Enhancing medical image classification with adaptive meta-learning and reinforcement learning-based data augmentation

Natalya Smith[1]

Faculty of Design and Creative Technologies,
School of Computer and Mathematical Sciences
Auckland University of Technology, Auckland, New Zealand
mnb8479@autuni.ac.nz

Abstract. The increasing demand for computer vision techniques has led to significant advancements in medical imaging. Image processing, a subset of computer vision, utilises algorithms for object recognition. Because manual screening for cancer is labour-intensive and prone to errors, better classification techniques are required to enhance diagnosis accuracy. With this in mind, this paper proposes the use of convolutional neural networks (CNNs) to classify histopathology images (HIs) of breast tissue into two classes (benign and malignant) using the BreakHis dataset. To optimise data augmentation (DA) policies and enhance model performance, this study incorporates a reinforcement learning (RL) component. The primary objective is to build and train high-performance classification models that are both sensitive and specific.
The study evaluates three CNN architectures, including ResNet50, DenseNet201, and EfficientNetB0, using stratified k-fold cross-validation to ensure robustness and generalisability. Advanced DA techniques are applied to improve the diversity of the training data. The integration of an RL agent dynamically adjusts augmentation policies based on model performance, aiming to find the optimal strategies that maximise model accuracy and robustness. The results demonstrate how DenseNet201 exhibits the highest overall performance, achieving the highest average accuracy of 0.981, precision of 0.9872, recall of 0.986, F1-score of 0.9865, and ROC-AUC of 0.9964, highlighting its superior capability to generalise across different data variations. This suggest RL-based DA significantly enhances the performance and robustness of DL models in the binary classification of breast cancer using HIs.

Keywords: Histopathological Image Analysis · Cancer Detection · Convolutional Neural Networks · Data Augmentation · Reinforcement Learning · Medical Imaging

1 Introduction

Breast cancer is a significant global health concern, accounting for 11.6% of the 19,964,811 cancer cases reported worldwide in 2022 [11]. Its early detection is critical for effective treatment and improved patient outcomes [62]. The use of histopathology images (HIs), which are microscopic images of tissue samples, plays a pivotal role in diagnosing breast cancer. The reliability of diagnoses using HIs, however, can vary considerably based on the quality of the tissue samples and the experience of the pathologist. Indeed, the overall concordance rate among pathologists diagnosing cancer through HIs is around 75.3% [33], with significant inconsistencies in specific breast cancer types.

Recent advancements in deep learning (DL) show promise in automating and enhancing the accuracy of histopathological image analysis (HIA), potentially leading to more reliable and rapid diagnoses. Despite numerous studies leveraging DL for HIA, there is still a need for innovative approaches to improve model robustness and generalisation. Many studies focus on static data augmentation (DA) techniques to enhance training data diversity, as surveyed, e.g., in [82, 88]. However, these methods may not fully adapt to the changing requirements of the training process, as pointed out, e.g., by [18], who highlight the limitations of static DA in adapting to different stages of model training.

Medical datasets, particularly those used in HIA, often face limitations such as small sample sizes and class imbalances. These issues can lead to over-fitting and poor generalisation of DL models. To address these challenges, dynamic DA can be employed to create more diverse and representative training datasets. Dynamic DA helps by automating the selection and application of augmentation techniques, such as zooming, rotating, flipping, and more, which otherwise require researchers to manually determine their effectiveness.

This paper aims to bridge this gap by incorporating a reinforcement learning (RL) component to dynamically adjust DA policies based on real-time model performance. While RL has been used in other contexts, its application to dynamically adjusting DA in HIA is relatively novel. Although there is potential for RL in automating DA [18, 38], studies that focus specifically on using RL for DA in the context of HIs or breast cancer detection are yet to be found.

The primary research question addressed in this study is: *How can RL-based DA improve the performance and robustness of DL models in the binary classification of breast cancer using HIs?*

This paper contributes to the filed by:

- Implementing and evaluating three most commonly used convolutional neural network (CNN) architectures (ResNet50, DenseNet201, and EfficientNetB0) to determine their efficacy in classifying HIs.
- Utilising stratified five k-fold cross-validation to ensure the robustness and generalisability of the developed models. This method divides the dataset into multiple folds, training and validating the model on different subsets to mitigate over-fitting and bias.
- Applying advanced DA techniques to enhance the diversity of the training data, thereby improving the model's ability to generalise to unseen data.
- Incorporating a RL agent to dynamically adjust augmentation policies based on model performance. This approach aims to find the optimal augmentation strategies that maximise the model's accuracy and robustness.
- Addressing the issue of class imbalance, which is prevalent in medical datasets, by computing and incorporating class weights during the training process.
- Evaluating the models using a comprehensive set of performance metrics, including accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (ROC-AUC), to provide a detailed assessment of their diagnostic capabilities.

This study is unique in several ways:

1. Unlike traditional static DA techniques, it implementes RL to dynamically adjust augmentation strategies based on real-time model performance. This approach allows for the optimisation of augmentation policies, leading to improved model robustness and generalisation.
2. By evaluating three distinct CNN architectures, it provides a comparative analysis that highlights the strengths and weaknesses of each model in the context of HIA.
3. Not only it looks at traditional performance metrics like accuracy but also provided a detailed analysis using precision, recall, specificity, F1-score, and ROC-AUC. This comprehensive evaluation helpes in understanding the nuanced performance of each model.
4. The application of these advanced models to HIs for breast cancer detection is a critical step towards improving diagnostic accuracy in medical imaging. The study's findings can be instrumental in guiding future research and clinical practice.

The findings reveal that DenseNet201 exhibits the highest overall performance, followed by ResNet50 and EfficientNetB0. DenseNet201 achieved the highest average accuracy (0.9812), precision (0.9872), recall (0.9858), F1-score (0.9865), and ROC-AUC (0.9964), reflecting its superior capability to generalise across different data variations. ResNet50 and EfficientNetB0 also performed robustly, with ResNet50 showing an average accuracy of 0.9603 and ROC-AUC of 0.9899, and EfficientNetB0 achieving an average accuracy of 0.9526 and ROC-AUC of 0.9924. These results suggest that RL-based DA significantly enhances the performance and robustness of DL models in the binary classification of breast cancer using HIs.

The outcomes of this research are expected to contribute to the ongoing efforts in leveraging artificial intelligence (AI) for medical image analysis. By identifying the most effective DL models and techniques for breast cancer classification, and by incorporating RL to optimise DA, this paper aims to support the development of automated diagnostic tools that can assist pathologists in making accurate and timely diagnoses, ultimately improving patient care and treatment 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 experimental results, which is followed by the conclusion and future perspectives in Section 5.

2 Related work

2.1 HIA and the role of DL in medical imaging

Histopathology, a crucial field in medical diagnostics, plays a vital role in the diagnosis and treatment of breast cancer [34, 74]. Histopathological analysis (HA) of breast biopsy specimens is considered the gold standard for confirming the presence of breast cancer [62, 102]. Defined as the microscopic examination of diseased tissue samples, histopathology provides invaluable insights into the underlying cellular and structural changes that characterise various pathological conditions, including breast cancer [5, 74, 89]. The significance of histopathology in breast cancer management cannot be overstated.

Digital pathology involves the digitisation of tissue samples, allowing for the storage, retrieval, and analysis of these images using specialised software [32, 43, 81]. Pathologists, highly trained specialists in the analysis of tissue samples, utilise a range of techniques, such as staining and microscopic examination, to identify the unique features and patterns associated with different types of breast cancer. This detailed analysis allows for the accurate classification of the disease, which is essential for determining the most appropriate course of treatment. Traditionally, the process of interpreting HIs has relied heavily on the expertise and subjective judgments of pathologists [26, 101]. However, manual interpretation of these images 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 [94, 100].

