



ORIGINAL

## Fusion enhancement and learning model for histopathological image analysis using learning approaches

### Mejora de la fusión y modelo de aprendizaje para el análisis de imágenes histopatológicas mediante enfoques de aprendizaje

N Hari Babu<sup>1</sup> , Enireddy Vamsidhar<sup>1</sup> 

<sup>1</sup>Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation. Guntur, Andhra Pradesh, India.

Cite as: Babu NH, Vamsidhar E. Fusion enhancement and learning model for histopathological image analysis using learning approaches. Data and Metadata. 2025; 4:511. <https://doi.org/10.56294/dm2025511>

Submitted: 08-04-2024

Revised: 11-08-2024

Accepted: 09-12-2024

Published: 01-01-2025

Editor: Adrián Alejandro Vitón Castillo 

Corresponding Author: N Hari Babu 

#### ABSTRACT

Breast cancer is the most prevalent form of the disease and the primary cause of cancer-related deaths among women globally. Early detection plays a pivotal role in substantially diminishing both the morbidity and mortality rates associated with this disease in women. Consequently, the development of an automated diagnostic system holds promise in enhancing the precision of diagnoses. To automatically classify breast cancer microscopy images stained into two distinct classifications—normal tissue and benign lesions—this study introduces a graph-based convolutional neural network with hybrid optimization (G-CNN) that makes use of a dataset that was specially selected for this purpose. The network layer is capitalized in our suggested model to extract reliable and abstract information from input photos. Initially, we used 5-fold cross-validation (CV) to optimize the suggested model on the original dataset. Our framework demonstrated a 98 % accuracy rate and a 0,969 kappa score. It also received an average AUC-ROC score of 0,998 and a mean AUC-PR value of 0,995. In specific terms, it displayed 96 % and 99 % sensitivity, respectively, about the supplied photographs. Examining normalized photos, the suggested architecture outperformed the other approaches in terms of colour normalization methodology performance. These findings underscore the superior performance of our proposed model compared to both the baseline approaches and established prevailing models using default settings. Furthermore, it becomes evident that while existing normalization techniques delivered competitive performance, they fell short of surpassing the results obtained from the original dataset.

**Keywords:** Histopathological Images; Prediction; Cancer; Deep Learning; Optimization.

#### RESUMEN

El cáncer de mama es la forma más prevalente de la enfermedad y la principal causa de muerte por cáncer entre las mujeres de todo el mundo. La detección precoz desempeña un papel fundamental en la disminución sustancial de las tasas de morbilidad y mortalidad asociadas a esta enfermedad en las mujeres. Por consiguiente, el desarrollo de un sistema de diagnóstico automatizado resulta prometedor para mejorar la precisión de los diagnósticos. Para clasificar automáticamente las imágenes de microscopía de cáncer de mama teñidas en dos clasificaciones distintas -tejido normal y lesiones benignas-, este estudio introduce una red neuronal convolucional basada en grafos con optimización híbrida (G-CNN) que hace uso de un conjunto de datos especialmente seleccionado para este fin. La capa de red se capitaliza en nuestro modelo sugerido para extraer información fiable y abstracta de las fotos de entrada. Inicialmente, utilizamos la validación cruzada (CV) quintuple para optimizar el modelo sugerido en el conjunto de datos original. Nuestro marco demostró una tasa de precisión del 98 % y una puntuación kappa de 0,969. También obtuvo una puntuación media AUC-ROC de 0,998 y un valor medio AUC-PR de 0,995. En concreto, mostró una sensibilidad del 96 % y el 99 %, respectivamente, sobre las fotografías suministradas.

Al examinar las fotografías normalizadas, la arquitectura propuesta superó a los demás enfoques en cuanto al rendimiento de la metodología de normalización del color. Estos resultados ponen de relieve el rendimiento superior de nuestro modelo propuesto en comparación tanto con los enfoques de referencia como con los modelos predominantes establecidos utilizando la configuración predeterminada. Además, resulta evidente que, aunque las técnicas de normalización existentes ofrecían un rendimiento competitivo, no llegaban a superar los resultados obtenidos con el conjunto de datos original.

**Palabras clave:** Imágenes Histopatológicas; Predicción; Cáncer; Aprendizaje Profundo; Optimización.

## INTRODUCTION

As per the Global Cancer Statistics 2023 report, breast cancer stands out as the most prevalent malignancy and holds the leading position in causing fatalities related to cancer among women globally. Specifically, 2,26 million cases, representing 11,7 % of all female cancer cases, were identified, resulting in 0,69 million deaths, representing 6,9 % of the overall cancer-related deaths in 2023. Consequently, an early and comprehensive understanding of the pathophysiology of breast tumours is crucial because it may help lower the incidence of morbidity and death among women globally. This cancer is understood to be a broad spectrum of illnesses with unique biochemical, clinical, and treatment-related characteristics. It mostly arises from abnormalities in the breast's epithelial tissues, although it can also spread into the stroma, mammary duct, or lobes.

Even while mammography, MRI, and ultrasound are commonly used in regular clinical tests for breast cancer, these non-invasive procedures may not be able to adequately capture the varied characteristics displayed by breast cancers. As such, pathological analysis continues to be a crucial reference point for understanding the pathogenesis of these malignancies. The procedure entails obtaining tissue samples on glass slides, and colouring the slides after that to get a more comprehensive representation of the tumors' morphological and immune phenotypical traits. Following this, pathologists meticulously examine these stained slides under a microscope to arrive at a potential diagnosis of breast cancer. The comprehensive steps involved in a detailed outline of the histological technique have been provided. However manually interpreting a histopathological image takes a lot of time and effort, and it might lead to biased results. Furthermore, the pathologists' topic experience plays a major role in the morphological standards applied in the manual examination. For instance, the research found that participating pathologists had about 75 % overall concordance rate.

