: recent advances and future research directions. *Healthcare Analytics* p. 100340 (2024)
38. Istighosah, M., Sunyoto, A., Hidayat, T.: Breast cancer detection in histopathology images using resnet101 architecture. *Sinkron: Jurnal dan Penelitian Teknik Informatika* **8**(4), 2138–2149 (2023)
39. Janowczyk, A., Madabhushi, A.: Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. *Journal of Pathology Informatics* **7** (2016)
40. Jiang, Y., Chen, L., Zhang, H., Xiao, X.: Breast cancer histopathological image classification using convolutional neural networks with small se-resnet module. *PLoS ONE* **14**(3), e0214587 (2019)
41. Kandel, I., Castelli, M.: A novel architecture to classify histopathology images using convolutional neural networks. *Applied Sciences* **10**(8), 2929 (2020)
42. Khalifa, N.E., Loey, M., Mirjalili, S.: A comprehensive survey of recent trends in deep learning for digital images augmentation. *Artificial Intelligence Review* **55**(3), 2351–2377 (2022)
43. Khamankar, V., Bera, S., Bhattacharya, S., Sen, D., Biswas, P.K.: Histopathological image analysis with style-augmented feature domain mixing for improved generalisation. In: Celebi, M.E., et al. (eds.) *Medical Image Computing and Computer Assisted Intervention Workshop (MICCAI)*. Lecture Notes in Computer Science, vol. 14393. Springer, Cham (2023)
44. Khan, A.M., Rajpoot, N., Treanor, D., Magee, D.: A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Transactions on Biomedical Engineering* **61**(6), 1729–1738 (2014)
45. Khan, A., Sohail, A., Zahoor, U., Qureshi, A.S.: A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review* **53**, 5455–5516 (2020)
46. Khosla, C., Saini, B.S.: Enhancing performance of deep learning models with different data augmentation techniques: a survey. In: International Conference on Intelligent Engineering and Management (ICIEM). pp. 79–85. IEEE (2020)
47. Komura, D., Ishikawa, S.: Machine learning methods for histopathological image analysis. *Computational and Structural Biotechnology Journal* **16**, 34–42 (2018)
48. Krishna, S., Krishnamoorthy, S., Bhavsar, A., et al.: Stain normalised breast histopathology image recognition using convolutional neural networks for cancer detection. *arXiv preprint arXiv:2201.00957* (2022)
49. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Communications of the ACM* **60**(6), 84–90 (2017)
50. Lakshmanan, B., Anand, S., Jenitha, T.: Stain removal through colour normalisation of haematoxylin and eosin images: a review. *Journal of Physics: Conference Series* **1362**(1), 012108 (2019)
51. Landini, G., Martinelli, G., Piccinini, F.: Colour deconvolution: stain unmixing in histological imaging (Nov 2020)
52. Langlotz, C.P.: Will artificial intelligence replace radiologists? *Radiology: Artificial Intelligence* **1**(3), e190058 (2019)
53. Laxmisagar, H., Hanumantharaju, M.: A survey on automated detection of breast cancer based histopathology images. In: International Conference on Inventive Computation and Informatics (ICICI). pp. 808–813. IEEE (2020)
54. Levine, A.B., Peng, L.H., Balachander, S., Najarian, K.: Review of the current landscape of artificial intelligence in digital pathology & clinical applications. *Archives of Pathology & Laboratory Medicine* (2022)
55. Li, X., Plataniotis, K.N.: A complete colour normalisation approach to histopathology images using colour cues computed from saturation-weighted statistics. *IEEE Transactions on Biomedical Engineering* **62**(7), 1862–1873 (2015)
56. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., et al.: A survey on deep learning in medical image analysis. *Medical Image Analysis* **42**, 60–88 (2017)