These challenges have led to the development of advanced analytical techniques, including the use of digital pathology and machine learning (ML) algorithms [19, 61, 75], which aim to enhance the accuracy and efficiency of HA. ML algorithms, trained on large datasets of HIs, can assist pathologists in the identification of specific cellular and structural patterns, thereby reducing the time and subjectivity associated with manual interpretation [26, 56, 72, 81, 93]. Moreover, the integration of these technologies has the potential to improve not only the diagnostic accuracy but also the consistency of breast cancer diagnosis and, ultimately, the effectiveness of treatment strategies [20, 79].

Recent research efforts have focused on the application of ML to medical imaging, specifically in the context of breast cancer diagnosis using HIs. Traditional ML models have garnered strong attention, as evidenced by the works of [6,22,102]. This integration of ML has shown significant promise in the realms of medical image diagnosis. Several studies, including those by [31, 40, 54, 78], have highlighted the potential and advancements in this field. DL, a subset of ML, has revolutionised numerous fields by enabling the creation of models that can learn from vast amounts of data. This technology, characterised by neural networks with many layers, excels at identifying patterns and making predictions, often surpassing human performance in specific tasks. DL's transformative impact is particularly evident in fields such as natural language processing, speech recognition, autonomous driving, and, importantly, medical imaging.

In medical imaging, DL has shown remarkable potential to enhance diagnostic accuracy and efficiency [86, 114]. Traditional diagnostic methods often rely on the expertise of radiologists and pathologists, who manually interpret medical images. However, DL models, particularly CNNs, have demonstrated their ability to assist and even outperform human experts in certain diagnostic tasks [20]. These models analyse medical images to detect abnormalities, classify diseases, and predict patient outcomes, thus offering a powerful tool to support medical professionals.

The advancements in computational methods, particularly DL, have shown great promise in enhancing the diagnostic accuracy and efficiency of breast cancer detection using HIs [13, 26, 70, 72, 81]. DL algorithms have demonstrated the ability to automatically extract and analyse relevant features from HIs, often outperforming traditional, hand-crafted feature-based approaches [102]. These techniques can potentially alleviate the workload of pathologists and improve the reproducibility and reliability of breast cancer diagnoses [21, 102].

The adoption of DL techniques has brought a revolutionary shift in the utilisation of ML for medical imaging, particularly in the detection of breast cancer using HIs. Pioneering research by [15,17,91] among others has emphasised the effectiveness of CNNs in accurately identifying cancerous tissues. Research shows the proficiency of CNNs in classifying breast cancer, aligning closely with the objectives of this paper. These foundational studies have paved the way for ongoing exploration and advancements in this domain.

2.2 Basics of CNNs and their suitability in HIA

As a class of DL models specifically designed for image analysis, CNNs have gained prominence due to their exceptional performance in tasks such as image classification [9,106], object detection [1,29,42], and image segmentation [85, 99]. CNNs consist of several layers, each with a distinct function in processing

visual information. Convolutional layers apply filters to the input image to extract features, generating feature maps that highlight various aspects of the input, such as edges, textures, and shapes [80]. These layers capture spatial hierarchies in the data, making them highly effective for image analysis [64, 105].

Following the convolutional layers, pooling layers reduce the spatial dimensions of the feature maps using techniques like max pooling or average pooling [41]. This process helps in down-sampling the data, reducing computational complexity, and mitigating over-fitting by retaining the most significant features [113]. Activation functions introduce non-linearity into the network, allowing it to learn complex patterns [64, 105]. The most commonly used activation function in CNNs is the Rectified Linear Unit (ReLU), which outputs the input if it is positive and zero otherwise, thus accelerating the convergence of the training process and mitigating the vanishing gradient problem [35].

After several convolutional and pooling layers, the feature maps are flattened and fed into fully connected layers, which perform the final classification based on the extracted features [64, 105]. The output is typically a probability distribution over the possible classes, achieved using a softmax function for multi-class classification or a sigmoid function for binary classification.

CNNs have revolutionised HA, excelling in processing complex visual data in stained tissue biopsies by autonomously recognising intricate diagnostic patterns, thus eliminating the need for manual feature extraction [7, 50, 68, 92, 110]. They effectively classify diverse tissue types, delineate specific histological structures such as tumour cells, and quantify areas of interest [50,71]. Their resilience to variable factors in histological imaging, including staining techniques and sample preparation, ensures uniform analyses [71, 92]. Moreover, CNNs' rapid processing capabilities surpass traditional methods, transforming operational workflows for pathologists [77, 110].

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 [47, 50, 110]. The scope of CNN applications has broadened across medical imaging, with early research demonstrating their superior performance in image classification challenges [16, 46, 59, 67, 90]. Studies on breast cancer detection utilising architectures such as EfficientNet, ResNet101, and DenseNet201 underscore their effectiveness in managing complex HIs [58, 91].

Tasks like nuclei segmentation benefit from CNNs' ability to differentiate various nuclear types and background areas effectively [37, 116]. Furthermore, the extraction of deep texture features from CNNs facilitates objective analysis of tumour HIs, enhancing visualisation, retrieval, and supervised learning applications [109]. Overall, CNNs' ability to analyse complex imaging data accurately and efficiently has led to significant strides in diagnostic precision within HA.

CNNs are trained using large labelled datasets through a process called backpropagation, which adjusts the weights of the network to minimise the difference between the predicted and actual outputs [64,80,105]. The optimisation is usually performed using gradient descent-based algorithms, with common variants including Stochastic Gradient Descent and Adam. CNNs are particularly well-suited for image analysis due to their ability to automatically and adaptively learn spatial hierarchies of features from input images [9]. Unlike traditional machine learning models, which require manual feature extraction, CNNs learn to identify relevant features directly from the raw pixel values [63]. This capability is crucial in medical imaging, where subtle variations in texture, shape, and structure can be indicative of different pathological conditions.

In medical imaging applications, CNNs have been employed to tackle a wide range of tasks, including:

- Image classification [9,106]: Identifying the presence or absence of specific diseases in medical images, such as detecting tumours or classifying skin lesions as benign or malignant.
- Object detection [1, 29, 42]: Locating and identifying specific structures within an image, such as detecting and delineating tumours in MRI scans.
- Image segmentation [85,99]: Partitioning an image into meaningful regions, such as segmenting organs or lesions in diagnostic's scans for precise measurement and analysis.

2.3 Comparative analysis of CNN architectures for HIA

The field of HIA has seen significant advancements with the introduction of various CNN architectures. Each architecture brings unique features and strengths suitable for different aspects of image analysis tasks such as tumour detection, classification, and segmentation. Comparing prominent CNN architectures highlights their key characteristics, advantages, and limitations in the context of HIA. This provides insights into the performance and applicability of these architectures for specific tasks, guiding researchers and practitioners in selecting the most appropriate model for their needs.

ResNet (Residual Networks), introduced by Microsoft Research, revolutionised DL by introducing residual connections that mitigate the vanishing gradient problem, enabling the training of very deep networks [42]. ResNet has shown excellent performance in various HIA tasks, including tumour detection [4], classification [8], grading of cancer severity [25], and segmentation of tissue components [44]. The key strength of ResNet lies in its residual connections that allow gradients to flow directly through the network, facilitating the training of deeper architectures and improving performance [24, 36, 111]. Moreover, ResNet architectures, such as ResNet50 and ResNet101, can be scaled to very deep networks without the degradation problem [42]. Despite the residual connections, very deep ResNet models can still be computationally intensive [12]. The performance of ResNet can be sensitive to the choice of hyperparameters and the specific implementation details [57].