To mitigate inter-pathologist variances throughout the breast cancer diagnostic process, Tools for computer-aided diagnosis (CAD) have been created to help pathologists improve diagnostic accuracy. However traditional computerized diagnostic approaches, which range from machine-learning methods to rule-based systems, may prove inadequate in handling the consistency between different classes and the variability within the same class of intricate pictures of breast cancer histology. Furthermore, these conventional approaches frequently depend on feature extraction methods such as local binary patterns, speed robust features, and scale-invariant feature transform. Yet, these methods depend on supervised information, potentially leading to biased outcomes when categorizing pictures like these. As a result, the need for a quick and accurate diagnostic answer has accelerated the creation of sophisticated computer models, or "deep learning," that are built on several layers of nonlinear processing units. These models aim to surpass the limitations of traditional approaches by leveraging intricate architectures capable of learning representations directly from the data without explicit feature engineering.

In the realm of activities involving eyesight, the Convolutional Neural Network (CNN) has gained prominence due to its translational equi-variance and invariance properties, characteristics obtained via pooling and sharing of parameters, respectively. Notably, CNNs have surpassed traditional multi-layer with deep CNN architectures demonstrating significant advancements in the last decade. Among these architectures, AlexNet stands as one of the earliest models that achieved commendable accuracy in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Subsequently, the VGG network introduced the concept of using smaller-sized convolutional filters in a deep network, securing the second position at ILSVRC. Szegedy et al. then presented the Inception architecture achieves an efficient receptive field by using many smaller convolutional filters claiming the top spot at ILSVRC in the same year. Following this, He et al. addressed the performance degradation that occurs with increased network depth by introducing residual connections, resulting in their first-place position at ILSVRC.

Using publically accessible datasets like BreakHis and BACH, several current research have used these previously described architectures—which were pre-trained on ImageNet—to successfully categorize photos related to breast cancer histopathology. On the BreakHis dataset, Jiang et al., for example, proposed an altered version of the ResNet model and achieved cutting-edge multiclass classification accuracy. Comparably, for the multiclass classification of breast microscope pictures, either a single pre-trained network or an ensemble of pre-trained architectures was employed in the top papers in the BACH challenge. On the BACH dataset, Elmannai et al. recently attained state-of-the-art performance by recognizing the efficacy of residual connections and Inception modules as feature extractors. We employed the Xception model (extreme inception) in our

methodology, which is based on a successful combination of residual connections and Inception. Serving as a feature extractor, the Xception model consistently delivers results in classifying photographs of histopathology at various magnifications. As a feature extractor, the trained Xception model was utilized our approach successfully applies the ideas presented in these earlier studies to the extraction of important characteristics from histopathological pictures. The following describes the purpose and importance of this study:

The primary aim was to annotate and create a specialized dataset intended for the categorization of pictures from breast cancer histology into several groups, such as invasive carcinoma, normal tissue, in situ cancer, and benign lesions. The created dataset builds on our earlier research on binary classification that was released. Another key objective was to assess the performance of four commonly used colour normalization techniques, aiming to standardize the appearance of histopathology images. The study aimed to introduce a deep learning model that utilizes multi-level to increase classification accuracy, features are taken out of the graph-based model's intermediate layers. The goal of the study was to refine the suggested model for accurate image categorization of breast cancer histology using hybrid cuckoo for swarm optimization (HCSO), both in their original form and after undergoing normalization procedures. Special emphasis was placed on enhancing accuracy, especially concerning the carcinoma classes. Notably, this study represents the first instance of annotating a new private dataset, introducing a comprehensive and effective computational model that uses a graph-based network as a feature extractor and evaluating the outcomes of widely used stain normalization methods. In conclusion, the suggested model produced dependable and consistent outcomes in accurately classifying breast cancer histopathology images into distinct categories, surpassing existing results that are cutting-edge. The parts that follow are the order in which this document is structured:

- 1) The principal aim of this investigation was to enhance breast histopathology classification accuracy at the image level by employing a patch-level G-CNN classification approach. To achieve this goal, we devised an effective approach for patch-level training that boasts computational speed. Our methodology involved fine-tuning a G-CNN model, initially trained on the cancer dataset, by incorporating convolutional, max-pooling and thick layers in the latter stages. With this model, we were able to extract patch characteristics.
- 2) To transform patch-level data into image-level predictions across four classes—normal, benign, in situ, and invasive—and two classes—cancerous and non-cancerous—our method used a two-stage model that leveraged a neural network.
- 3) The histopathology dataset, which is accessible to the public, was used to assess our model's performance. With four classes and two classes, State-of-the-art classification accuracy of 97,50 % and 98,6 %, respectively, were obtained using our technique.

This essay's remaining sections are organized as follows: Section 2 provides a detailed description of the existing methodology. Section 3 elaborates on the process. The experimental findings and a critique of our methodology are included in Section 4 including a discussion of the outcomes. Finally, in Section 5, we conclude our work, highlight its limitations, and outline potential future directions for research.

### Related works

From the existing literature, it is noted that several researchers have undertaken reviews about distinct aspects of histopathological image analysis, including studies on histopathological image analysis itself,<sup>(16)</sup> stain normalization techniques,<sup>(17)</sup> segmentation methods,<sup>(18)</sup> and classification methodologies.<sup>(19)</sup> This section specifically focuses on summarizing and detailing the findings derived from the review conducted on histopathological image analysis. The literature encompasses a multitude of reviews and surveys focused on different aspects of histopathological image analysis, each shedding light on distinct facets of this field: Choudhary et al.<sup>(20)</sup> provided an overview of the latest Computer-Aided Diagnosis specifically designed for histopathology image analysis (CAD) technology. They stressed the importance of employing standard datasets for evaluating CAD systems, making comparison and analysis simpler. A review of many techniques for the examination of Breast Cancer Histological Images (BCHI) was conducted by Ezhil et al.<sup>(21)</sup> They discussed the intricacy of tissue characteristics that demand examination to enhance the system's robustness.

