Annals of Data Science
https://doi.org/10.1007/s40745-026-00708-1

ORIGINAL ARTICLE

CAE-StackedNet: A Hybrid Convolutional Autoencoder and Stacked Ensemble Framework for Histopathological Breast Cancer Image Classification

Roseline Oluwaseun Ogundokun[1,2,5] · **Pius Adewale Owolawi**[1] · **Etienne van Wyk**[1] · **Chunling Du**[1] · **Cheng-Chi Lee**[3,4]

Received: 8 July 2025 / Revised: 6 May 2026 / Accepted: 14 June 2026

Abstract

The accurate classification of breast cancer histopathology images is critical for early diagnosis, risk stratification, and treatment planning. However, traditional supervised learning approaches are hindered by the limited availability of annotated medical datasets. At the same time, single-model classifiers often struggle to generalize effectively across subtle morphological variations in tissue structures. To address these challenges, this study proposes CAE-StackNet, a hybrid framework that combines self-supervised feature learning via a Convolutional Autoencoder (CAE) with a heterogeneous stacked-ensemble classifier. The CAE is first trained to reconstruct breast cancer histopathology images from the BACH dataset, enabling it to learn compact, morphology-preserving features without requiring class labels. These self-learned features are then classified using a stacked ensemble of Random Forest (RF), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP), with a meta-learner aggregating their predictions to produce the final diagnosis. Experimental results demonstrate that CAE-StackNet achieves an AUC of 0.99 and an average precision of 0.96, outperforming individual classifiers trained on the same features. MLP achieved the highest standalone accuracy (0.89), but the stacked ensemble exhibited superior discriminative power, reflecting the advantage of combining complementary classifiers in a diverse ensemble framework. For breast cancer classification in histopathology, CAE-StackNet provides a data-efficient and clinically meaningful solution by reducing the need for manual annotation and improving robustness through ensemble learning. The developed method enables more reproducible and objective histopathological grading and has excellent potential for incorporation into computer-aided diagnostic (CAD) systems. To increase clinical interpretability and gain the trust of pathologists, future efforts will involve external validation on multi-center datasets, integrating pathology-aware constraints into the CAE training process, and creating explainability tools. Recent transformer-based frameworks, such as DFViT, fuse convolutional and

Extended author information available on the last page of the article

Published online: 26 June 2026

vision-transformer blocks to capture both local and global tissue patterns and have achieved outstanding accuracy on the BACH and BreakHis datasets. By contrast, CAE-StackNet achieves comparable AUC and average precision while relying primarily on self-supervised features and requiring far fewer annotated images, underscoring its potential for scalable clinical deployment.

Keywords Self-supervised learning · Convolutional autoencoder · Stacked ensemble · Breast cancer · Histopathology · Medical image classification · Computer-aided diagnosis (CAD)

1 Introduction

Timely and accurate labelling of medical images is essential in modern healthcare, e.g., for the early diagnosis of fatal diseases such as cancer, neurological, and cardiovascular diseases [1]. Medical images often exhibit complex shapes, subtle textures, and texture variations, which, together with inter-patient variability, make hand interpretation time-consuming, subjective, and error prone [2]. Hence, deep learning (DL) and machine learning (ML) methods have gained extensive interest, as they offer reproducible, automated, and scalable diagnostic support to physicians [3].

However, deep learning models often require large amounts of labelled data, which in the medical setting is often scarce due to costly annotation procedures involving expert radiologists or pathologists [4]. To bypass this constraint, self-supervised learning (SSL) has emerged as a strong paradigm in which models pre-train on unlabelled data by solving auxiliary tasks, thereby gaining transferable representations [5]. In medical imaging, the CAEs have excelled as self-supervised feature learners, leveraging reconstruction-based learning to capture spatial, textural, and morphological patterns required for downstream classification [6]. CAEs have performed well in histopathology and radiology image analysis by preserving diagnostic features in low-dimensional latent representations [7].

Of all the applications of medical image analysis, the diagnosis of breast cancer is perhaps one of the most significant and most studied intensively due to the incidence of the disease and deaths globally. Breast cancer is the most prevalent cancer diagnosis and among the leading causes of cancer deaths in females worldwide [8]. Early and accurate diagnosis, particularly by histopathological study of breast tissue, is highly significant for effective treatment planning and improved patient outcomes. Histopathology images contain complicated cell patterns, which pathologists analyse to categorise them into benign, in situ carcinoma, and invasive carcinoma tissue classes [9]. The classification process is time-consuming and subjective, particularly when classes differ only slightly in morphology. To address these challenges, the ICIAR 2018 BACH (Breast Cancer Histology) Challenge Dataset provides a carefully curated benchmark for

developing automated diagnostic tools [9]. Successful computational methods that can learn powerful features from breast cancer histology images and perform precise multi-class classification have the potential to significantly assist pathologists in reaching concordant, unbiased diagnoses.

Despite self-guided feature learning, achieving classification accuracy remains challenging due to heterogeneity in medical images, class imbalance, and conflicting class boundaries [10]. Ensemble learning algorithms, the hallmark of robustness promotion via Classifier diversity, have proved to be first-rate solutions [11, 12]. Among ensemble methods in which predictions from multiple base classifiers are hierarchically combined by a meta-learner, stacked ensemble learning has proven highly effective for medical image classification [13, 14]. Stacking utilises the different strengths of individual classifiers, e.g., the high interpretability of decision trees, the boundary-optimising ability of SVM, and the multi-layer perceptrons' capacity for deep feature abstraction [12].

This work introduces a novel Self-Supervised Feature Learning model, CAE-StackNet, for medical image classification, including breast cancer histopathological image classification based on the BACH dataset. CAE is an autoencoder that serves as a self-supervised feature learner, learning informative embeddings from unlabeled raw breast tissue images. These acquired features are then passed to a stacked ensemble classifier, which aggregates heterogeneous base-learner predictions (e.g., RF, SVM, and MLP) into a meta-learner for the final prediction. This hybrid design combines the deep autoencoder's feature-learning strengths with the ensemble's predictive robustness.

The novelty of this framework lies in its synergistic integration of unsupervised deep representation learning and heterogeneous ensemble classification. It is specifically tailored for breast cancer histopathology, where subtle morphological differences between benign, malignant, and normal tissues make classification inherently challenging. By enabling self-supervised pretraining, the framework reduces dependence on large, labelled datasets. At the same time, the stacked ensemble classifier mitigates the weaknesses of individual base learners, resulting in a more generalisable, interpretable, and robust diagnostic system.

This hybrid framework is highly relevant to histopathological breast tissue classification, where feature learning and robust classification are paramount. Through comprehensive experiments on the BACH dataset, the proposed framework demonstrates superior classification performance compared with conventional deep learning models, single-model classifiers, and traditional ensemble methods.

In summary, this study improves the diagnosis of breast cancer through:

1. Suggesting a self-supervised convolutional autoencoder-based feature learning approach customised for histopathological breast tissue examination.
2. Developing a stacked heterogeneous ensemble classifier that exploits the diversity of complementary learners to improve predictive robustness through boosting.
3. Providing large-scale experimental results on the BACH dataset, beating baseline approaches.