DenseNet (Densely Connected Networks), proposed by researchers at the University of Toronto and Tsinghua University, introduces dense connections where each layer receives input from all preceding layers, enhancing feature reuse and reducing the number of parameters [3, 49, 55]. DenseNet has achieved state-of-the-art results in HIA, including cancer detection and classification tasks [87]. Its efficiency and improved learning capabilities make it a preferred choice for many researchers. Dense connections improve feature propagation and reuse, leading to more efficient learning. Despite being deep, DenseNet has fewer parameters than comparable architectures, reducing the risk of over-fitting [42,65]. However, the dense connectivity can introduce computational overhead due to the concatenation operations [55,65,97]. Implementing and optimising DenseNet can be more complex compared to simpler architectures [115].

EfficientNet, developed by Google AI, employs a compound scaling method that uniformly scales network depth, width, and resolution, achieving better performance with fewer parameters [96–98]. EfficientNet has demonstrated excellent performance in HIA tasks [48, 104], often outperforming other architectures in terms of both accuracy and computational efficiency [103]. Its scalability and efficiency make it suitable for deployment in resource-constrained environments. EfficientNet strikes a fine balance between accuracy and computational efficiency, achieving high performance with fewer parameters and reduced computational cost. Its compound scaling method allows EfficientNet to scale effectively across varying computational budgets [96]. However, this method can be complex to implement and necessitates careful tuning of scaling coefficients [23]. EfficientNet also heavily relies on pre-training on large datasets to attain optimal performance in specific domains such as histopathology [104].

Each of these architectures has its own strengths and weaknesses. For example, ResNet, with its residual connections, enables the training of very deep networks, leading to significant performance improvements, but at the cost of increased complexity. DenseNet's dense connectivity improves learning efficiency and reduces the number of parameters, making it effective for histopathological tasks. EfficientNet's compound scaling method achieves a balance between accuracy and efficiency, making it a versatile choice for various applications. The choice of CNN architecture for HIA, therefore, depends on the specific requirements of the task, including the available computational resources, the size and diversity of the dataset, and the need for model interpretability.

2.4 DA in DL

DA is a critical technique for increasing the diversity of training data without the need for additional data collection. By creating modified versions of existing images through transformations such as rotation, translation, flipping, scaling, and colour adjustment (among other techniques), DA enhances the robustness and generalisation capabilities of ML models. It introduces realistic variations that mirror potential occurrences during data acquisition, helping to prevent model over-fitting and ensuring adaptability to a diverse range of imaging conditions.

In medical imaging, DA is particularly vital due to the scarcity and high cost of obtaining labelled datasets. DL models require substantial amounts of data for effective training, which is often difficult to acquire in the medical field [2, 28, 45]. By generating augmented images, the effective size of the available dataset is increased, facilitating the training of more accurate and generalisable models [14,28,45,76]. This is especially important given the variability in medical images due to differences in imaging equipment, patient positioning, and other factors. Augmented data helps ensure that the model remains invariant to such discrepancies, a necessity for clinical applications [14,30,45,76]. The significance of DA in training DL models cannot be overstated, particularly due to its impact on model generalisation and robustness [53]:

- *Increased training data*: DA expands the amount of training data by introducing modified versions of the data through transformations. This variety aids in better generalisation from the training data to new, unseen data, preventing over-fitting, where the model learns the training data too well, including its noise and outliers, thus reducing its performance on unseen data [69].

– *Improved robustness*: By exposing the model to a wider range of variations, DA enhances its robustness [60, 69]. This ensures the model is less likely to be confused by minor changes or distortions in the input data, common in real-world scenarios.
– *Class imbalance mitigation*: Many datasets suffer from under-representation of certain classes [30, 45]. DA can help balance the dataset by creating more examples of minority classes, leading to improved performance in classifying these classes.
– **Transfer learning**: When applying a model to a new but related problem, DA can be used to tailor the model to the new task by incorporating domain-specific variations into the training process [53].

Several studies have underscored the utility of DA in leveraging the limited availability and variability of medical datasets while addressing challenges such as class imbalance and over-fitting. For instance, [82] explored various DA techniques on different datasets, including medical images, and demonstrated significant improvements in model performance. They found that rotation, translation, and flipping were particularly effective, significantly enhancing classification accuracy and reducing over-fitting.

A comprehensive survey by [88] reviewed numerous DA techniques and their applications in different fields, including medical imaging. The survey discusses the effectiveness of geometric transformations and colour space augmentations, highlighting that DA enhances model generalisation, with techniques like rotation and flipping being particularly effective for medical images. Similarly, [27] used extensive DA, including rotations, translations, and colour adjustments, to train a CNN for skin cancer classification, achieving dermatologist-level accuracy.

[14] systematically examined various DA methods used in medical imaging and their impact on model performance. Their review confirmed that DA consistently improves model accuracy and robustness, with geometric transformations being the most effective. [83] explored the effectiveness of combining transfer learning with DA techniques in medical image classification tasks, finding that this combination significantly improves model performance, especially in limited data scenarios. Similarly, [95] demonstrated that integrating transfer learning with a DA strategy enhances model performance on classification tasks across various datasets.

Research consistently shows that implementing diverse DA techniques helps synthetically expand the dataset, providing the model with a plethora of scenarios to learn from. This reduces the risk of over-fitting by preventing the model from memorising specific patterns in the training data and instead improves its ability to generalise to new, unseen data. This is essential for creating robust models that perform well across diverse datasets and in real-world applications.

3 Methodology

3.1 Reinforcement learning in DA

HIs are high-dimensional and complex, making manual identification of relevant features time-consuming and prone to error. Reinforcement learning (RL), a field of ML focused on how agents should take actions in an environment to maximise cumulative rewards, offers a promising approach to address this challenge. An RL agent can be trained to autonomously select and extract features from augmented data, thereby enhancing the model's ability to generalise across different variations of the images. By receiving rewards for correctly identifying cancerous regions, the agent refines its feature selection strategy, potentially uncovering novel biomarkers. Standard DA techniques apply fixed transformations, which may not always be optimal for all images. RL can be employed to dynamically adjust augmentation policies. An RL agent can learn which augmentations improve model performance and apply them adaptively, maximising the effectiveness of DA. This approach ensures that the most beneficial transformations are applied, improving the robustness of the model.

To effectively measure the performance of the RL agent, the area under the receiver operating characteristic curve (AUC) is chosen as the key performance metric in this paper. AUC is a widely recognised metric for evaluating the performance of binary classification models, especially in medical imaging. It provides a single scalar value that summarises the trade-off between sensitivity (true positive (TP) rate) and specificity (false positive (FP) rate) across different threshold settings. A higher AUC indicates better overall performance, as it reflects the model's ability to distinguish between classes regardless of the decision threshold. Therefore, using AUC as the reward metric ensures that the RL agent is optimised for both sensitivity and specificity, which are crucial for the accurate identification of cancerous regions in HIs. In this RL framework, the components are defined as follows:

– *State* (S): The current state of the environment, represented by the set of current augmentation policies applied to the training images.
– *Action* (A): The action taken by the agent, which involves selecting a new augmentation policy or adjusting the parameters of the current policy.
– *Reward* (R): The feedback received from the environment after taking an action, represented by the performance metric (e.g., AUC) of the trained model.
– *Policy* (π): The strategy used by the agent to decide the next action based on the current state. The policy can be deterministic or stochastic.

The RL agent's goal is to learn an optimal policy π^* that maximises the expected cumulative reward over time. The RL problem can be formalised as a Markov Decision Process (MDP) with components such as states, actions, state transition probabilities, reward functions, and discount factors:

– S is a finite set of states.
– A is a finite set of actions.
– P are the state transition probabilities, $P(s'|s,a)$, representing the probability of transitioning to state s' from state s after taking action a.
– R is the reward function, $R(s,a)$, representing the immediate reward received after taking action a in state s.
– γ is the discount factor, $0 \leq \gamma \leq 1$, which determines the importance of future rewards.