- 
57. Litjens, G., Sánchez, C.I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., Hulsbergen-Van De Kaa, C., Bult, P., Van Ginneken, B., Van Der Laak, J.: Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific Reports* **6**(1), 26286 (2016)
58. Liu, Y.H.: Feature extraction and image recognition with convolutional neural networks. *Journal of Physics: Conference Series* **1087**(6), 062032 (2018)
59. Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.: A method for normalising histology slides for quantitative analysis. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. pp. 1107–1110. IEEE (2009)
60. Madusanka, N., Jayalath, P., Fernando, D., Yasakethu, L., Lee, B.: Impact of h&e stain normalisation on deep learning models in cancer image classification: performance, complexity, and trade-offs. *Cancers* **15**(16), 4144 (2023)
61. Mahbod, A., Dorffner, G., Ellinger, I., Woitek, R., Hatamikia, S.: Improving generalisation capability of deep learning-based nuclei instance segmentation by non-deterministic train time and deterministic test time stain normalisation. *Computational and Structural Biotechnology Journal* **23**, 669–678 (2024)
62. Maleki, F., Ovens, K., Gupta, R., Reinhold, C., Spatz, A., Forghani, R.: Generalisability of machine learning models: quantitative evaluation of three methodological pitfalls. *Radiology: Artificial Intelligence* **5**(1), e220028 (2022)
63. Matos, J.d., Ataky, S.T.M., Britto, A.d.S., Oliveira, L.S., Koerich, A.L.: Machine learning methods for histopathological image analysis: a review. *Electronics* **10**(5), 562 (2021)
64. Matos, J.D., Ataky, S.T., Britto, A.D., Oliveira, L.S., Koerich, A.L.: Machine learning methods for histopathological image analysis: a review. *Electronics* **10**(5), 562 (2021)
65. Mishra, J., Kumar, B., Targhotra, M., Sahoo, P.K.: Advanced and futuristic approaches for breast cancer diagnosis. *Future Journal of Pharmaceutical Sciences* **6**(1), 1–14 (2020)
66. Moxley-Wyles, B., Colling, R., Verrill, C.: Artificial intelligence in pathology: an overview. *Diagnostic Histopathology* **26**(11), 498–502 (2020)
67. Mumuni, A., Mumuni, F.: Data augmentation: a comprehensive survey of modern approaches. *Array* **16**, 100258 (2022)
68. Mun, S.K., Wong, K.K.H., Lo, S.B., Li, Y., Bayarsaikhan, S.: Artificial intelligence for the future radiology diagnostic service. *Frontiers in Molecular Biosciences* **7**, 614258 (2021)
69. Munir, K., Elahi, H., Ayub, A., Frezza, F., Rizzi, A.: Cancer diagnosis using deep learning: a bibliographic review. *Cancers* **11**(9), 1235 (2019)
70. Mzoughi, H., Njeh, I., Slima, M.B., Hamida, A.B.: Histogram equalisation-based techniques for contrast enhancement of mri brain glioma tumour images: comparative study. *International Conference on Advanced Systems and Electric Technologies (ICASET)* (2018)
71. Ng, A.Y., Oberije, C., Ambrozay, E., Szabo, E., Serfozo, O., Karpati, E., et al.: Prospective implementation of ai-assisted screen reading to improve early detection of breast cancer. *Nature Medicine* **29**(12), 3044–3049 (2023)
72. Nitya, K., Amberkar, V.S., Nadar, B.G.: Vital staining - pivotal role in the field of pathology. *Annals of Cytology and Pathology* **5**(1), 058–063 (2020)
73. Pallua, J.D., Brunner, A., Zelger, B., Schirmer, M., Haybaeck, J.: The future of pathology is digital. *Virchows Archiv* **477**(4), 509–516 (2020)
74. Pontalba, J.T., Gwynne-Timothy, T., David, E., Jakate, K., Androutsos, D., Khademi, A.: Assessing the impact of colour normalisation in convolutional neural network-based nuclei segmentation frameworks. *Frontiers in Bioengineering and Biotechnology* **7**, 300 (2019)