2 Related Works

The recent acceleration of artificial intelligence (AI) methods in medical image analysis has greatly enhanced the precision of diagnosis [15], risk prediction, and personalised treatment planning. This section describes relevant work across four major research topics: self-supervised learning-based medical image feature extraction, convolutional autoencoder-based medical image analysis, ensemble learning approaches for medical image classification, and automatic classification of breast cancer histopathological images.

High-accuracy classification, achieved through conventional supervised learning, requires large amounts of annotated data. Medical annotation is costly and time-consuming to acquire, typically needing specialist radiologists or pathologists. This restricts the quantity of labelled data [1]. Self-supervised learning has also emerged as a potential solution to this bottleneck, leveraging models' ability to learn dense feature representations from unlabelled data via pretext tasks [5].

Self-supervised methods, such as context prediction, image inpainting, and contrastive learning, have been applied to medical images in recent research [16]. Azizi et al. [16], for instance, demonstrated that contrastive pretraining on chest X-ray datasets yields significant downstream classification benefits even with limited labelled data. In pathology, pretext tasks such as rotation prediction and jigsaw puzzle completion have proved helpful for learning representations of histopathological images [17]. These self-supervised techniques are gaining traction in label-efficient diagnostic systems, especially for underrepresented groups and rare diseases.

Among self-supervised architectures, CAEs are ideal for medical image processing because they can learn hierarchical representations of spatial information, texture patterns, and fine-grained structures [7]. Within CAEs, the encoder maps the input image to a lower-dimensional latent space. The decoder reconstructs the input image, with the regularizer preserving the latent space to maintain crucial diagnostic information [13].

CAEs have been used for unsupervised radiology anomaly detection [18], pretraining segmentation in pathology [19], and clustering tumor subtypes in oncology imaging [20]. Hussein et al. [7] demonstrated that histopathological image features learned by CAEs significantly enhanced downstream classification accuracy. Additionally, compared with supervised deep learning models, the lower computational cost of CAEs makes them highly attractive for implementation in resource-scarce clinical settings.

Ensemble learning, which combines predictions from a collection of distinct classifiers, has been particularly valuable in medical image classification, where data variation, class imbalance, and noise are common challenges [11, 12]. Ensemble methods such as bagging, boosting, and stacking utilise the complementary strengths of a collection of distinct individual classifiers to generalise and resist better.

Medical image analysis also made effective use of ensemble techniques for disease detection, including tuberculosis classification [21], diabetic retinopathy

grading [13], and lung cancer screening [10]. Stacking achieved state-of-the-art performance using classifier diversity and a meta-learner to optimally combine base classifier predictions [22]. These techniques find applications in healthcare settings, where heterogeneity in imaging protocols, equipment, and populations may limit the performance of individual-model classifiers.

Breast cancer has the highest incidence and is the largest killer of women worldwide, accounting for approximately 11.7% of all cancer cases and 6.9% of cancer deaths [8]. The histopathology of breast tissue biopsy remains the gold standard for definitive diagnosis because it allows pathologists to distinguish normal tissue from benign lesions, carcinoma in situ, and invasive carcinoma [9]. As it is a subjective process, it is prone to inter-observer discordance, especially in borderline cases [23].

Computational approaches have been widely applied to classify breast cancer histopathology images, thereby overcoming these challenges. Traditional machine learning techniques employed handcrafted features, such as Haralick textures, Gabor filters, and wavelet coefficients, and classifiers such as SVMs or RFs [23]. While these methods offered interpretability, they did not generalise to new imaging sites and staining protocols.

Recently, deep learning approaches have also shown great promise for classifying breast histopathology [9, 13]. CNNs could be trained directly on breast histology images to automatically acquire hierarchical features and achieve high classification precision. The ICIAR 2018 BACH Challenge provided a benchmark database to stimulate development in the area [9]. Several competing teams developed combinations of CNN architectures, transfer learning, and ensemble models.

Recently, Ewers et al. revealed a robust and transparent trend in research on hybrid deep learning and ensemble methods for medical image classification, with an emphasis on histopathology. Ogundokun, Owolawi, and Tu [24] are well-aligned with this work and demonstrate expertise in both convolutional autoencoders and ensemble learning for breast cancer classification. However, label scarcity and class imbalance remain major obstacles to identifying benign (non-cancerous) lesions or early-stage cancers. The use of CAEs as a self-supervised learning technique is a viable alternative to expert-based annotation, enabling the extraction of features directly from unlabelled histopathology without heavy reliance on a domain expert. Deep features extracted using CAEs can also enhance classifier robustness by reducing model-specific bias in small or imbalanced datasets [20]. Folorunso et al. [25] build on that foundation through transfer learning with EfficientNet architectures, and Ogundokun et al. [26] demonstrate the versatility of hybrid deep-learning methodologies for brain tumour detection. Ogundokun et al. [27] further extend the autoencoder/ensemble framework for lymphoma subtype classification. Taken together, all these projects present a cohesive, forward-thinking, and domain-specific contribution to AI-assisted medical imaging analysis.

In the last two years, several studies have extended beyond convolutional autoencoders to adopt contrastive learning, masked image modelling and transformer architectures in histopathology. Almubarak et al. [28] proposed a hybrid self-supervised segmentation framework that integrates masked autoencoder reconstruction, multi-scale contrastive learning, and an adaptive augmentation policy. Their approach significantly improves the Dice coefficient and mIoU on

diverse pathology datasets, using only 25% of the annotations, highlighting the importance of domain-specific augmentation and multi-scale representations [28].

Contrastive learning has also been used for classification tasks. Sani et al. [29] developed a modified supervised contrastive learning method for breast cancer histopathology that reduces false positives and improves image-level accuracy to 93.63%. Jain and Lynn [30] proposed an interpretable self-supervised contrastive learning framework for colorectal histopathology that pre-trains a ResNet50 encoder with contrastive objectives, achieves 85.86% classification accuracy and employs Grad-CAM to highlight salient tissue regions. Contrastive objectives and saliency-based interpretability can lead to the development of powerful, label-efficient models.

Recently, histopathological classification using vision transformers has been adapted. Fiaz et al. [31] presented DFViT (Deep–Fusion Vision Transformer), a method that utilizes CNN feature extraction to extract both local and global tissue patterns from images using both replica and CNN to provide high levels of accuracy, precision, and recall on the BreakHis, BACH, and UCSC databases. He et al. [32] did so for breast cancer recognition via a model that integrated deformable filters and wavelet-Fourier frequency components in the spatial-frequency domain on a vision transformer. Both methods demonstrated improved performance over classical CNNs on benchmark datasets; however, they require substantial amounts of labelled training data and computational resources to achieve comparable accuracy.

Across the surveyed literature, authors have shown that combining self-supervised deep feature learning with ensemble methods can improve medical image classification, particularly when annotated datasets are limited. In breast cancer histopathology, where the acceptable discrimination between benign, pre-cancerous, and cancerous tissue is critical, such fusion strategies are extremely promising. However, little work has explicitly combined self-supervised CAE-based feature learning with stacked ensemble classifiers for breast cancer image classification. This work addresses this by proposing CAE-StackNet, a novel hybrid approach tailored for effective and robust breast cancer histopathology classification on the BACH dataset. The present study complements these developments by using a convolutional autoencoder for self-supervised feature extraction and demonstrating that ensemble methods can achieve competitive performance without extensive annotation or transformer architectures.