Fondón I et al.<sup>(22)</sup> delved into Machine Learning (ML) techniques utilized for histopathological image analysis. A list of accessible datasets for the investigation of breast cancer and generalized image classification techniques were presented in Hameed, Z.<sup>(23)</sup> including supervised, unsupervised, and DL classifiers. Irshad and colleagues<sup>(24)</sup> examined various methods for histopathological image analysis, including the identification of nuclei, segmentation, extraction of features, and classification. Benchmark datasets, difficulties with topics covered included the value of resilience in technical and therapeutic settings, as well as microscopic image segmentation. In their investigation of cutting-edge methods for segmenting images for feature extraction and disease classification, Mostafa et al.<sup>(25)</sup> addressed the features of histological images. Ozaki et al.<sup>(26)</sup> explored the use of DL approaches in digital pathology across segmentation, detection, and classification tasks. They suggested enhancing classifier quality by combining hand-crafted features with DL approaches. A study of

several image analysis techniques in histological image analysis was conducted by Spanhol et al.<sup>(27)</sup> focusing on the cell detection problem and listing pertinent limitations. Aresta et al.<sup>(28)</sup> talked about potential research areas in diagnostic machine learning (ML) and identified difficulties in computational pathology procedures. He et al.<sup>(29)</sup> assembled the most recent methods and uses for large-scale medical image analytics. In their overview of DL approaches in medical image analysis, Holzinger et al.<sup>(30)</sup> addressed difficulties in assessing BCHI.

Aksac et al.<sup>(31)</sup> investigated magnification-independent multi-category classification issues and conducted a thorough assessment of DL approaches for automated breast cancer detection using the BreakHis dataset. Additional evaluations by Aksac, A. et al.<sup>(32)</sup> covered a wide range of subjects, such as deep neural network designs for histopathological image analysis, DL algorithms for breast cancer identification, lymph node support for breast cancer pictures, ML and DL approaches for breast carcinoma diagnosis, and issues involved in developing CAD systems for cancer diagnosis. These reviews collectively contribute a comprehensive understanding of the various methodologies, challenges, and advancements in histopathological image analysis using a spectrum of computational techniques.<sup>(33)</sup> Indeed, the existing literature highlights several reviews concentrating on distinct facets of automating histopathological image analysis. However, a comprehensive and detailed review encompassing it appears that every facet of histological image analysis, including colour normalization, feature extraction, possible Region of Interest (ROI) identification and segmentation, and classification—are absent.<sup>(34,35)</sup>

While numerous review papers exist, they often tend to focus on specific segments or aspects within breast histopathological image analysis. These reviews provide in-depth insights into individual elements such as detection algorithms, classification methodologies, or segmentation techniques, rather than offering a comprehensive overview covering the entire spectrum of histopathological image analysis.<sup>(36)</sup> This observation underscores the need for a comprehensive and integrated review that synthesizes every facet of histopathology image analysis, including segmentation, ROI detection, colour normalization feature extraction, and classification within a singular, encompassing framework.<sup>(37,39)</sup> Such a comprehensive review would be instrumental in consolidating diverse methodologies and advancements across the entire pipeline of histopathological image analysis, providing a more holistic understanding of this field's current state and future directions.<sup>(40)</sup>

## METHOD

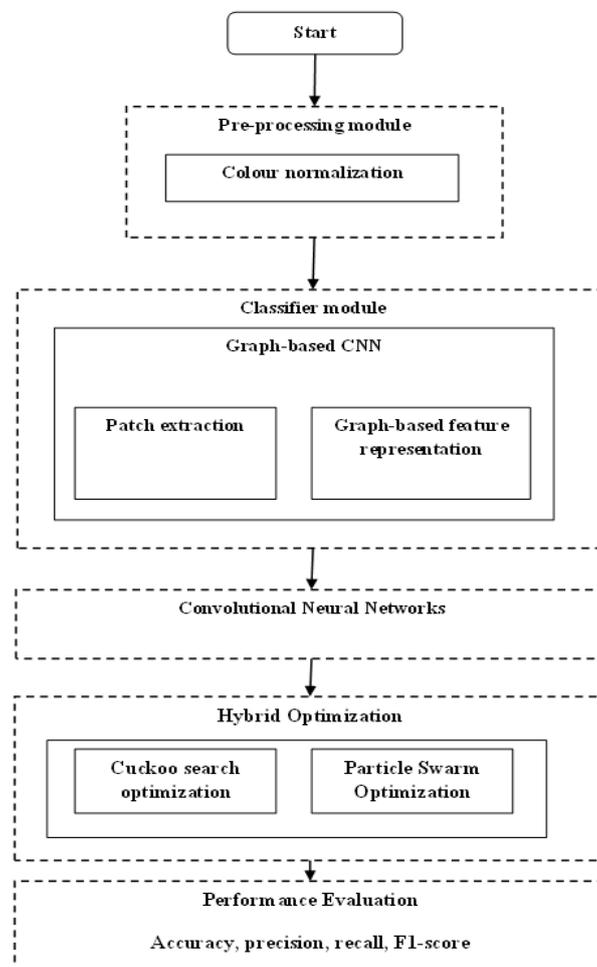


Figure 1. Flow diagram of the proposed model

Histopathological image analysis serves the purpose of categorizing images into benign and malignant classifications, serving as an essential tool for decision-making. Several stages are included in this analysis utilizing image processing techniques, encompassing colour normalization, classification and optimization. While traditional methods follow a step-by-step pipeline, modern approaches, particularly based on DL adopt a complete learning approach. The conventional method uses colour normalization as a first step to standardize colour and lighting variances between photos. Using the characteristics that were retrieved, DL algorithms are trained to categorize the photos into several groups. The methods established for colour normalizing are described in depth in this section, which is followed by the discussion of methods for detecting and classifying potential regions of interest within histopathological images. Additionally, an overview of diverse techniques employed for classification specifically tailored for Breast Cancer Histological Images (BCHI) will be presented. Finally, hybrid optimization approaches are adopted to attain global outcomes. These techniques constitute essential components in the comprehensive analysis of histopathological images, aiming to enhance the accuracy and reliability of image-based diagnosis and decision-making processes. Figure 1 shows the workflow.