75. Prezja, F., Ayramo, S., Polonen, I., Ojala, T., Lahtinen, S., Ruusuvuori, P., Kuopio, T.: Improved accuracy in colorectal cancer tissue decomposition through refinement of established deep learning solutions. *Scientific Reports* **13**(1), 15879 (2023)
76. Raghu, M., Zhang, C., Kleinberg, J., Bengio, S.: Transfusion: understanding transfer learning for medical imaging. *Advances in Neural Information Processing Systems* **32** (2019)
77. Ramadan, S.Z.: Using convolutional neural network with cheat sheet and data augmentation to detect breast cancer in mammograms. *Computational and Mathematical Methods in Medicine* **2020** (2020)
78. Reinhard, E., Ashikhmin, M., Gooch, B., Shirley, P.: Colour transfer between images. *IEEE Computer Graphics and Applications* **21**(5), 34–41 (2001)
79. Reyes, H.E.Z., Barreto, H.P., Ramos, J.A.C., Armas, G.D.C.L.: Comparative analysis of stained normalisation in h&e histopathological images of breast cancer for nuclei segmentation improvement. In: *IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*. vol. 7, pp. 1–6. IEEE (2023)
80. Saraswat, M., Arya, K.: Colour normalisation of histopathological images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging Visualisation* **1**(4), 185–197 (2013)
81. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. *Journal of Big Data* **6**(1), 1–48 (2019)
82. Sigirci, I.O., Albayrak, A., Bilgin, G.: Detection of mitotic cells in breast cancer histopathological images using deep versus handcrafted features. *Multimedia Tools and Applications* **81**, 13179–13202 (2022)
83. Singh, J., Magudeeswaran, V.: A machine learning approach for brain image enhancement and segmentation. *International Journal of Imaging Systems and Technology* **27**(4), 334–341 (2017)

- 
84. Sirinukunwattana, K., Raza, S.E.A., Tsang, Y.W., Snead, D.R., Cree, I.A., Rajpoot, N.M.: Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Transactions on Medical Imaging* **35**(5), 1196–1206 (2016)
85. Sivaroopan, N., Watawana, H., Jayanga, C., Ekanayake, C., Rodrigo, R., Edussooriya, C.U., Wadduwage, D.N.: Contrastive deep encoding enables uncertainty-aware machine-learning-assisted histopathology. In: *Microscopy Histopathology and Analytics*. pp. MM3A–5. Optica Publishing Group (2024)
86. Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L.: A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering* **63**(7), 1455–1462 (2016)
87. Srinidhi, C.L., Ciga, O., Martel, A.L.: Deep neural network models for computational histopathology: a survey. *Medical Image Analysis* **67**, 101813 (2021)
88. Steiner, D.F., Macdonald, R.D., Liu, Y., Truszkowski, P., Hipp, J., Gammie, C., et al.: Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. *The American Journal of Surgical Pathology* **42**(12), 1636–1646 (2018)
89. Su, J., Yu, X., Wang, X., Wang, Z., Chao, G.: Enhanced transfer learning with data augmentation. *Engineering Applications of Artificial Intelligence* **129**, 107602 (2024)
90. Tabatabaei, Z., Pérez Bueno, F., Colomer, A., Moll, J.O., Molina, R., Naranjo, V.: Advancing content-based histopathological image retrieval pre-processing: a comparative analysis of the effects of colour normalisation techniques. vol. 14, p. 2063. MDPI (2024)
91. Tellez, D., Litjens, G., Bárdi, P., Bulten, W., Bokhorst, J.M., Ciompi, F., Van Der Laak, J.: Quantifying the effects of data augmentation and stain color normalisation in convolutional neural networks for computational pathology. *Medical Image Analysis* **58**, 101544 (2019)