3 Materials and Methods

This paper proposes a Self-Supervised Feature Learning with a Convolutional Autoencoder and a Hybrid Stacked Ensemble framework for data-efficient, robust breast cancer histopathology image classification. This pipeline provides three successive steps: (1) self-supervised deep feature learning by a CAE; (2) hybrid stacked ensemble classification with various base classifiers and a meta-classifier; and (3) assessment and performance evaluation.

3.1 The BACH Dataset

The ICIAR 2018 BACH Challenge Dataset, an open-access benchmark dataset for breast cancer histopathological image classification, is employed in the study [9, 31]. The BACH dataset includes 400 H&E-stained breast histopathology images, categorised into four classes: Normal Tissue (200 images), Benign Lesions (200 images), In Situ Carcinoma (200 images), and Invasive Carcinoma (200 images). All images are provided in 2048×1536-pixel resolution, making feature extraction difficult due to the tissue's high textural and morphological intricacy. To standardise input for the proposed framework, images were resized to 128×128 pixels during preprocessing.

3.2 Self-Supervised CAE

The CAE is a self-supervised feature extractor that learns a compact representation (bottleneck) from each input image without requiring class labels. The CAE consists of:

- Encoder: A deep convolutional stack compressing the input image into a low-dimensional latent vector.
- Decoder: A symmetrical transposed convolutional stack that reconstructs the input image from the latent vector, ensuring the preservation of diagnostic features.

The encoder consists of three convolutional layers with increasing filter sizes (32, 64, 128), followed by ReLU activation and stride-based downsampling. The decoder reverses this process using transposed convolutional layers to upsample the latent vector back to the input resolution. The final output layer uses a sigmoid activation to reconstruct pixel intensities within the range [0, 1].

The CAE is trained to minimise the Mean Squared Error (MSE) between input and reconstructed images:

$$\mathcal{L}_{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \hat{x}_i \right)^2 \tag{1}$$

This forces the encoder to capture compact, meaningful representations that preserve essential morphological and textural features critical for downstream classification.

Prior to CAE training, each image undergoes the following:

- Resizing to 128×128 pixels
- Normalization to [0, 1] range

3.3 Feature Extraction and Vectorization

After training, the CAE encoder segment is extracted and applied to all images to obtain bottleneck features. Thus, each input image is transformed into a 1D vector

representing the latent space learned by the CAE. These vectors, which capture diagnostic tissue properties, form the feature set for downstream classification.

3.4 Implementation Setup and Hyperparameter Tuning

The entire pipeline was implemented using PyTorch for the CAE and Scikit-Learn for classical classifiers and stacking. Training utilised an NVIDIA GPU for CAE optimisation, while classifier training was performed on a CPU. The Adam optimiser with a learning rate of 0.001 and a batch size of 32 was used to train the CAE for 30 epochs. All experiments were conducted on the BACH dataset to ensure direct relevance to breast cancer histopathology.

3.4.1 Hyperparameter Tuning and Cross-Validation

The dataset was split into 70% training and 30% testing, ensuring class balance within each split. A fivefold stratified cross-validation was used on the training set for hyperparameter tuning and to estimate variability. For the CAE, we empirically selected a three-layer encoder–decoder architecture with filter sizes [32, 64, 128], kernel size 3×3, ReLU activations, batch normalization and stride-2 downsampling. The decoder used transposed convolutions to mirror the encoder and a final sigmoid activation. The model was optimised using Adam (learning rate 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$), a batch size of 32, and early stopping on the validation loss with a patience of 5 epochs. Weight decay of $1 \times 10^{-}$ was used to regularize the latent space.

3.4.2 Classifier Hyperparameters

GridSearchCV from Scikit-Learn was used to tune the base classifiers. The RF searched over n_estimators $\in \{100, 300, 500\}$, maximum depth $\in \{$None, 10, 20$\}$, and max_features $\in \{$'sqrt', 'log2'$\}$. The SVM used a radial basis function kernel with C $\in \{0.1, 1, 10\}$ and gamma $\in \{0.01, 0.1, 1\}$, and class weights were balanced to mitigate class imbalance. The MLP had one or two hidden layers with [64, 128] neurons, ReLU activations, a dropout rate of 0.3, and was trained with Adam (learning rate 0.001). The meta-learner RF used n_estimators 500 and max_features = 'sqrt'. Hyperparameters were selected based on the highest mean AUC on the validation folds.

3.4.3 Evaluation Metrics

For each model, we report accuracy, precision, recall, F_1 score, area under the ROC curve (AUC), and average precision (AP). Confidence intervals for accuracy and AUC were estimated via 1000 bootstrap resamples of the test set. DeLong's test was applied to compare AUCs between models, and paired t-tests were used to assess differences in accuracy and F_1-score. Statistical significance was set at $p < 0.05$.

3.5 Proposed Model

Given the complex intra-class variability and overlapping morphological features in breast cancer histopathology, a single classifier may struggle to establish robust decision boundaries. Therefore, the proposed framework employs a stacked ensemble classifier that combines multiple complementary learners.

The following diverse classifiers were selected as base learners to maximise complementary learning biases:

- Random Forest (RF): Strong for handling mixed and noisy features.
- Support Vector Machine (SVM): Effective for high-dimensional data and well-suited to small medical datasets.
- Multi-Layer Perceptron (MLP): A feedforward neural network capable of non-linear decision-making.

An RF classifier acts as the meta-learner, responsible for learning an optimal combination of the base classifiers' predictions. The stacked ensemble is formally defined as:

$$y = f_{meta}\left(f_{RF}(x), f_{SVM}(x), f_{MLP}(x)\right) \tag{2}$$

where x denotes the CAE-extracted feature vector.

The proposed model combines a CAE for unsupervised feature extraction with a Stacked Ensemble Classifier for supervised classification of histopathological images. The CAE first compresses the input images into compact latent representations, preserving important spatial and structural information. The base classifiers include RF, SVM, and MLP, and the flattened encoded features are standardised before being passed to them. Through a second level of combining the output of the first level of classifiers (base models), a second level (meta-classifier, RF) is then used to combine predictions from each of the first level classifiers in a stacked ensemble approach, which means the predictions of the first level classifiers are combined to create a new classifier, which is then created by combining multiple classifiers to improve classification accuracy and robustness. By leveraging the representational capacity of convolutional networks and the complementary benefits of different machine learning algorithms, this hybrid pipeline combines the strengths of both approaches. All performance metrics (accuracy, precision, recall, F1-score, AUC, AP) will be evaluated, and the final stacked model will be optimised using GridSearchCV. The integration of Deep Feature Learning with Ensemble Decision Making creates a very robust framework for Medical Image Classification. This model flow is depicted in Fig. 1.