### **Dataset**

The Breast Cancer Histology images (BACH) collection from the ICIAR 2018 was used in this investigation. It consists of whole-slide pictures and breast histology microscopy Hematoxylin and eosin (H&E) staining. Depending on which type of cancer is most common or whether there is no cancer at all in each microscope picture, the pictures are categorized as invasive, benign, in situ, or normal. Two medical specialists annotated these photos, and those with inconsistent annotations were removed from the dataset. There are 400 microscope photos in total in this collection, 100 images for each of the four classes. The microscope pictures have particular characteristics and are saved in a.tiff format. RGB Pixel Scale: 0,42  $\mu\text{m}$   $\times$  0,42  $\mu\text{m}$ ; Dimensions: 1536  $\times$  2048 pixels; Memory Space: Approximately 10-20 MB per image, Label: Image-wise classification. These images serve as a valuable resource for training and validating models regarding the field of image analysis in histopathology of breast cancer, providing a diverse set of samples representative of various histological classes, annotated by expert medical professionals.

### **Pre-processing**

The process of slide preparation significantly impacts the results of histopathological images due to potential variations in colour distribution. To address these variations, colour normalization processes are commonly employed. However, a notable observation in the literature is the lack of suitable evaluation metrics for assessing the efficacy of colour normalization techniques. Identifying appropriate metrics for evaluating the effectiveness of the colour normalization process is thus an important endeavour in this field. Additionally, graph-based CNN (G-CNN) present a promising avenue for future exploration. G-CNN can learn and generate images based on specified constraints. This potential enables G-CNN to facilitate the transfer of colour distributions from a reference image to an input image. Leveraging G-CNN for such colour distribution transfer tasks holds promise for enhancing the standardization and normalization of histopathological images, thereby potentially improving the consistency and reliability of subsequent analyses. Exploring the utilization of G-CNN in this context remains an intriguing area for future research and development within histopathological image analysis.

### **Methodology description**

The methodology employed in this study encompassed several key steps in the analysis of histopathological images: initially, the dataset of 400 histology images was separated into sets for testing, validation, and training. This division involved an 80/20 split, with 80 % of the images allocated to the intermediate set (further divided for training and validation) and 20 % reserved for the test set. Within the training and validation sets; images were segmented into patches for more granular analysis. A patch-level classifier was trained using these patches and a pre-processing step involving color stain normalization was applied to normalize the images. To extract characteristics from these patches, several deep-learning models that had already been trained were used. Patch-level predictions were processed through different classifiers, and ensemble techniques were employed to merge and analyze these predictions. This process aimed to improve accuracy and robustness in classifying patches into malignant (invasive and in situ) and benign (normal and non-cancerous) classifications. Instead of using traditional ensemble methods, a graph-based neural network was devised. This network managed the shift from patch-level to image-level classification, effectively categorizing the final images into four groups: normal, benign, in situ, and invasive. This suggested methodology's pipeline consisted of many steps, including dataset partitioning, pre-processing, classification, and optimization model for image-level classification. This approach aimed to streamline and enhance the accuracy of the classification process for histopathological images, potentially offering a more efficient and accurate analysis framework.

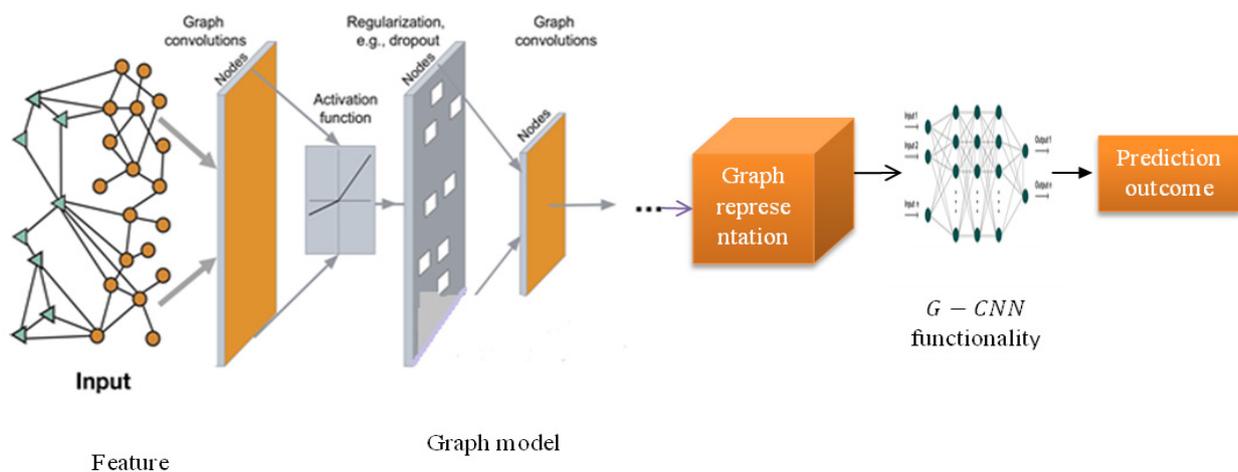
## Classification

Once the extraction of image candidates is completed, an initial classification of images is conducted for each particle utilizing G-CNN. To evaluate the effectiveness, we compared the proposed with distinct architectures. These architectures varied based on the dimensionality of patches (2D/3D) and the consideration of image strength. In the subsequent sections, this work will initially delineate the suggested method, which emphasizes the G-CNN architecture and patch extraction. Following this, this section provides comprehensive details about the alternative architectures employed for comparison purposes. These architectures are aimed at discerning differences based on patch dimensions and the incorporation of image strength in the classification process.