The objective is to find the optimal policy π^* that maximises the expected cumulative reward G_t:

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \tag{1}$$

The action-value function $Q^\pi(s,a)$ represents the expected cumulative reward starting from state s, taking action a, and following policy π:

$$Q^\pi(s,a) = \mathbb{E}_\pi \left[G_t \mid S_t = s, A_t = a \right] \tag{2}$$

The optimal action-value function $Q^*(s,a)$ satisfies the Bellman optimality equation:

$$Q^*(s,a) = \mathbb{E}\left[R_{t+1} + \gamma \max_{a'} Q^*(S_{t+1}, a') \mid S_t = s, A_t = a \right] \tag{3}$$

The optimal policy π^* can be derived by choosing the action that maximises $Q^*(s,a)$:

$$\pi^*(s) = \arg\max_a Q^*(s,a) \tag{4}$$

Meta-learning and its application to data augmentation Meta-learning, often called "learning to learn," is a paradigm in ML where models are designed to rapidly adapt to new tasks or environments by leveraging prior experience [10]. In the context of DA, meta-learning provides a framework to optimise augmentation strategies that generalise well across different datasets and models. In traditional ML, models are trained on a fixed dataset with predefined augmentations. However, in meta-learning, the focus is on learning a meta-policy that can adaptively determine the most effective augmentations based on the current state of the model and data [84]. The key challenge lies in systematically learning from prior experience in a data-driven way, which requires collecting meta-data that describe previous learning tasks and models [39]. This approach enhances the model's ability to generalise and perform well on unseen data.

Auto-augmentation as a meta-learning approach Under the meta-learning umbrella, auto-augmentation through RL represents an advanced approach to adaptively optimise DA strategies. The auto-augmentation technique seeks to automatically identify the most effective augmentations using RL. Here, RL serves as the meta-learning strategy, where the policy includes various sub-policies representing different types of augmentations and their magnitudes.

– *Adaptation*: The agent adapts its augmentation policy based on the classifier's performance, continually refining the augmentations to improve the training process.
– *Exploration and exploitation*: The agent balances exploring new augmentation sub-policies and exploiting known effective ones to enhance learning.

– *Policy optimisation*: The RL framework optimises the policy to generalise across different datasets and classifiers, ensuring robust performance improvements.

That is, the auto-augmentation method aims to automatically identify the most effective augmentations by leveraging RL [51, 66]. In this framework, RL employs a policy that encompasses all feasible actions or options the agent (augmentation generator) can take to achieve a specific goal. For auto-augmentation, this policy consists of sub-policies, each representing a particular type of augmentation and its magnitude, such as a 90-degree rotation [108,112]. Consequently, the agent explores all sub-policies or augmentation options to minimise the loss function of a classifier, which is its primary objective [88]. By leveraging meta-learning, the auto-augmentation process can be viewed as a form of meta-learning where the agent learns to quickly adapt augmentation strategies based on the feedback received [66]. This approach enhances the robustness and generalisability of the augmentation policies, leading to improved performance of the CNN model across various datasets and tasks [69, 73].

Application to DA In this study, the RL agent adapts DA policies to maximise the performance of the CNN models. The RL agent iteratively updates its policy based on the feedback (reward) from the model's performance. This dynamic adjustment ensures that the data augmentation strategies are optimised to improve the model's generalisation ability and robustness. The process involves:

1. *Initialisation*: The RL agent starts with an initial set of augmentation policies.
2. *Training*: The CNN model is trained using the augmented data, and the performance metric (e.g., validation accuracy) is recorded.
3. *Policy update*: The RL agent receives the performance metric as a reward and updates its augmentation policy accordingly.
4. *Iteration*: Steps 2 and 3 are repeated, allowing the RL agent to iteratively improve the augmentation policies based on the model's performance.

Incorporating RL into the DA process aims to achieve more effective and efficient augmentation strategies, ultimately leading to better-performing CNN models for breast cancer classification. This approach leverages the strengths of RL to adapt and optimise DA in a way that static augmentation techniques cannot, providing a significant advantage in improving model accuracy and robustness.

Dataset For the experiments, the Breast Cancer Histopathological Database (BreakHis) dataset is utilised (for a description, see [91]). This well-established public dataset[1] has been widely used in the development and evaluation of computer-aided diagnosis (CAD) systems for breast cancer. It presents, however, 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.

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 5:

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N}\left[w_p y_i \log(p_i) + w_n(1-y_i)\log(1-p_i)\right], \tag{5}$$

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 stratified split ensures 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. This approach allows for an unbiased evaluation of the model's performance on unseen data, providing a robust assessment of its generalisability.

[1] BreakHis dataset is available here.

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 strategy Three distinct CNN architectures were chosen due to their proven effectiveness in image classification tasks: ResNet50, DenseNet201, and EfficientNetB0. Each architecture brings unique strengths that make them particularly suitable for medical image analysis, where fine-grained details are crucial for accurate diagnosis. Despite their architectural differences, the core structure and training process for these models share common elements. The models include several convolutional layers, batch normalisation, max-pooling, dropout, and a dense layer with a sigmoid activation function for binary classification.

- *Backbone Networks*:
  - ResNet50 utilises residual connections to improve gradient flow and allows the training of deeper networks without the issues of vanishing or exploding gradients.
  - DenseNet201 incorporates dense connections that improve feature propagation and reuse, leading to efficient learning.
  - EfficientNetB0 employs compound scaling to balance network depth, width, and resolution, enhancing performance with fewer parameters.
- *Additional Convolutional Layers*: With ReLU activation, these layers refine the features extracted by the backbone networks for the specific task of binary cancer classification.
- *Batch Normalisation*: Applied after each convolution to stabilise learning and reduce the number of epochs needed for convergence.
- *MaxPooling and Dropout*: The former reduces spatial dimensions to decrease computational load, while the latter mitigates over-fitting by randomly omitting units during training.

The models employ a binary classification output using a *sigmoid activation function*, which outputs a probability score indicating the likelihood of the presence of cancer. The training process involves compiling the model with an Adam optimiser, a custom weighted binary cross-entropy loss function, and evaluation metrics such as accuracy, sensitivity, specificity, and AUC. DA is applied to the training images to improve generalisation, using techniques such as rotation, width and height shifts, shear, zoom, and horizontal flip. Early stopping and learning rate reduction callbacks are employed to enhance training efficiency and prevent over-fitting. This structured approach ensures that each architecture is optimally trained for the task of binary cancer classification, leveraging their unique strengths to achieve high performance and robustness.

DA To enhance the diversity and robustness of the dataset, various augmentation functions were defined. These functions include transformations such as rotation, translation, shear, scale, and flips:

- *Rotation (20-degree limit)*: The image is rotated by a specified angle to handle the random orientation of tissues on pathology slides.
- *Translation*: The image is shifted in the horizontal or vertical direction.
- *Shear (20% range)*: A shear transformation is applied to the image to simulate changes in viewing angle by displacing parts of the image, mimicking variations in microscope slide placement.
- *Scale (20% range)*: The image is resized to help the model learn from varying scales and proportions of image features, which is crucial in histopathology where pathological feature sizes differ.
- *Flips (horizontal and vertical)*: The image is flipped horizontally or vertically to increase data variability and promote a model invariant to tissue and cell orientation.
- *'Nearest fill' mode*: Missing pixels are filled with nearby values to help maintain histopathological image integrity.