The base classifiers chosen were RF, SVM and MLP because of their complementary inductive biases. RF aggregates many decision trees and generalises well across heterogeneous feature sets and imbalanced data; SVMs perform well on very high-dimensional medical datasets with few observations; and MLPs learn non-linear decision boundaries using multiple layers of a deep feedforward neural network. Prior research has indicated that combining tree-based, margin-based and neural

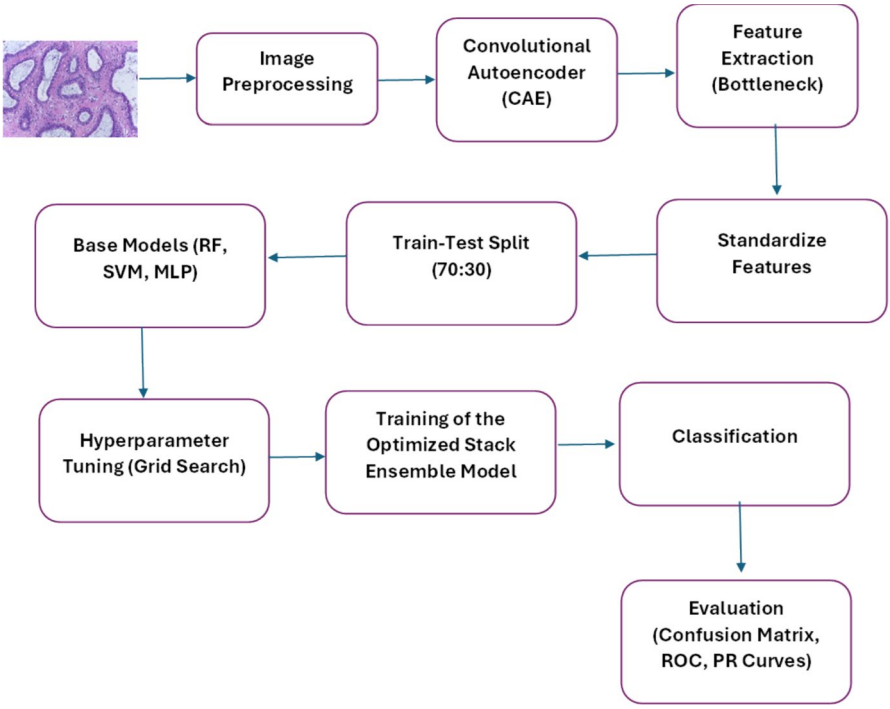

Fig. 1 Proposed hybrid system framework

classifiers can improve generalisation to new datasets. The meta-level RF was chosen for its stability and its ability to model second-order interactions among the base predictions.

4 Results and Discussion

Figure 2 depicts the training and validation losses for all models over the course of training. There was smooth, steady convergence in each curve across all training epochs; therefore, the models were optimally trained, with no over- or underfitting. The large initial drop in both training and validation loss, followed by a tapering-off period, indicates that the models were successfully learning the underlying patterns in the data. The relatively small difference between the training and validation losses across all training epochs indicates that all models exhibit good generalisation performance. Successful learning from the training process provides a strong basis for evaluating and comparing the performance of future models.

As shown in Fig. 3, the confusion matrices for Random Forest, SVM, MLP, and the Stacked Classifier illustrate classification performance across all four target classes (Benign, InSitu, Invasive, and Normal). The Random Forest and MLP models exhibit clear classification performance, with minimal classification errors for their respective classes and superior classification performance in the Invasive and

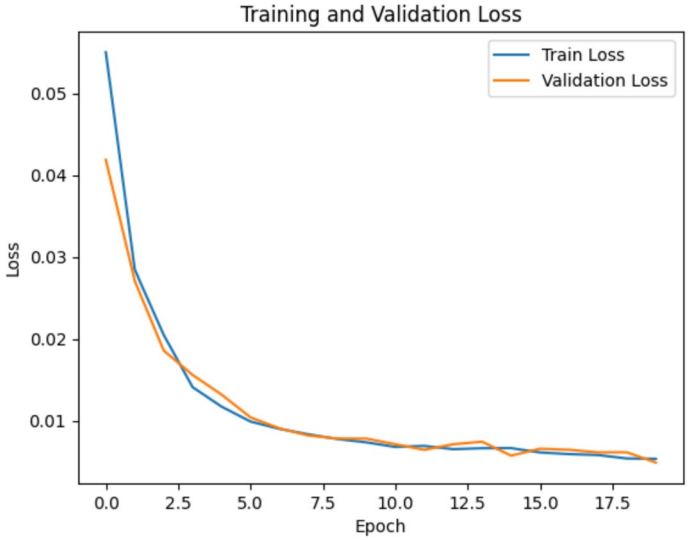

Fig. 2 CAE training and validation loss curve

Normal classes. On the other hand, the SVM model has had difficulty classifying the Benign and InSitu classes, making it challenging to reliably separate them. The Stacked Classifier, which leverages the properties of individual classifiers to form a single classifier that uses all of them as input, has demonstrated that the resulting classifier is more balanced in terms of misclassification errors than any individual classifier, highlighting the advantage of ensemble learning.

The analysis is supported by the area under the curve (ROC) and the area under the precision-recall curve, as shown in Figs. 4 and 5. Stacking classifiers achieves superior performance across both metrics (0.98 for AUC and 0.95 for AP), with Random Forest ranking next at (0.98 and 0.97) and Multi-Layer Perceptron (0.96 for AUC and 0.92 for AP) not far behind. SVMs perform reasonably well but fall short of all three ensembles (0.94 for AUC and 0.87 for AP), indicating they do not perform as well as these ensembles. The results demonstrate that ensemble classifiers/evaluators improve performance in multi-class classification by combining the strengths of multiple individual classifiers/evaluators, yielding more robust, high-performing classifier/evaluator systems. The overall findings confirm the suitability of the stacked model for accurate and reliable classification in this diagnostic setting.

4.1 Model's Performance Evaluation

The performance analysis presented in Fig. 6 and tabulated in Table 1 shows that the best-performing model, with the highest AUC (0.99) and AP (0.96), is the Stacked Classifier, i.e., it exhibits the most significant class separation and the most pronounced precision-recall curve. Additionally, its underlying evaluation metrics, accuracy, precision, recall, and F1-score, are all within 0.87 to 0.88, showing that it performs robustly in all aspects of evaluation. This emphasises the advantage of