**Patch Extraction:** only 3D patches taken from the CT scan surrounding the tumour under study are used in the suggested G-CNN. On the reformatted plane, a centred patch encompassing a 32x32 pixel region is removed along the particle's principal axis. This main axis, which is created by resampling the original picture using cubic interpolation, is defined by the first eigenvector of the Hessian matrix and isotropically spaced at 0,625 mm. The particular centre particle's patch and the patches surrounding four neighbouring particles that belong to the same picture in its reformatted direction are taken into consideration to create the 3D patch. As a result, on the reformatted plane along the picture point of interest, every patch is a 32 x 32 x 5 voxel small picture section.

**Architecture:** the architecture of the proposed G-CNN is depicted in figure 2. Three convolutional layers, two dropout layers, one max-pooling layer, and three fully-connected layers make up the network. Our hypothesis posits that employing a 3D approach yields superior performance compared to 2D CNNs. This superiority arises from the crucial role of connectivity information in distinguishing between the two vascular trees, a factor that 2D-CNNs can only emulate through specific workarounds. Additionally, this work conjectures that the inclusion of additional information does not significantly enhance the network's ability to learn image characteristics. To validate this hypothesis, we formulated different architectures for comparison alongside the proposed approach. For this purpose, with several approaches, we were able to extract 2D and 3D patches surrounding every particle location.

**Feature representation:** in our approach for extracting 3D patches, we maintained the same method as the one used in the suggested approach. In particular, we took into account the surrounding 32x32 pixel area on the particles' reconstructed plane for 2D patches. However since we knew that the connection between picture spots may offer crucial information for classification, we came up with a method to merge the 2D patch of the target sample using the two nearest particles whose orientations were the most similar. This image amalgamation was achieved by merging the extracted 2D patches with the original patch with more channels designed to make a more reliable comparison and mimic 3D representation. These patches are referred to as 2.5D. As a result, a three-channel, 32x32-pixel 2.5D patch surrounding a potential vessel is created using the CT scans of the centre point and the two nearest spots. To explore whether G-CNN could benefit from structural information in differentiating between normal and benign, we incorporated tumour-enhanced images in alternative architectures. Given that the study focused on particles from the given structure, the inclusion of the tumour-enhanced image would appear redundant, but we hypothesized that the G-CNN might use this image to identify and evaluate extra local information.



**Figure 2.** Architecture of G-CNN

To improve, we used a Frangi filter by adjusting the parameters through grid search to optimize the enhancement of CT images. Subsequently, from these improved pictures, we took 2D and 3D patches from the vicinity of the potential locations and we used two different approaches to merge these patches into the CT

patches. Using the first method, we added the improved photos as additional patch channels. In the second approach, we concatenated the three patches at the fully connected level and let the network learn from them separately. For instance, a 2,5D patch with enhanced images added as extra channels comprises nine channels total, including the centre point's CT image plus the two closest ones and 32x32 pixels surrounding the particle point. In contrast, better photographs used as separate inputs for the 2D network employ three patches of 32x32 pixels and three channels (the centre point and the two adjacent ones).

The 2D architecture, shown in figure 2, shows the G-CNN when it uses 2.5D patches, using either the histopathological picture alone or with augmented images added as additional channels or separate inputs. Despite variations in inputs, the fundamental CNN architecture remains consistent across all three cases. This architecture comprises five 2D convolutional layers, the network topology is shown by three fully-connected layers, dotted with two max-pooling and two dropout layers. Convolutions and max-pooling techniques are applied to the three inputs simultaneously to integrate picture strength as independent inputs with the local information obtained from the histopathology analysis, all while maintaining the same architecture. This allows the network to automatically pick up new features from each patch. Therefore, just before the fully linked processes are started, the resulting weights and biases are concatenated. For each of these networks, the hyper-parameters have been meticulously chosen through empirical experimentation to optimize performance specifically for the problem at hand. This careful parameter selection aims to achieve the best results for the classification task.

### Graph-based feature representation

The cancer classification process may exhibit inconsistencies due to intertwined and touching areas within the vascular trees, particularly when conducting classification on individual particles without explicitly considering smoothness at the tree level. To mitigate these inconsistencies observed during the initial G-CNN-based classification, using an automated graph-cuts (GC) technique, we improve the classification outcomes. This GC strategy combines graph theory principles with energy minimization methods to identify minimum cuts in the graph, thereby defining an optimal solution. The graph is composed of a set of vertices  $V = \{v_i | i=1, \dots, N_{nodes}\}$  and edges connecting different nodes  $e = \{(v_i, v_j) | i, j=1, \dots, N_{nodes}\}$ . It includes two terminal (or virtual) nodes, namely the source,  $s$ , and the sink,  $t$ , denoted as  $V_t = \{s\} \cup \{t\}$ , alongside a collection of real non-terminal nodes,  $V_{n-t}$ . The edges consist of connections between pairs of non-terminal nodes ( $n_{links}$ ),  $\epsilon_{n-t} = \{(v_i, v_j) | v_i, v_j \in V_{n-t}\}$ , and connections between a terminal node and a non-terminal node ( $t_{links}$ ),  $\epsilon_t = \{(s, v_i) \cup (v_i, t) | v_i \in V_{n-t} \text{ and } s, t \in V_t\}$ .