It is important to note that, during the experimental stage, various more and less aggressive DA approaches were tested. These included smaller and larger rotations, lower and higher zoom levels, and additional types of transformations. However, the DAs reduced the model's accuracy, possibly because they introduced features into the images that are not representative of true histopathological scenarios [107], [52]. This could confuse the model and lead it to learn incorrect patterns. For instance, very high zoom levels could distort cellular structures beyond realistic proportions, and excessive rotation might present the tissues in orientations that are never encountered in actual diagnostic settings.

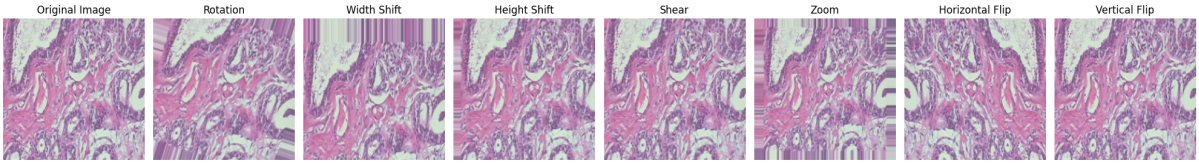

Fig. 1: Examples of augmented images.

Figure 1 shows examples of augmented images from the training set, highlighting the diversity introduced through data augmentation while avoiding the creation of misleading artifacts.

An auto-augment function was implemented to apply a series of these augmentations to each image, guided by the policies defined by the *Augmentation Agent*, which manages and updates augmentation policies. It is initialised with a set of potential augmentations and generates policies specifying the type and magnitude of each augmentation. The agent can update these policies based on the model's performance, following a RL approach where the agent learns the best policies over time. During training, DA was applied to the training images using the augmentation policies generated by the Augmentation Agent.

Five-fold cross-validation approach To ensure the robustness and generalisability of the findings, a k-fold cross-validation approach is 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.

The model and its training history were saved after each fold, allowing for a detailed analysis of the performance across different cross-validation runs. By following this methodology, the study ensured that the model was rigorously evaluated and optimised, providing reliable results for the task of breast cancer classification using HIs.

4 Results and discussion

Figure 2 presents the training and validation curves for the three architectures tested in this study: ResNet50, DenseNet201, and EfficientNetB0. These curves provide insights into the models' learning processes and their performance over the training epochs with RL guided DA.

For ResNet50, the training and validation accuracy curves show a steady increase, with the model achieving high accuracy early in the training process. The loss curves indicate that the model's training loss decreases smoothly, while the validation loss stabilises, suggesting good generalisation without significant over-fitting. The RL guided DA likely contributed to this steady improvement by dynamically adapting the augmentation strategies based on the model's performance.

DenseNet201 exhibits a similar pattern, with both training and validation accuracy improving consistently throughout the epochs. The loss curves for DenseNet201 also demonstrate a gradual decrease in training loss and stabilisation of validation loss. This model shows the highest overall accuracy and lowest loss, reflecting its superior performance compared to the other architectures. The effectiveness of RL in optimising data augmentation policies is evident in DenseNet201's performance, highlighting the model's ability to generalise well across different variations in the data.

EfficientNetB0's training curves indicate rapid initial learning, with accuracy curves showing quick improvement early in the epochs. The loss curves for EfficientNetB0, like the other models, demonstrate a steady decline in training loss and stable validation loss. However, EfficientNetB0's validation accuracy plateaus slightly lower than DenseNet201, indicating a marginally lower generalisation capability. The RL guided DA still proves beneficial in maintaining performance and reducing over-fitting tendencies.

In summary, the training curves illustrate that all three models effectively learn from the data, with DenseNet201 showing the best overall performance, followed by ResNet50 and EfficientNetB0. The

stability of the validation curves across all models suggests that the RL guided DA and regularisation techniques applied were successful in preventing over-fitting, leading to robust and generalisable models.

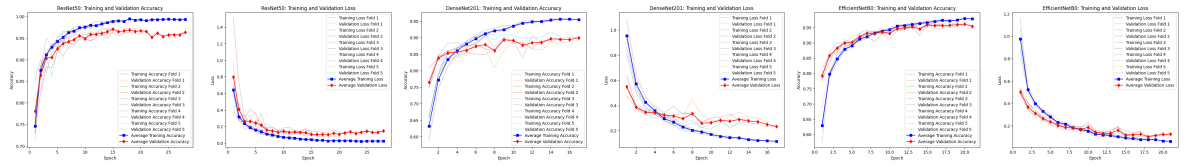

Fig. 2: Training and validation curves for the three architectures (ResNet50, DenseNet201, and EfficientNetB0) tested in this study.

The performance of ResNet50, DenseNet201, and EfficientNetB0 was evaluated based on several metrics: accuracy, precision, recall, specificity, F1-score, and ROC-AUC. Table 1 summarises the average performance metrics for each model. With an average accuracy of 0.9812, DenseNet201 demonstrated superior performance in correctly classifying images. It also achieved the highest precision (0.9872) and F1-score (0.9865), indicating a high proportion of true positive identifications and a strong balance between precision and recall. DenseNet201's recall of 0.9858 and specificity of 0.9710 suggest that it effectively identified cancerous regions while maintaining a low FP rate. The ROC-AUC score (in Figure 3) of 0.9964 further underscores DenseNet201's excellent performance, reflecting its capability to distinguish between cancerous and non-cancerous images accurately.

Table 1: Performance metrics for different architectures.

Architecture	Accuracy	Precision	Recall	Specificity	F1-score	ROC-AUC
ResNet50	0.9603	0.9740	0.9687	0.9413	0.9713	0.9899
DenseNet201	0.9812	0.9872	0.9858	0.9710	0.9865	0.9964
EfficientNetB0	0.9526	0.9837	0.9473	0.9645	0.9652	0.9924

ResNet50 also performed well, with an average accuracy of 0.9603 and a high ROC-AUC score of 0.9899 (in Figure 3) . This model's performance metrics indicate that it is a robust option for HIA, though it slightly lags behind DenseNet201 in terms of overall accuracy and F1-score. EfficientNetB0, while achieving a slightly lower average accuracy (0.9526) compared to ResNet50 and DenseNet201, demonstrated strong precision (0.9837) and specificity (0.9645). Its ROC-AUC score (in Figure 3) of 0.9924 indicates a high level of performance in differentiating between cancerous and non-cancerous images. However, its recall (0.9473) and F1-score (0.9652) suggest it may have a slightly higher tendency to miss some true positives compared to the other models.

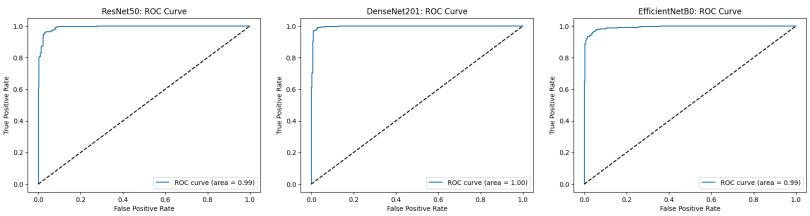

Fig. 3: Receiver Operating Characteristic (ROC) curves.

The Precision-Recall curves illustrate the trade-off between precision and recall across different thresholds, offering insights into each model's performance in handling the positive class. Figure 4 presents these curves, highlighting the models' ability to distinguish between cancerous and non-cancerous classes, especially in imbalanced class scenarios.