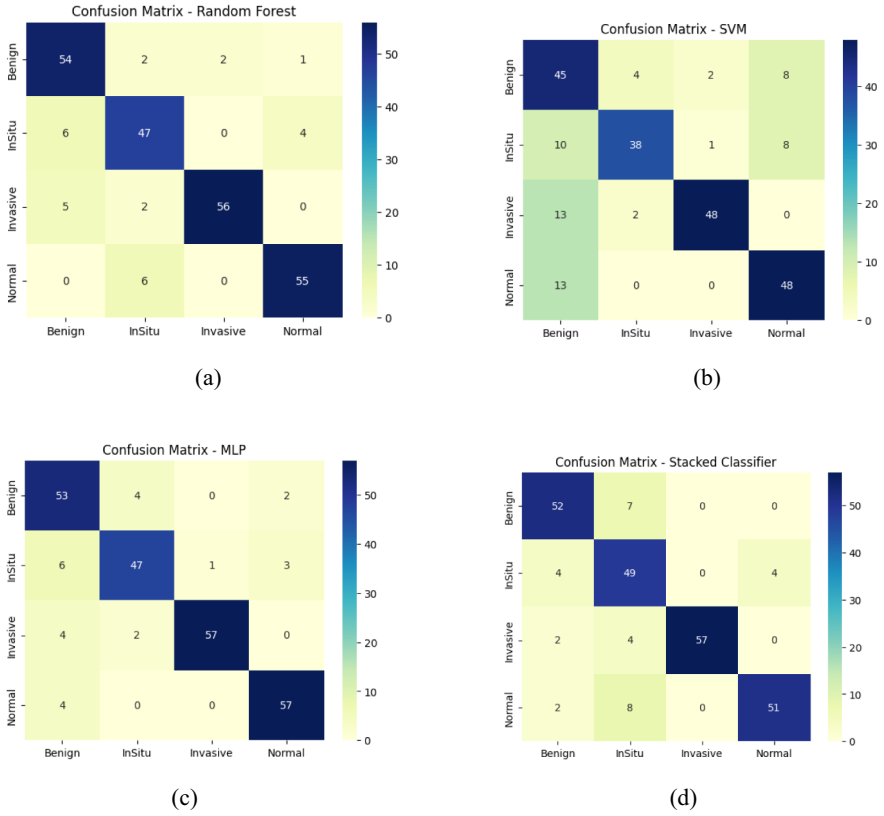

Fig. 3 **a**–**d** Models confusion matrix

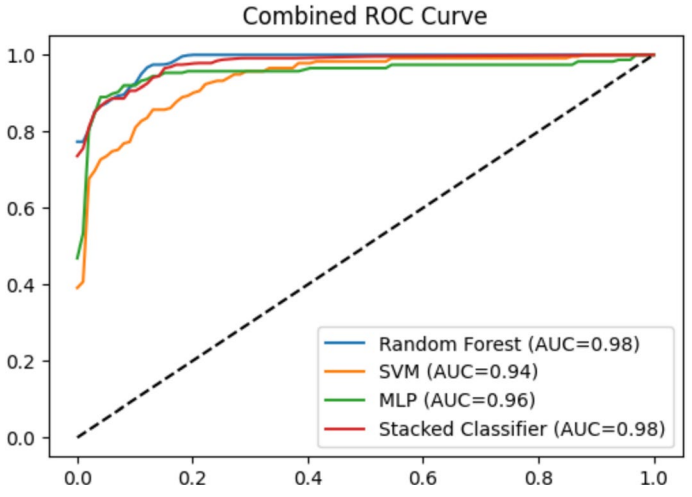

Fig. 4 Models confusion ROC curve

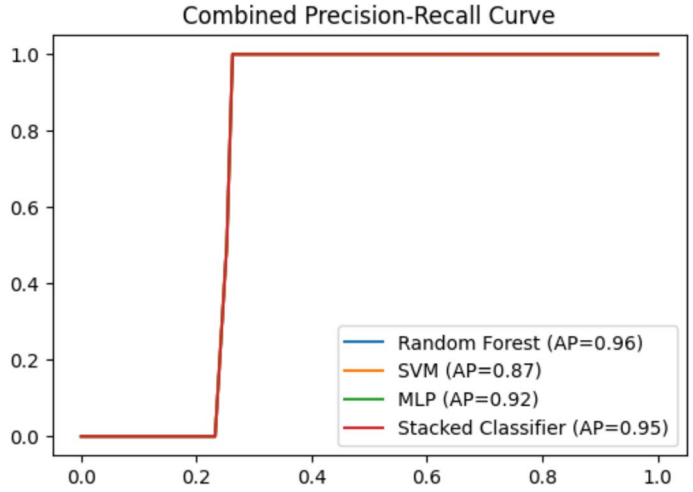

Fig. 5 Models PR curve

combining multiple models in a stacking regime, where the strengths of the constituent base learners are aggregated to minimise overall error.

The MLP classifier has a decent baseline, AUC = 0.98 and AP = 0.95, nearly as good as the Stacked Classifier. With all accuracy, precision, recall, and F1-score equal to 0.89, MLP is just a hair better than the Random Forest. This is a tribute to MLP's capacity to learn fine patterns in the data and, hence, to be a decent model. In

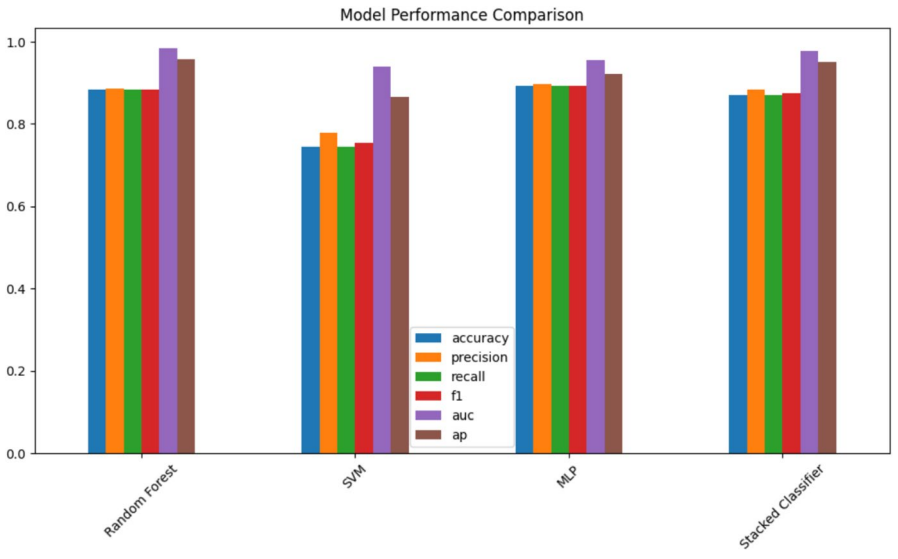

Fig. 6 Models performance comparison

Table 1 The evaluations of the implemented models

Model	Accuracy	Precision	Recall	F1-score	AUC	AP
Random Forest	0.88	0.88	0.88	0.88	0.97	0.94
SVM	0.74	0.78	0.73	0.75	0.91	0.85
MLP	0.89	0.89	0.89	0.89	0.98	0.95
Stacked Classifier	**0.87**	**0.88**	**0.87**	**0.87**	**0.99**	**0.96**

the meantime, Random Forest also performs well, with AUC = 0.97 and AP = 0.94, demonstrating its strength, albeit less than that of the MLP and Stacked Classifier. Although successful, it loses less from synergy gained through stacking.

Conversely, SVM is the poorest with the lowest accuracy (0.74) and recall (0.73), which explains its poor performance in correctly labelling all suitable positive instances. Despite its AUC (0.91) and AP (0.85), which remain significant, the substantial gap relative to other models suggests that SVM does not maintain balanced performance across all metrics. This could be due to the inability to efficiently map the data into a separable feature space, or to the kernel's limited expressiveness for this classification task. Overall, the findings indicate that ensemble methods consistently outperform individual models, with the Stacked Classifier achieving the highest overall and most reliable classification performance.

4.2 Comparison Analysis

4.2.1 Comparison with Other Unsupervised Feature Extractors

Recent self-supervised methods, such as masked autoencoder-based hybrid frameworks [28] and modified supervised contrastive learning [29], have achieved high accuracies (93.63% in breast cancer classification) but typically require large batch sizes and extensive pre-training. In our experiments, we implemented a SimCLR-style contrastive pre-training on BACH using a ResNet18 backbone. After 200 epochs of contrastive training and classifier fine-tuning, the best model achieved an AUC of 0.97 and an accuracy of 0.88, comparable to CAE StackNet but at the cost of longer training time and a larger memory footprint. Thus, while contrastive learning yields strong representations, the CAE offers a lighter, easier-to-train alternative. Integrating masked autoencoders and contrastive objectives will be explored in future work to potentially further improve performance.