Since energy is specified as a function, the minimal cut problem is handled from the perspective of energy reduction.

$$\xi = \xi_{bound} + \xi_{reg} \quad (1)$$

The obtained by minimizing this energy function, which effectively determines the optimal partitioning of the graph, enhancing the accuracy of the classification. The formulation of the energy function involves two essential components: the boundary term ( $\xi_{bound}$ ) and the regional term ( $\xi_{reg}$ ).  $\xi_{bound}$  signifies the coherence among neighbouring nodes reflecting the details of the connection. The weights connected to the links decide it. The regional term, represented by  $\xi_{reg}$  represents the probability of each class (in this case, the similarity between arteries and veins). This term is denoted by the weights assigned to the links. The objective of minimizing the combined energy function is to achieve a classification that optimizes both the coherence among neighbouring nodes and the likelihood estimation for different classes. In this particular problem scenario, the probabilities obtained from the CNN's pre-classification step are used to generate the arterial-venous similarity score that forms the regional term, which is defined by the edges  $\epsilon_t$ . As a result, the weights links of are directly set to these probabilistic estimations: they represent the probability-based assessments obtained from the initial G-CNN-based classification step.

$$w(s, p_i) = p_{image}(p_i) \quad (2)$$

$$w(p_i, t) = p_{tumour}(p_i) = 1 - P_{non-tumour}(p_i) \quad (3)$$

The boundary term, represented by  $\epsilon_{n-t}$ , requires information about the connectivity among particles ( $P(p_i)$ ). As individual particles themselves don't offer information on direct connectivity, a conservative structural connectivity approach is employed. Using a radius of  $neigh = 3$  mm, this technique entails creating connections between every particle and every other particle inside a cylinder formed along the vessel's main direction. Additionally, to manage concerns in densely populated areas, a limit is set on the number of allowable

connections per node ( $N_{con} = 5$ ). These parameters were adjusted through empirical tweaking following multiple tests. To determine the weight of the links, three primary characteristics are taken into consideration to indicate the strength of the bonds between particles:

- 1) Scale Consistency ( $w_{\sigma}(p_1, p_2)$ ): Particles with similar scale are more likely to be neighbours.
- 2) Particle Proximity ( $w_{dist}(p_1, p_2)$ ).
- 3) Direction Consistency ( $w_k(p_1, p_2)$ ): The greater the likelihood that two particles belong to the same tree, the closer they are in terms of Euclidean distance. determined by the local direction of the particles under consideration and the alignment of the connection vector connecting two particles.

$$w(p_1, p_2) = f(w_{\sigma}(p_1, p_2); w_{dist}w_{\sigma}(p_1, p_2), w_{||}w_{\sigma}(p_1, p_2)) \quad (4)$$

By considering these three characteristics, various weighting functions can be formulated to define the strength of the links: these functions evaluate the relationships between particles based on their scale, proximity, and direction consistency. Occasionally, the method produces isolated sub-trees, which might make categorization more difficult. To address this, an additional step is taken after the initial graph-cut classification to iteratively join all of these separated sub-trees edges. Until the complete graph is made up of a single linked component, the isolated sub-trees are joined iteratively using the following steps:

- a) Designating the primary component as the one with the greatest connectedness.
- b) calculating the Euclidean distance between the points inside the isolated sub-trees and the main component.
- c) Adding an edge to join particles from the isolated sub-trees to the main component, with a weight derived from equation (4).

Using the suggested min-cut/max-flow conversion, the minimal cut, or  $C_{min}$ , is calculated after the final graph has been obtained through this iterative process. The graph  $G$  is divided into  $G_1$  and  $G_2$ , two linked components, as a consequence of this algorithm. In the equation mentioned, the parameter  $\alpha$  which has a value of  $\alpha$  was determined as the optimal value through grid search. This parameter serves to balance the contributions of the regional and boundary terms within the graph-cut method.

$$C_{min} = arg \min_{G_1, G_2} (\xi(G_1, G_2)) = arg \min_{G_1, G_2} (\xi_{bound}(G_1, G_2) + \alpha \cdot \xi_{reg}(G_1, G_2)) \quad (5)$$

$$\xi_{bound}(S, T) = \sum_{p_1 \in S \cap v_{n-t}} w(p_1, p_2) \quad (6)$$

$$\xi_{reg}(S, T) = \sum_{p_1 \in S \cap v_{n-t}} w(p_1, t) + \sum_{p_2 \in S \cap v_{n-t}} w(p_2, s) \quad (7)$$

## Optimization

### Cuckoo search

Using a population-based methodology, a stochastic global search method is the Cuckoo Search (CS) algorithm draws inspiration from specific behaviors observed in birds, particularly the flying patterns and breeding habits of specific cuckoo species known as Levy flights in various bird species. Utilizing CS presents two primary advantages: first, it requires minimal configuration of parameters during the initial search phase, and second, it offers ease of interaction even for inexperienced users. However, it's commonly observed that CS might display relatively sluggish convergence speed and lower search accuracy due to diminished diversity within the population. Three fundamental rules emulate the behaviours of cuckoos, rendering them suitable for application as a tool for optimization: First, every cuckoo lays one egg, which is then arranged in a nest at random. Secondly, nests containing superior-quality eggs are perpetuated into the subsequent generation. Thirdly, there is a finite number of host nests, and a host bird has a chance  $pa \in [0, 1]$  of identifying an egg laid by a cuckoo. When a host bird finds an egg, it can either discard it or abandon the nest entirely to construct a new one. To put it simply, every egg in a nest in the CS algorithm is a possible solution, and each cuckoo's egg represents a novel answer. By iteratively replacing the candidate solutions with more advantageous created solutions (cuckoo's eggs) depending on their fitness scores, the CS algorithm aims to improve the results. This method consists of two phases: a local random walk controls the exploitation phase, while the switching parameter  $pa$ , a global Levy flight random walk, controls the exploration phase. Throughout the whole procedure, this parameter controls the population update. In the first step, we create new solutions around the best nest,  $g_{Best}$ , in the current generation using the global Levy flight random walk. Concerning the CS algorithm, let's consider a population consisting of  $N_p$  eggs denoted as  $X = \{X_1, \dots, X_{N_p}\}$ , where each egg

comprises C decision variables represented as  $X_i = \{x_1, \dots, x_C\}$ . Mantegna’s algorithm is used to identify the step in the Levy flight, as seen below:

$$stepsize_i^{(k)} = 0.006 \frac{u_k}{|v_k|^{\frac{1}{\alpha}}} (X_i^k - g_{Best}) \quad (8)$$