For ResNet50, the Precision-Recall curve shows high precision across a wide range of recall values, indicating that the model is effective in maintaining a low false positive rate while correctly identifying a

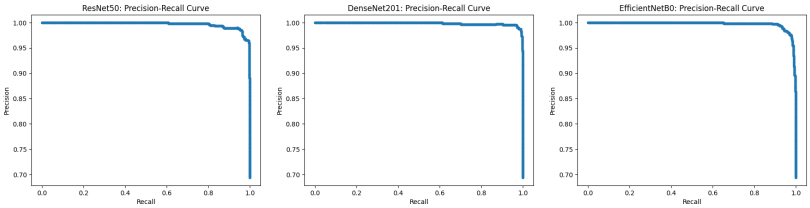

Fig. 4: Precision-Recall curves.

large proportion of true positives. This performance is reflected in the model's high precision and recall metrics.

DenseNet201 exhibits an even more robust Precision-Recall curve, maintaining both high precision and high recall across almost the entire range of threshold settings. This curve highlights DenseNet201's superior ability to correctly identify positive cases while keeping false positives to a minimum, which aligns with its high overall accuracy and F1-score.

EfficientNetB0's Precision-Recall curve also demonstrates strong performance, though it shows a slight drop in precision at the highest recall values compared to DenseNet201. Despite this, EfficientNetB0 still maintains high precision and recall, indicating its effectiveness in correctly identifying positive cases while minimising false positives.

That is, the Precision-Recall curves for all three models indicate strong performance in handling the positive class, with DenseNet201 showing the best overall performance, followed by ResNet50 and EfficientNetB0. These results reinforce the findings from the other performance metrics, confirming that DenseNet201 is the most effective model for this breast cancer classification task.

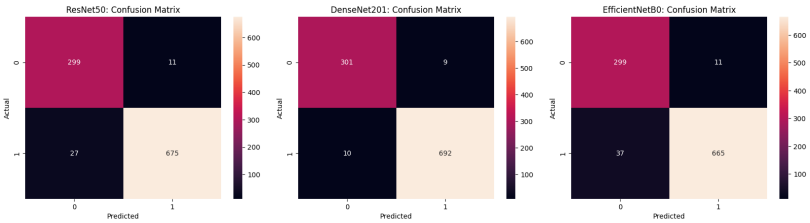

Fig. 5: Confusion Matrices for ResNet50, DenseNet201, and EfficientNetB0 architectures.

The *Confusion Matrices* presented in Figure 5 offer a granular look at each model's predictive performance, detailing the number of TPs, TNs, FPs, and FNs.

For ResNet50, the confusion matrix shows that the model correctly classified 675 instances as TPs and 239 as TNs, with only 27 FNs and 11 FPs. This indicates a strong performance in identifying both cancerous and non-cancerous cases, reflecting the high precision and recall observed in the earlier metrics.

DenseNet201 demonstrates the best performance among the three models, with 692 TPs and 302 TNs, while only mis-classifying 13 instances as FNs and 9 as FPs. The high number of correctly classified cases and the low number of mis-classifications highlight DenseNet201's superior ability to accurately detect cancerous regions, aligning with its highest overall accuracy, precision, and recall metrics.

EfficientNetB0's confusion matrix shows 665 TPs and 239 TNs, with 37 FNs and 11 FPs. While the performance is still robust, the slightly higher number of FNs compared to DenseNet201 suggests a marginally lower recall. However, the model maintains a high precision, indicating it is effective at minimising FPs.

In summary, the confusion matrices corroborate the performance metrics, with DenseNet201 exhibiting the most balanced and accurate classification, followed by ResNet50 and EfficientNetB0. These results underscore the importance of choosing the right architecture based on the specific requirements of the task, with DenseNet201 showing clear advantages in this binary breast cancer classification study.

5 Conclusion

This study evaluated the performance of three most popular CNN architectures - ResNet50, DenseNet201, and EfficientNetB0 - in the context of binary breast cancer classification using HIs. The architectures were enhanced with RL guided DA, which dynamically adjusted augmentation strategies based on model performance. The results were assessed using a range of metrics including accuracy, precision, recall, specificity, F1-score, and ROC-AUC.

The training and validation curves indicated effective learning across all models, with DenseNet201 demonstrating the most consistent and highest performance. DenseNet201 achieved the highest average accuracy (0.9812), precision (0.9872), recall (0.9858), F1-score (0.9865), and ROC-AUC (0.9964), reflecting its superior capability to generalise across different data variations. ResNet50 and EfficientNetB0 also performed robustly, with ResNet50 showing an average accuracy of 0.9603 and ROC-AUC of 0.9899, while EfficientNetB0 achieved an average accuracy of 0.9526 and ROC-AUC of 0.9924.

The Precision-Recall curves further confirmed DenseNet201's outstanding performance in identifying positive cases while minimising false positives, with ResNet50 and EfficientNetB0 also showing strong but slightly lower performance. The confusion matrices provided a granular view of the models' predictive capabilities, with DenseNet201 again leading in the number of correctly classified instances and the lowest number of mis-classifications.

Overall, DenseNet201 emerged as the most effective architecture for breast cancer classification in this study, followed by ResNet50 and EfficientNetB0. The use of RL guided DA proved beneficial in enhancing model performance and preventing over-fitting. These findings highlight the potential of advanced CNN architectures combined with RL strategies to significantly improve the accuracy and robustness of cancer detection models, offering valuable insights for future research and clinical applications.

Despite the promising results and a novel approach, this study has several limitations which suggest directions for future research:

- The results presented here are based on a specific dataset, and the performance of the models might vary with different datasets. Future studies should consider using larger and more diverse datasets to evaluate the models' generalisation capabilities comprehensively.
- While cross-validation was used to ensure robustness, it might not fully capture the models' performance on entirely unseen data. Further validation on independent test sets is necessary to confirm the findings and provide a clearer picture of the models' real-world applicability.
- Training DL models with RL guided DA requires significant computational resources, which might limit the accessibility and scalability of the proposed methods. Future research could explore more efficient algorithms or hardware solutions to mitigate this constraint.
- The integration of RL for DA introduces additional complexity in the training process. Developing and fine-tuning RL policies can be challenging and time-consuming. Simplifying the RL frameworks or developing more user-friendly interfaces could encourage wider adoption in practice.
- While the models demonstrated high performance, their interpretability remains a challenge. Future studies should focus on improving the interpretability of these models to ensure trust and reliability in clinical applications. Techniques such as explainable AI could be explored to make the decision-making process of the models more transparent.

6 Disclosure of Interests

The author has no competing interests to declare.

References

1. Abarghouei, A.A., Ghanizadeh, A., Sinaie, S., Shamsuddin, S.M.: A survey of pattern recognition applications in cancer diagnosis. In: International Conference of Soft Computing and Pattern Recognition. pp. 448–453. IEEE (2009)
2. 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)
3. AbdelMaksoud, E., Barakat, S., Elmogy, M.: A computer-aided diagnosis system for detecting various diabetic retinopathy grades based on a hybrid deep learning technique. Medical & Biological Engineering & Computing **60**(7), 2015–2038 (2022)