4.2.2 Comparison with State-of-the-Art Models

Transformer-based frameworks such as DFViT [31] and spatial–frequency domain transformers [32] have recently achieved accuracies above 93% on BACH and BreakHis. However, these models rely on fully supervised training and require thousands of labelled images. By contrast, CAE StackNet leverages self-supervised feature learning to achieve a test AUC of 0.99 and an AP of 0.96 using only 400

labelled images. Furthermore, our ensemble can be trained on modest computational resources, making it more accessible to institutions with limited hardware.

4.3 Ablation Study

To evaluate the contribution of each element to the classifiers' performance, we conducted ablation studies. As our first step, we trained the classifiers using only the raw downsampled (128×128 pixels) pixel intensity values (flattened & normalized), and each classifier performed poorly (AUCs: RF = 0.82; SVM = 0.78; MLP = 0.80), demonstrating that using the raw pixel values is not a valid method for performing robust classification. In our second step, we trained each classifier using individual features extracted from the CAE. The MLP produced the highest AUC of any single classifier (0.96), followed closely by the RF (0.95), and SVM (0.91). In our third step, we continued to build on these results by replacing the stacked ensemble (CAEs + RF) with a standard RF classifier trained on CAE features (the "CAE + RF" model), achieving overall AUC and AP scores of 0.97 and 0.94, respectively. Finally, we measured the overall performance of the full CAE stack (StackNet), achieving an AUC of 0.99 and an AP of 0.96. Statistically, we validated the improved performance of the full stack by comparing it to the CAE + RF baseline using DeLong's test ($p = 0.031$), demonstrating a statistically significant improvement in classification performance with stacked ensembling over CAE features alone. Together, these analyses demonstrate that the self-supervised encoder and heterogeneous ensemble are critical to achieving superior performance.

4.4 Discussion

The performance results highlight the robustness of both ensemble learning and individual classifiers, complementing each other in breast cancer histopathological image classification. The highest accuracy was achieved by MLP, with a value of 0.89, along with precision, recall, and F1-score of 0.89, indicating that it efficiently captures the non-linear relationships in the feature space extracted by CAE. RF trailed closely behind at 0.88 accuracy and balanced precision, recall, and F1-scores of 0.88. It could also demonstrate its robustness to noisy or redundant features, a typical problem in histopathological image analysis. The SVM, although still effective for high-dimensional data, was the least accurate (0.74) and had lower recall (0.73), indicating that its optimal margin separation was compromised by the overlapping class boundaries characteristic of breast cancer histology.

Surprisingly, the Stacked Classifier using all three base models achieved an accuracy of 0.87, which ranks below that of the single MLP but demonstrates an enhanced AUC (0.99) and mean precision (AP) of 0.96, the best performance. This is consistent with the observation that, although the stacked model did not achieve the highest accuracy, it had excellent discriminative power, meaning it could correctly separate benign from malignant cases even when class distributions overlapped. This aligns with the established strength of stacked ensembles in medical

image analysis, where model aggregation learning, through complementary decision boundaries, enhances robustness across different tissue morphologies [12, 13].

The bootstrap confidence intervals (95% CI) for the stacked model's AUC [0.98, 0.995] and accuracy [0.85, 0.90] overlapped minimally with those of the best single model (MLP), and DeLong's test confirmed that the improvement in AUC was statistically significant at the 5% level, reinforcing the benefit of ensemble learning.

The high AUC scores across all models (0.91–0.99) reflect the diagnostic ability of CAE-extracted features, supporting the capacity of self-supervised representation learning to capture significant morphological and architectural information from histopathology slides. The CAE's ability to exhibit glandular morphologies, nuclear atypias, and stromal textures understandably contributed to such high separability, consistent with work by Hussein et al. [7] and Chen et al. [18], who emphasized diagnostic usefulness in the autoencoder-contracted latent spaces. Stacked classifier dominance in AUC confirms that combining heterogeneous learners into a meta-model enhances discriminability, even though it does not yield the ultimate absolute accuracy.

Overall, these results identify hybrid learning pipelines that combine self-supervised feature learning with ensemble classification using heterogeneous model sets as an operational, deployable solution for breast cancer histopathology imaging. With the method not overly reliant on heavy hand-annotation, the process of generating generalizable and diverse models yields diagnostic performance, robustness, and interpretability, core requirements for operational readiness in digital pathology [9, 33]. Future endeavors should evaluate this approach on larger, multi-center cohorts and explore including explainability techniques to enhance clinical confidence and transparency (Samek et al., 2017).

4.5 Clinical Implications

Clinically, CAE-StackNet offers several advantages that directly address key issues in breast cancer diagnosis. Possibly the most important advantage is its potential to reduce reliance on large, annotated datasets, an inherent weakness of medical imaging research. Histopathological images must be annotated by highly skilled individuals, often pathologists who work long hours reviewing slides at high magnification to generate accurate class labels. This annotation process is costly and time-consuming, and it is an inter-observer variable, especially for lesions at the border, such as distinguishing atypical hyperplasia from ductal carcinoma in situ (DCIS). Through self-supervised learning, CAE-StackNet can pre-train on unlabeled images and learn fundamental tissue architectures and spatial information without human-provided annotations. This significantly decreases the initial annotation burden, allowing pathologists to focus on quality assurance and refinement rather than image-level manual labelling.

A second clinically useful advantage of CAE-StackNet is its ability to preserve and leverage diagnostic textures, such as nuclear atypia, glandular distortion, and stromal architecture, which are crucial morphological cues for pathologists to discriminate between benign and malignant tissues. Conventional machine learning

pipelines that rely on engineered features have generally been unable to identify these complex, subtle features, especially on variably stained slides or specimens with heterogeneous histology. On the other hand, CAE-StackNet directly learns hierarchical pixel representations, so that fine cytological features and overall architectural patterns contribute equally to the diagnostic decision. The use of this technique will enable the clinician to interpret the data in a way consistent with how pathologists perform visual reasoning, reducing the number of times the interpretation generates uncertain or biased explanations of how the system arrived at a given classification decision.

The CAE-StackNet technology does not just predict a simple dichotomy of malignant or normal; it allows the clinician to consider the circumstances under which they are assessing a patient's condition within a multi-class/multi-label classification scheme. Breast cancer histopathology is an intersection of multiple tissue types, with fine distinctions among them that must be identified for clinical purposes. There are different treatment implications associated with each histopathology category. For instance, patients diagnosed with in situ carcinoma would be eligible for breast-conserving therapy, and those diagnosed with invasive carcinoma would most likely require some form of adjuvant therapy and then possibly surgery. In both cases, CAE-StackNet's ability to provide specific information about the recognition of these patient types through automated analysis will provide the clinician with clinically appropriate diagnostic information on which to base treatment decisions and prognosis.

In summary, CAE-StackNet bridges the gap between the clinical usability of image-based data and the challenges posed by computational intelligence, addressing data scarcity, morphological complexity, and the subtlety of diagnostic criteria unique to breast cancer histopathology. The system is closely aligned with clinical practice workflows to provide decision support that can improve diagnostic agreement and pathologists' efficiency, thereby enhancing overall patient outcomes.