Where  $X_i^k$  represents the  $i^{th}$  egg within the population during the  $k^{th}$  iteration. In our study, the value of  $\alpha$  is fixed at 1.5. The variables  $u$  and  $v$  follow a normal distribution, with  $u \sim N(0, \sigma_u^2)$  and  $v \sim N(0, \sigma_v^2)$ . The definition of the randomly generated matrix’s standard deviation is as follows:

$$\sigma_u(\alpha) = \left[ \frac{\Gamma(1 + \alpha) \sin\left(\frac{\pi\alpha}{2}\right)}{\Gamma\left(\frac{1+\alpha}{2}\right) \alpha 2^{\frac{\alpha-1}{2}}}\right]^{\frac{1}{\alpha}} \quad (9)$$

$$\sigma_v = 1 \quad (10)$$

Subsequently, the new egg denoted as  $X_i^{(k+1)}$  can be obtained through the following equation:

$$X_i^{k+1} = X_i^k + stepsize_i^k \cdot randn(C) \quad (11)$$

where random scalars selected from the standard normal distribution are denoted by the symbol,  $randn(C)$ . To increase variety and boost global exploration capabilities in a finite number of generations by exposing hitherto unexplored regions inside the search space, we suggest taking an extra step: generating new eggs when superior ones cannot be discovered. Specifically, if a newly generated egg fails to surpass the previous one about the significance of fitness, the suggested Differential Evolution (DE) method to produce another egg, is denoted as  $X_i^{(k+1)}$ . This scheme is formulated as follows:

$$X_i^{(k+1)} = X_r^k + F_m (X_{best}^k - X_{worst}^k) \quad (12)$$

Where,  $X_{best}^k$  and  $X_{worst}^k$  represent the best and worst options, respectively, depending on fitness value within the current population.  $X_i^k$  denotes a solution whose fitness value falls within the range of  $[\text{round}(\varepsilon N_p), N_p - \text{round}(\varepsilon N_p)]$ , where  $\varepsilon \in (0, 0.5)$ . For our study, the value of  $\varepsilon$  is designated as 0.15, and  $F_m$  is uniformly generated within the interval (0,1,1) for each iteration. Employing this scheme yields two significant advantages. Firstly, the target vector gravitates toward superior solutions as it consistently follows the path of the superior ones. Second, it directs the search to areas of the search space that show promise by avoiding the direction of the worst answer. In the second stage, CS continues by employing a biased/selective random walk technique to generate additional eggs. With consideration for the likelihood that cuckoos will be found, the following new egg is created using a crossover operator:

$$X_i^{(k+1)} = \begin{cases} X_i^k + F_c(X_{r_1}^k - X_{r_2}^k) & \text{if } rand[0,1] > p_a \\ X_i^k & \text{else} \end{cases} \quad (13)$$

In this context,  $r_1$  and  $r_2$  represent distinct random integers and  $F_c$  signifies the scaling factor, an evenly distributed random number in the range [0,1]. The operator,  $X_{jump}^k = X_i^k + F_c (X_{r_1}^k - X_{r_2}^k)$  is intended to facilitate a leap to evade local entrapment. However, if  $X_{r_1}^k$  and  $X_{r_2}^k$  are in relative proximity, this method may fail to explore new potential zones. To tackle this issue, in this study, instead of utilizing  $X_{jump}^k = X_i^k + F_c (X_{r_1}^k - X_{r_2}^k)$ , we implement an operator defined as follows:

$$X_{jump}^k = \begin{cases} X_i^k + F_c(X_{r_1}^k - X_{r_2}^k) & \text{if } FDR_i > 10^{-3} \\ X_i^k + F_c[(X_{r_1}^k - X_{r_2}^k) + (X_{r_3}^k - X_{r_4}^k)] & \text{else} \end{cases} \quad (14)$$

In this context,  $r_1, r_2, r_3,$  and  $r_4$  represent distinct random integers. The symbol  $FDR_i$  represents the fitness difference ratio (FDR) of the  $i$ th solution inside the current population, which has the following definition:

$$FDR_i = \left| \frac{f_i - f_{g\text{best}}}{f_{g\text{best}}} \right| \quad (15)$$

Where  $f_i$  represents the fitness value of the  $i^{\text{th}}$  solution, and  $f_{g\text{best}}$  denotes the fitness value of the best solution discovered thus far. Additionally, to augment exploration within the search space during the initial stages and to emphasize exploitation of the best solutions identified as the algorithm progresses towards completion, the parameter  $p_a$  is updated as follows:

$$p_a^{(k)} = pa_{\text{max}} - (pa_{\text{max}} - pa_{\text{min}}) \cdot (k/N_{\text{iter}}) \quad (16)$$

Where  $p_a(k)$  represents the switching parameter  $p_a$  at the  $k^{\text{th}}$  iteration;  $[pa_{\text{min}}, pa_{\text{max}}]$  signifies the range within which  $pa$  fluctuates with  $pa_{\text{min}}$  set at 0,01 and  $pa_{\text{max}}$  at 0,5;  $N_{\text{iter}}$  denotes the highest number of iterations that are permitted. Finally, a greedy technique is used to create the next-generation solution. Every cycle ends with an update to the best solution found thus far. The pseudo-code in Algorithm 1 can be used to explain the steps of the improved CS algorithm.