92. Thyagaraj, T., Prasanna, K., Hariprasad, S.A.: Harnessing convolutional neural networks for histopathological breast cancer classification. *International Journal on Recent and Innovation Trends in Computing and Communication* **11**(9) (2023), ISSN: 2321-8169
93. Tormey, C.A., Gehrie, E.A., Pham, H.P., Bucy, R.P., Lorenz, R.G., Zheng, X., Hendrickson, J.E.: Data interpretation in laboratory medicine. In: *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics*, pp. 489–510. Elsevier (2018)
94. Jimenez-del Toro, O., Otalora, S., Andersson, M., Euren, K., Hedlund, M., Rousson, M., Muller, H., Atzori, M.: Analysis of histopathology images: From traditional machine learning to deep learning. In: *Biomedical Texture Analysis*. pp. 281–314. Elsevier (2017)
95. Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A.M., Esposito, I., Navab, N.: Structure-preserving colour normalisation and sparse stain separation for histological images. *IEEE Transactions on Medical Imaging* **35**(8), 1962–1971 (2016)
96. Veta, M., Pluim, J.P., van Diest, P.J., Viergever, M.A.: Breast cancer histopathology image analysis: a review. *IEEE Transactions on Biomedical Engineering* **61**(5), 1400–1411 (2014)
97. Vidhya, G.R., Ramesh, H.: Effectiveness of contrast limited adaptive histogram equalisation technique on multi-spectral satellite imagery. In: *10th International Conference on Soft Computing and Pattern Recognition* (2017)
98. Vu, H.: Integrating pre-processing methods and convolutional neural networks for effective tumour detection in medical imaging. arXiv preprint arXiv:2402.16221 (2024)
99. Xu, J., Hou, J., Zhang, Y., Feng, R., Ruan, C., Zhang, T., Fan, W.: Data-efficient histopathology image analysis with deformation representation learning pp. 857–864 (2020)
100. Xu, M., Yoon, S., Fuentes, A., Park, D.S.: A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognition* **137**, 109347 (2023)
101. Yee, W.C., Jian, T.X., Ab Rahman, K.S., Hoe, T.L., Min, L.J., Hang, Q.Y., Ling, T.C.: Performance analysis of colour normalisation methods in histopathology images. In: *IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*. pp. 147–151. IEEE (2022)
102. Yıldız, S., Memiş, A., Varlı, S.: Nuclei segmentation in colon histology images by using the deep cnns: a u-net based multi-class segmentation analysis. In: *Medical Technologies Congress (TIPTEKNO)*. pp. 1–4. IEEE (2022)
103. Yu, H., Yang, L.T., Zhang, Q., Armstrong, D., Deen, M.J.: Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives. *Neurocomputing* **449**, 412–450 (2021)
104. Yıldırım, Z., Hançer, E., Samet, R., Mali, M.T., Nemati, N.: Effect of colour normalisation on nuclei segmentation problem in he stained histopathology images. In: *30th Signal Processing and Communications Applications Conference (SIU)*. pp. 1–4. IEEE (2022)
105. Zheng, Y., Jiang, Z., Zhang, H., Xie, F., Shi, J., Xue, C.: Adaptive colour deconvolution for histological wsi normalisation. *Computer Methods and Programs in Biomedicine* **170**, 107–120 (2019)
106. Zhou, K., Greenspan, H., Davatzikos, C., Duncan, J., Van Ginneken, B., Madabhushi, A., Prince, J., Rueckert, D., Summers, R.: A review of deep learning in medical imaging: imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE* **109**(5), 820–838 (2021)
107. Önder, D., Zengin, S., Sarıoğlu, S.: A review on colour normalisation and colour deconvolution methods in histopathology. *Pathology - Research and Practice* **210**(11), 713–719 (2014)
108. Štifanic, J., Štifanić, D., Zulijani, A., Car, Z.: Application of ai in histopathological image analysis. In: *Serbian International Conference on Applied Artificial Intelligence*. pp. 121–131. Springer (2022)