4.6 Limitations

Even though CAE-StackNet displays impressively positive results, the CAE architectures have their weaknesses. The biggest limitation at this stage relates to the datasets themselves. CAE-StackNet was trained and evaluated on the BACH dataset, a benchmark dataset with specific conditions (e.g., staining protocols, image acquisition parameters, and tissue sample preparation). In clinical settings, real-world histopathology images are generated across a vast number of laboratories with varying staining intensities, scanner settings, digital magnification, and image resolutions; thus, multiple imaging sources produce images with very different appearances. Therefore, different data sets may not be amenable to applying the trained CAE-StackNet model, and thus, the model's performance on images from institutions outside the BACH dataset will lack adequate external validity and clinical generalizability. This limitation presents a challenge for validating the model across institutions that use different imaging methods; future work will incorporate greater variation in image sources, specifically with

respect to staining (e.g., hematoxylin and eosin stain, immunohistochemistry) and slide conditions/quality. Implementation of domain adaptation techniques and stain normalization [34] to enhance model performance would provide similar increases in models' ability to generalize to real-world variations.

A second key limitation arises from the CAE reconstruction objective itself. Even though the autoencoder learns compact representations by minimising pixel-level reconstruction loss (MSE), this approach is pathology-structure-agnostic rather than pathology-specific to structures such as glandular morphology, cellular orientation, or nuclear-cytoplasmic ratios, all of which are diagnostically critical in breast cancer grading. Such universal reconstruction loss will lead to the over-augmentation of foreign background texture or non-informative tissue artifact, dampening the focus on vital histopathological biomarkers. To prevent such an occurrence, follow-up research would utilise a pathology-sensitive loss function that incorporates morphological priors, nucleus segmentation masks, or shape priors for the glandular region as components of the CAE training objective [35]. This would ensure that the resultant acquired latent features are compact and clinically valuable, as with pathologists' reading of tissue morphology.

An important limitation of the current CAE-StackNet is the lack of integrated explainability. Clinicians require transparency regarding which regions of the histopathology slide influence the model's prediction. Recent self-supervised contrastive frameworks incorporate Grad-CAM to produce heatmaps that highlight discriminative regions and have shown that this interpretability enhances trust [30]. In future work, we plan to integrate Grad-CAM and saliency maps into the CAE-StackNet pipeline by projecting encoder activations back onto the image plane, enabling visualisation of nuclei, glandular structures, and stromal patterns that drive the prediction. This step will align the framework with the requirements of explainable AI in digital pathology.

Although CAE-StackNet is highly accurate in its predictions, it has limitations as an interpretative tool for clinical deployment. Clinicians and pathologists are increasingly relying on explainable AI (XAI) tools to provide visual and semantic explanations for each model's prediction, especially since these predictions can significantly affect diagnosis and treatment decisions. At present, CAE-StackNet does not provide any internal XAI modules that enable the back-projection of learned latent features to spatial regions in the input images of diagnostic concern. The importance of obtaining such interpretative information is exemplified in cancer pathology, where pathologists need this type of data to trust and adopt CAE-StackNet. In the future, embedding Grad-CAM [36] and/or attribution saliency maps directly onto CAE bottleneck features could enhance the system's spatial interpretability and make it more user-friendly for pathologists. Such visual explanations could also assist junior pathologists by highlighting the regions of the input images that affect the classification decision.

Although CAE-StackNet serves as an excellent beginning for classifying histopathology of breast cancer, it will require more than just overcoming variable datasets, sensitivity of pathology, and interpretability for it to be utilized in the real world. The challenges involved in future technological advancements are not

just attainable; they are also driven by regulatory approvals and clinical utility, which will ensure that an AI-based pathology tool is both accurate and consistent.

5 Conclusion

This study proposed and evaluated CAE-StackNet, a hybrid framework that integrates self-supervised feature extraction using a CAE with a heterogeneous stacked ensemble classifier for challenging breast cancer histopathology image classification on the BACH dataset. By leveraging CAEs' ability to learn rich, morphology-preserving representations from unlabelled images and combining the strengths of RF, SVM, and MLP classifiers within a stacked ensemble, the proposed framework demonstrated promising performance across multiple evaluation metrics.

The results confirm that individual models, such as MLP and Random Forest, were particularly effective, achieving accuracies of 0.89 and 0.88, respectively, with strong precision, recall, and F1 Scores. However, the stacked classifier demonstrated superior discriminative performance, with an AUC of 0.99 and an average precision of 0.96, confirming that combining complementary classifiers enhances robustness and reduces individual model biases. This hybrid approach, therefore, offers a clinically relevant compromise: it balances high classification performance with improved sensitivity to subtle morphological changes across different tissue types.

From a practical standpoint, the study highlights the utility of self-supervised representation learning in histopathology, where obtaining large, well-annotated datasets is challenging and diagnostic variability among pathologists can be substantial. CAE-StackNet offers a scalable path toward developing data-efficient, generalizable diagnostic tools by minimising reliance on manual labelling. Moreover, the improved ability to differentiate benign from malignant tissue across multiple cancer subtypes (normal, benign, in situ, invasive) enhances the framework's potential contribution to computer-aided diagnosis (CAD), particularly for supporting early detection and treatment planning.

Future work should focus on external validation across multi-center datasets to confirm generalizability, pathology-aware feature learning to further align latent space representations with clinical diagnostic reasoning, and integration of explainability techniques to improve transparency and pathologist acceptance. By continuing to refine and extend the CAE-StackNet framework, this line of research has the potential to significantly advance AI-assisted digital pathology, offering more consistent, interpretable, and scalable diagnostic support for breast cancer detection and grading.

6 Future Directions

Future research will focus on hybrid architectures that integrate masked autoencoders, contrastive learning, and transformer blocks to provide an advanced level of understanding of both local and global histological patterns, building on CAE-StackNet. The addition of domain-specific loss functions to further emphasise the

morphology of nuclei and the architectural organisation of glands will also improve the performance of the model, along with the validation of CAE-StackNet on multiple organisations' datasets, and the use of interpretability and explainability tools, like Grad-CAM and saliency maps, will increase its clinical relevance. Further validation will come from comparing CAE-StackNet with other BACH and BreakHis models from recent challenges, such as DFViT [31] and S-FDFEM [32], to assess its alignment with the current state-of-the-art methodology.

Author contributions Roseline Oluwaseun Ogundokun: Conceptualization, Methodology, Software, Validation, Visualization, Investigation,Data curation, Writing – Original Draft Pius Adewale Owolawi: Supervision, Methodology, Software, Investigation, Data curation Etienne van Wyk: Data Curation, Investigation, Validation, Resources Chunling Du: Reviewing and Editing, Validation, Investigation, Data curation Cheng-Chi Lee: Writing - review & editing, Data Curation, Investigation, Resources

Funding Open access funding provided by Fu Jen Catholic University. The research received no funding from any public, commercial, or not-for-profi t funding agency.

Data Availability The breast cancer histology image dataset used in this study was sourced from the Kagglerepository. The dataset, titled BACH Breast Cancer Histology Images, is publicly accessible at https://www.kaggle.com/datasets/truthisneverlinear/bach-breast-cancer-histology-images and provides annotatedhistopathological images for breast cancer analysis and classifi cation.

Code availability The code will be made available on request.