4. Ahmed, H.A., Mohammed, E.A.: Detection and classification of the osteoarthritis in knee joint using transfer learning with convolutional neural networks (cnns). International Journal of Sciences **63**(11), 40 (2022)
5. Alkabban, F.M., Ferguson, T.: Breast cancer. StatPearls (9) (2022)
6. 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)
7. 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)
8. Ashraf, F.B., Alam, S.M., Sakib, S.M.: Enhancing breast cancer classification via histopathological image analysis: leveraging self-supervised contrastive learning and transfer learning. Heliyon **10**(2) (2024)
9. Azad, R., Kazerouni, A., Heidari, M., Aghdam, E.K., Molaei, A., Jia, Y., Jose, A., Roy, R., Merhof, D.: Advances in medical image analysis with vision transformers: a comprehensive review. Medical Image Analysis p. 103000 (2023)
10. Beck, J., Vuorio, R., Liu, E.Z., Xiong, Z., Zintgraf, L., Finn, C., Whiteson, S.: A survey of meta-reinforcement learning (2023)
11. 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. A Cancer Journal for Clinicians (2024)
12. Brock, A., De, S., Smith, S.L.: Characterising signal propagation to close the performance gap in unnormalised resnets (2021)
13. Chang, M.C., Mrkonjic, M.: Review of the current state of digital image analysis in breast pathology. The Breast Journal **26**(6), 1208–1212 (2020)
14. Chlap, P., Min, H., Vandenberg, 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)
15. Ciresan, D., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. Advances in Neural Information Processing Systems **25** (2012)
16. Cireșan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Mitosis detection in breast cancer histology images with deep neural networks. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI): 16th International Conference. pp. 411–418. Springer (2013)
17. Cruz-Roa, A., Basavanhally, A., Gonzalez, F., Gilmore, H., Feldman, M., Ganesan, S., Shih, N., Tomaszewski, J., Madabhushi, A.: Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In: SPIE Medical Imaging. vol. 9041, p. 904103. International Society for Optics and Photonics (2014)
18. Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V.: Autoaugment: learning augmentation strategies from data. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 113–123 (2019)
19. Cui, M., Zhang, D.Y.: Artificial intelligence and computational pathology. NPJ Precision Oncology **5**(1), 1–15 (2021)
20. 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)
21. Das, A., Nair, M., Peter, S.: Computer-aided histopathological image analysis techniques for automated nuclear atypia scoring of breast cancer: a review. Journal of Digital Imaging **33**(5), 1091–1121 (2020)
22. Das, L., 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)
23. Dollar, P., Singh, M., Girshick, R.: Fast and accurate model scaling. In: International Conference on Computer Vision and Pattern Recognition (IEEE/CVF). pp. 924–932 (2021)
24. Ebski, S.J., Arpit, D., Ballas, N., Verma, V., Che, T., Bengio, Y.: Residual connections encourage iterative inference. In: International Conference on Learning Representations (2018)
25. Elazab, N., Gab-Allah, W.A., Elmogy, M.: A multi-class brain tumour grading system based on histopathological images using a hybrid yolo and resnet networks. Scientific Reports **14**(1), 4584 (2024)
26. 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)
27. 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)
28. 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)
29. Farid, A.A., Khater, H., Selim, G.: A cnn classification model for diagnosis covid19 (2020)
30. 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)
31. Ghaffar Nia, N., Kaplanoglu, E., Nasab, A.: Evaluation of artificial intelligence techniques in disease diagnosis and prediction. Discover Artificial Intelligence **3**(1), 5 (2023)
32. Go, H.: Digital pathology and artificial intelligence applications in pathology (2022)

33. 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)
34. Hanahan, D.: Hallmarks of cancer: new dimensions. Cancer Discovery **12**(1), 31–46 (2022)
35. Hara, K., Saito, D., Shouno, H.: Analysis of function of rectified linear unit used in deep learning. In: International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2015)
36. He, F., Liu, T., Tao, D.: Why resnet works? residuals generalise. IEEE Transactions on Neural Networks and Learning Systems **31**(12), 5349–5362 (2020)
37. Herdiantoputri, R.R., Komura, D., Fujisaka, K., Ikeda, T., Ishikawa, S.: Deep texture representation analysis for histopathological images. STAR Protocols **4**(2), 102161 (2023)
38. Ho, D., Liang, E., Chen, X., Stoica, I., Abbeel, P.: Population based augmentation: efficient learning of augmentation policy schedules. In: Chaudhuri, K., Salakhutdinov, R. (eds.) 36th International Conference on Machine Learning. vol. 97, pp. 2731–2741. PMLR (6 2019)
39. Hospedales, T., Antoniou, A., Micaelli, P., Storkey, A.: Meta-learning in neural networks: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(9), 5149–5169 (2021)
40. 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)
41. Huang, G., Liu, Z., Pleiss, G., Van Der Maaten, L., Weinberger, K.Q.: Convolutional networks with dense connectivity. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(12), 8704–8716 (2019)
42. Huang, X., Yao, C., Xu, F., Chen, L., Wang, H., Chen, X., Ye, J., Wang, Y.: Mac-resnet: knowledge distillation based lightweight multiscale-attention-crop-resnet for eyelid tumours detection and classification. Journal of Personalised Medicine **13**(1), 89 (2022)
43. Huss, R., Coupland, S.E.: Software-assisted decision support in digital histopathology. Journal of Pathology **250**(5), 685–692 (2020)
44. Ing, N., Ma, Z., Li, J., Salemi, H., Arnold, C., Knudsen, B.S., Gertych, A.: Semantic segmentation for prostate cancer grading by convolutional neural networks. In: Medical Imaging 2018: Digital Pathology. vol. 10581, pp. 343–355. SPIE (2018)
45. 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)
46. 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)
47. 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)
48. Kajala, A., Jaiswal, S.: Classification of breast cancer histopathology images using efficientnet architectures. In: Advances in Information Communication Technology and Computing Conference (AICTC), pp. 639–653. Springer (2022)
49. Kalinathan, L., Sivasankaran, D., Jeyasingh, J.R., Sudharsan, A.S., Marimuthu, H.: Classification of hepatocellular carcinoma using machine learning (2021)
50. Kandel, I., Castelli, M.: A novel architecture to classify histopathology images using convolutional neural networks. Applied Sciences **10**(8), 2929 (2020)
51. 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)
52. Khamankar, V., Bera, S., Bhattacharya, S., Sen, D., Biswas, P.K.: Histopathological image analysis with style-augmented feature domain mixing for improved generalisation. In: Medical Image Computing and Computer Assisted Intervention (MICCAI) Workshop. Lecture Notes in Computer Science, vol. 14393. Springer (2023)
53. Khamis, T., Mokayed, H.: Optimising the ai development process by providing the best support environment (2023)
54. Khan, A., Sohail, A., Zahoora, U., Qureshi, A.S.: A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review **53**, 5455–5516 (2020)
55. Kiani, A., Uyumazturk, B., Rajpurkar, P., Wang, A., Gao, R., Jones, E., Yu, Y., Langlotz, C.P., Ball, R.L., Montine, T.J., et al.: Impact of a deep learning assistant on the histopathologic classification of liver cancer. NPJ Digital Medicine **3**(1), 23 (2020)
56. Komura, D., Ishikawa, S.: Machine learning methods for histopathological image analysis. Computational and Structural Biotechnology Journal **16**, 34–42 (2018)
57. Kornblith, S., Shlens, J., Le, Q.V.: Do better imagenet models transfer better? In: International Conference on Computer Vision and Pattern Recognition (IEEE/CVF). pp. 2661–2671 (2019)
58. Krishna, S., Krishnamoorthy, S., Bhavsar, A., et al.: Stain normalised breast histopathology image recognition using convolutional neural networks for cancer detection (2022)
59. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Communications of the ACM **60**(6), 84–90 (2017)
60. Kumar, T., Mileo, A., Brennan, R., Bendechache, M.: Image data augmentation approaches: a comprehensive survey and future directions (2023)

61. Langlotz, C.P.: Will artificial intelligence replace radiologists? Radiology: Artificial Intelligence **1**(3), e190058 (2019)
62. 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)
63. LeCun, Y., Bengio, Y., et al.: Convolutional networks for images, speech, and time series. The Handbook of Brain Theory and Neural Networks **3361**(10) (1995)
64. Lee, H., Song, J.: Introduction to convolutional neural network using keras; an understanding from a statistician. Communications for Statistical Applications and Methods **26**(6), 591–610 (2019)
65. Li, G., Zhang, M., Li, J., Lv, F., Tong, G.: Efficient densely connected convolutional neural networks. Pattern Recognition **109**, 107610 (2021)
66. Lin, S., Yu, T., Feng, R., Li, X., Yu, X., Xiao, L., Chen, Z.: Local patch autoaugment with multi-agent collaboration. IEEE Transactions on Multimedia (2023)
67. 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)
68. Liu, Y.H.: Feature extraction and image recognition with convolutional neural networks. Journal of Physics: Conference Series **1087**(6), 062032 (2018)
69. 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)
70. Mandair, D., Reis-Filho, J.S., Ashworth, A.: Biological insights and novel biomarker discovery through deep learning approaches in breast cancer histopathology. NPJ Breast Cancer **9**(1), 1–10 (2023)
71. 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)
72. 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)
73. Mikolajczyk, A., Grochowski, M.: Data augmentation for improving deep learning in image classification problem. In: International Interdisciplinary PhD workshop (IIPhDW). pp. 117–122. IEEE (2018)
74. 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)
75. Moxley-Wyles, B., Colling, R., Verrill, C.: Artificial intelligence in pathology: an overview. Diagnostic Histopathology **26**(11), 498–502 (2020)
76. Mumuni, A., Mumuni, F.: Data augmentation: a comprehensive survey of modern approaches. Array **16**, 100258 (2022)
77. 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)
78. Munir, K., Elahi, H., Ayub, A., Frezza, F., Rizzi, A.: Cancer diagnosis using deep learning: a bibliographic review. Cancers **11**(9), 1235 (2019)
79. Ng, A.Y., Oberije, C.J., Ambrozay, E., Szabo, E., Serfozo, O., Karpati, E., Fox, G., Glocker, B., Morris, E.A., Forrai, G., et al.: Prospective implementation of ai-assisted screen reading to improve early detection of breast cancer. Nature Medicine **29**(12), 3044–3049 (2023)
80. O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G.V., Krpalkova, L., Riordan, D., Walsh, J.: Deep learning vs. traditional computer vision. In: Computer Vision Conference (CVC): Advances in Computer Vision. vol. 1, pp. 128–144. Springer (2020)
81. Pallua, J.D., Brunner, A., Zelger, B., Schirmer, M., Haybaeck, J.: The future of pathology is digital. Virchows Archiv **477**(4), 509–516 (2020)
82. Perez, L., Wang, J.: The effectiveness of data augmentation in image classification using deep learning (2017)
83. Raghu, M., Zhang, C., Kleinberg, J., Bengio, S.: Transfusion: understanding transfer learning for medical imaging. Advances in Neural Information Processing Systems **32** (2019)
84. Rajendran, J., Irpan, A., Jang, E.: Meta-learning requires meta-augmentation. Advances in Neural Information Processing Systems **33**, 5705–5715 (2020)
85. Ridzuan, M., Bawazir, A., Gollini Navarrete, I., Almakky, I., Yaqub, M.: Self-supervision and multi-task learning: challenges in fine-grained covid-19 multi-class classification from chest x-rays. In: Annual Conference on Medical Image Understanding and Analysis. pp. 234–250. Springer (2022)
86. Sarah, M., Mathieu, L., Philippe, Z., Guilcher, A.L., Borderie, L., Cochener, B., Quellec, G.: Generalising deep learning models for medical image classification (2024)
87. Shoaib, M., Shah, B., Ei-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., Gechev, T., Hussain, T., Ali, F.: An advanced deep learning models-based plant disease detection: a review of recent research. Frontiers in Plant Science **14**, 1158933 (2023)
88. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. Journal of Big Data **6**(1), 1–48 (2019)
89. Sikandar, A.: Histopathology: an old yet important technique in modern science. In: Histopathology, pp. 3–24. IntechOpen (2018)

90. 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)
91. 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)
92. Srinidhi, C.L., Ciga, O., Martel, A.L.: Deep neural network models for computational histopathology: a survey. Medical Image Analysis **67**, 101813 (2021)
93. Srinidhi, C.L., Ciga, O., Martel, A.L.: Deep neural network models for computational histopathology: a survey (2019)
94. Steiner, D.F., Macdonald, R.D., Liu, Y., Truszkowski, P., Hipp, J., Gammage, C.: 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)
95. Su, J., Yu, X., Wang, X., Wang, Z., Chao, G.: Enhanced transfer learning with data augmentation. Engineering Applications of Artificial Intelligence **129**, 107602 (2024)
96. Tan, M., Le, Q.: Efficientnet: rethinking model scalling for convolutional neural networks. In: International Conference on Machine Learning. pp. 6105–6114. PMLR (2019)
97. Tan, M., Le, Q.: Efficientnetv2: smaller models and faster training. In: International Conference on Machine Learning. pp. 10096–10106. PMLR (2021)
98. Tan, M., Pang, R., Le, Q.V.: Efficientdet: scalable and efficient object detection. In: International Conference on Computer Vision and Pattern Recognition (IEEE/CVF). pp. 10781–10790 (2020)
99. Tian, K., Wei, Q., Li, X.: Co-teaching for unsupervised domain adaptation and expansion. arXiv preprint arXiv:2204.01210 (2022)
100. 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)
101. 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)
102. 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)
103. Voon, W., Hum, Y.C., Tee, Y.K., Yap, W.S., Salim, M.I.M., Tan, T.S., Mokayed, H., Lai, K.W.: Performance analysis of seven convolutional neural networks (cnns) with transfer learning for invasive ductal carcinoma (idc) grading in breast histopathological images. Scientific Reports **12**(1), 19200 (2022)
104. Wang, J., Liu, Q., Xie, H., Yang, Z., Zhou, H.: Boosted efficientnet: detection of lymph node metastases in breast cancer using convolutional neural networks. Cancers (13), 661 (2021)
105. Wu, J.: Introduction to convolutional neural networks. National Key Lab for Novel Software Technology **5**(23), 495 (2017)
106. Xia, K., Wang, J.: Recent advances of transformers in medical image analysis: a comprehensive review. MedComm–Future Medicine **2**(1), e38 (2023)
107. Xu, J., Hou, J., Zhang, Y., Feng, R., Ruan, C., Zhang, T., Fan, W.: Data-efficient histopathology image analysis with deformation representation learning. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 857–864 (2020)
108. Yang, Z., Sinnott, R.O., Bailey, J., Ke, Q.: A survey of automated data augmentation algorithms for deep learning-based image classification tasks. Knowledge and Information Systems **65**(7), 2805–2861 (2023)
109. Yildiz, S., Memiş, A., Varl, 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)
110. 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)
111. Zaeemzadeh, A., Rahnavard, N., Shah, M.: Norm-preservation: why residual networks can become extremely deep? IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(11), 3980–3990 (2020)
112. Zhang, J., Li, D., Wang, L., Zhang, L.: Auto machine learning for medical image analysis by unifying the search on data augmentation and neural architecture (2022)
113. Zhao, C., Xing, F., Yeo, Y.H., Jin, M., Le, R., Le, M., Jin, M., Henry, L., Cheung, R., Nguyen, M.H.: Only one-third of hepatocellular carcinoma cases are diagnosed via screening or surveillance: a systematic review and meta-analysis. European Journal of Gastroenterology & Hepatology **32**(3), 406–419 (2020)
114. Zhou, S.K., Greenspan, H., Davatzikos, C., Duncan, J.S., Van Ginneken, B., Madabhushi, A., Prince, J.L., Rueckert, D., Summers, R.M.: 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)
115. Zhu, Y., Newsam, S.: Densenet for dense flow. In: IEEE International Conference on Image Processing (ICIP). pp. 790–794. IEEE (2017)
116. 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)