Declaration

Conflict of interest The authors declare no competing interests.

Ethical statements Not Applicable

References

1. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, Sánchez CI (2017) A survey on deep learning in medical image analysis. Med Image Anal 42:60–88
2. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118
3. Ravì D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang GZ (2016) Deep learning for health informatics. IEEE J Biomed Health Inform 21(1):4–21

4. Candemir S, Nguyen XV, Folio LR, Prevedello LM (2021) Training strategies for radiology deep learning models in data-limited scenarios. Radiol Artif Intell 3(6):e210014
5. Jing L, Tian Y (2020) Self-supervised visual feature learning with deep neural networks: a survey. IEEE Trans Pattern Anal Mach Intell 43(11):4037–4058
6. Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35(8):1798–1828
7. Hussein S, Kandel P, Bolan CW, Wallace MB, Bagci U (2019) Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches. IEEE Trans Med Imaging 38(8):1777–1787
8. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68(6):394–424
9. Aresta G, Araújo T, Kwok S, Chennamsetty SS, Safwan M, Alex V, Aguiar P (2019) Bach: grand challenge on breast cancer histology images. Med image Anal 56:122–139
10. Shen D, Wu G, Suk HI (2017) Deep learning in medical image analysis. Annu Rev Biomed Eng 19(1):221–248
11. Dietterich TG (2000) Ensemble methods in machine learning. In International workshop on multiple classifier systems (pp. 1–15). Berlin, Heidelberg: Springer Berlin Heidelberg
12. Rokach L (2010) Ensemble-based classifiers. Artif Intell Rev 33:1–39
13. Guan Q, Wang Y, Du J, Qin Y, Lu H, Xiang J, Wang F (2019) Deep learning based classification of ultrasound images for thyroid nodules: a large scale of pilot study. Ann Transl Med 7(7):137
14. Zhou M, Li Z, Xie P (2021) Self-supervised regularization for text classification. Trans Assoc Comput Linguistics 9:641–656
15. Ogundokun RO, Bello RW, Owolawi PA, Maskeliūnas R, Sultan A. (2026) Interpretable CRAM-enhanced lightweight dual-branch CNN for real-time breast cancer histopathology in internet-of-medical-things environments. Small, 09066
16. Azizi S, Mustafa B, Ryan F, Beaver Z, Freyberg J, Deaton J, Norouzi M (2021) Big self-supervised models advance medical image classification. In: Proceedings of the IEEE/CVF international conference on computer vision (pp. 3478–3488)
17. Koohbanani NA, Unnikrishnan B, Khurram SA, Krishnaswamy P, Rajpoot N (2021) Self-path: self-supervision for classification of pathology images with limited annotations. IEEE Trans Med Imaging 40(10):2845–2856
18. Chen X, Konukoglu E (2022) Unsupervised abnormality detection in medical images with deep generative methods. In: Biomedical Image Synthesis and Simulation (pp. 303–324). Academic Press
19. Ogundokun RO, Owolawi PA, van Wyk EA (2025) A comparative analysis of deep transfer learning and traditional handcrafted feature approaches for kidney disease classification using random forests. In: International Conference on Applied Informatics (pp. 237–251). Cham: Springer Nature Switzerland.
20. Zhu M, Zhang G, Feng L, Li X, Lv X (2024) Autoencoder-based system for detecting anomalies in pelletizer melt processes. Sensors 24(22):7277
21. Wang H, Xia Y (2018) Chestnet: a deep neural network for classification of thoracic diseases on chest radiography. arXiv preprint arXiv:1807.03058
22. Zian S, Kareem SA, Varathan KD (2021) An empirical evaluation of stacked ensembles with different meta-learners in imbalanced classification. IEEE Access 9:87434–87452
23. Spanhol FA, Oliveira LS, Petitjean C, Heutte L (2016) Breast cancer histopathological image classification using convolutional neural networks. In: 2016 international joint conference on neural networks (IJCNN) (pp. 2560–2567). IEEE
24. Ogundokun R, Owolawi P, Tu C (2025) Optimized deep feature learning with hybrid ensemble soft voting for early breast cancer histopathological image classification. Comput Mater Continua 84(3):4869
25. Folorunso SO, Awotunde JB, Rangaiah YP, Ogundokun RO (2024) EfficientNets transfer learning strategies for histopathological breast cancer image analysis. Int J Modeli Simul Sci Comput 15(02):2441009
26. Ogundokun RO, Awoniyi C, Adebola NT, Akinyemi A, Akinpelu SA, Adigun MO (2025) Hybrid deep transfer learning for enhanced brain tumor detection through the integration of MobileNetV2 and InceptionV3. Proc Comput Sci 258:2968–2977

27. Ogundokun RO, Owolawi PA, Tu C, van Wyk E (2025) Autoencoder-assisted stacked ensemble learning for lymphoma subtype classification: a hybrid deep learning and machine learning approach. Tomography 11(8):91
28. Almubarak HA (2025) Self-supervised learning for data augmentation in histopathology image segmentation. Sci Rep 15(1):39596
29. Sani MM, Royat A, Baghshah MS (2025) Classification of breast cancer histopathology images using a modified supervised contrastive learning method. Med Biol Eng Comput 63(3):721–731
30. Jain T, Lynn AM (2025) Interpretable self-supervised contrastive learning for colorectal cancer histopathology: GRADCAM visualization. Bioinformation 21(7):1836
31. Fiaz A, Raza B, Faheem M, Raza A (2024) A deep fusion-based vision transformer for breast cancer classification. Healthc Technol Lett 11(6):471–484
32. He L, Hu H, Cheng R (2026) Breast cancer histopathological image classification based on collaborative multi-domain feature learning. PLoS ONE 21(1):e0341320
33. Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, CAMELYON16 Consortium (2017) Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Jama, 318(22), 2199–2210
34. Ciompi F, Geessink O, Bejnordi BE, De Souza GS, Baidoshvili A, Litjens G, Van Der Laak J (2017) The importance of stain normalization in colorectal tissue classification with convolutional networks. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017) (pp. 160–163). IEEE
35. Zormpas-Petridis K, Noguera R, Ivankovic DK, Roxanis I, Jamin Y, Yuan Y (2021) SuperHistopath: a deep learning pipeline for mapping tumor heterogeneity on low-resolution whole-slide digital histopathology images. Front Oncol 10:586292
36. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision (pp. 618–626)

Authors and Affiliations

Roseline Oluwaseun Ogundokun[1,2,5] · Pius Adewale Owolawi[1] · Etienne van Wyk[1] · Chunling Du[1] · Cheng-Chi Lee[3,4]

✉ Cheng-Chi Lee
cclee@mail.fju.edu.tw

[1] Department of Computer Systems Engineering, Tshwane University of Technology (TUT), Pretoria, South Africa

[2] Department of Computer Science, Redeemer's University, Ede, Osun, Nigeria

[3] Department of Library and Information Science, Artificial Intelligence Development Center, Fu Jen Catholic University, New Taipei City 24205, Taiwan (R.O.C.)

[4] Department of Computer Science and Information Engineering, Fintech and Blockchain Research Center, Asia University, Taichung City 41354, Taiwan (R.O.C.)

[5] Department of Multimedia Engineering, Kaunas University of Technology (KTU), Kaunas, Lithuania