### Particle Swarm Optimization (PSO)

The Particle Swarm Optimization (PSO) algorithm is a population-based stochastic optimization technique that is employed as a global search strategy, much like the Cuckoo Search algorithm. It takes inspiration from the gregarious and cooperative behaviours seen in a flock of birds. Owing to its simple graphical representation and very few tunable parameters, PSO has become one of the most used methods, effectively solving a wide range of optimization issues. However, one significant disadvantage of PSO is its susceptibility to being trapped within local optimal solution regions, a concern that has been extensively discussed in comprehensive reviews. In the PSO algorithm, each individual, referred to as a particle, embodies a potential solution within a given population termed as a swarm. These particles are continually updated based on their own experiences and the experiences of their neighbouring particles. The corresponding fitness value is used to assess the quality of a potential solution. Considering a swarm comprising  $N_p$  particles, each particle possesses a position vector denoted as  $X_i = \{x_{i1}, \dots, x_{ic}\}$ , a velocity vector  $V_i = \{v_{i1}, \dots, v_{ic}\}$ . It interacts with nearby particles using the best position  $g_{\text{Best}}$  found in the neighbourhood and its own best position  $p_{\text{Best}}$  found so far. Particles are relocated in the search process by the following equations during the  $k^{\text{th}}$  iteration:

$$V_i^{(k+1)} = w^{(k)}V_i^{(k)} + c_1r_1[pBest^k - X_i^k] + c_2r_2[gBest^k - X_i^k] \quad (17)$$

$$X_i^{(k+1)} = X_i^{(k)} + V_i^{(k+1)} \quad (18)$$

The equations for updating particles in the PSO algorithm involve stochastic weighting of various components using random variables  $r_1$  and  $r_2$ , uniformly distributed within the range  $[0, 1]$ . These components influence velocity adjustment through acceleration coefficients  $c_1$  and  $c_2$  which respectively scale the impact of cognitive and social factors. The inertia weight is denoted as  $w$  plays a significant role and the flying velocity is constrained within a reasonable range  $[V_{\text{min}}, V_{\text{max}}]$ . For this study,  $V_{\text{min}}$  is set to -3 while  $V_{\text{max}}$  is set to 3 imposing limits on particle movement. To enhance the PSO algorithm's performance is adopted to generate a diverse population that provides effective guidance for particles. This approach aids in mitigating premature convergence issues and bolstering the algorithm's exploitation capabilities. Specifically, a differential mutation scheme proposed is formulated as follows:

$$X_i^{(k+1)} = \bar{X}_i^k + F_{p1}(X_{\text{best}}^k - X_{\text{better}}^k) + F_{p2}(X_{\text{best}}^k - X_{\text{worst}}^k) + F_{p3}(X_{\text{better}}^k - X_{\text{worst}}^k) \quad (19)$$

$X_{\text{best}}^k$ ,  $X_{\text{better}}^k$ , and  $X_{\text{worst}}^k$  represent, respectively, the top three randomly chosen particles from the current population for the tournament. For every iteration, the mutation factors  $F_{p1}$ ,  $F_{p2}$ , and  $F_{p3}$  are individually produced from an even distribution throughout the range of  $(0,1)$ . As a result, throughout the optimization process, this method creates triangles in the feasible zone that are different in size and form. Moreover,  $X_i^{(k)}$  denotes a convex combination vector with the triangle as its source and has the following definition:

$$\bar{X}_i^k = \delta_1 X_{\text{best}}^k + \delta_2 X_{\text{better}}^k + \delta_3 X_{\text{worst}}^k \quad (20)$$

The actual weights  $\delta_i$ , where  $i = 1, 2, 3$  are determined by  $\delta_i = p_i / (\sum_{i=1}^3 p_i)$ . Here,  $p_1$ ,  $p_2$ , and  $p_3$  are set as 1,  $\text{rand}(0, 75, 1)$ , and  $\text{rand}(p_2, 1)$ , respectively. The function  $\text{rand}(a, b)$  generates a real number between  $a$  and  $b$ . Considering equation (20), two primary advantages emerge. Firstly, by constructing different-sized and shaped triangles throughout the optimization process, global exploration capabilities are considerably enhanced and a variety of sub-regions around the optimum vectors may be explored. Second, since the convex combination  $X_i^k$  mostly contains the best vector because of its larger weight, it is easier to obtain the global solution assuming the optimal vectors are the ones that all the vectors point toward. Additionally, to keep exploration and exploitation in a correct balance during the optimization process, improving exploration capacity initially and shifting towards increased exploitation later—a parameter control strategy is implemented. In this strategy,  $c_1$  and  $c_2$  are both set to 2 while the updating scheme for the inertia weight denoted as  $w$  is expressed as follows:

$$\bar{X}_i^k = \delta_1 X_{best}^k + \delta_2 X_{better}^k + \delta_3 X_{worst}^k \quad (21)$$

Where  $w(k)$  represents the inertia weight during the  $k^{\text{th}}$  iteration. The range for the inertia weight, denoted as  $[w_{\min}, w_{\max}]$  is specified with  $w_{\min}$  set to 0,4 and  $w_{\max}$  set to 0,9. The steps outlining the enhanced PSO algorithm are detailed in the fictitious code that Algorithm 1 shows.

### Hybrid optimization

Theoretical and experimental studies do not guarantee the prevention of optimization algorithms from being mired in mediocre fixes. To increase the accuracy of the findings and lessen the likelihood of being trapped in small solution zones, it is usual practice to hybridize or mix several methodologies. We present an effective hybrid optimization approach in this research, which we call hybrid leveraging the strategies outlined previously. This algorithm is constructed upon two primary steps. The initial step involves identifying diversity by using hybrid algorithms, and promising areas inside the search space are located. The second phase then concentrates on choosing and revising solutions while maintaining the advantages of both methods. As a result, this strategy seeks to accomplish quick convergence while also somewhat reducing the risk of being caught up in less-than-ideal solutions. Until the termination requirements are satisfied, these two processes are again performed recursively. Algorithm 3 describes the main steps of the hybrid algorithm. Hybridizing CS and PSO algorithms has been used in several research to solve global optimization issues. On the other hand, the suggested approach has unique features that set it apart from other algorithms in the literature. First, the search space's hitherto undiscovered regions are now better probed by the improved CS and PSO algorithms. Secondly, as the hybrid algorithm operates independently during the optimization process, the strengths of both approaches are retained. Thirdly, through a simple yet effective selection mechanism, Potential solutions avoid going in the wrong way as well as going in the optimal one. Additionally, the sharing of solutions aids the hybrid algorithms in compensating for their respective weaknesses.

#### Algorithm 1: Hybrid optimization

//Standard CS

**Input:** Initialize population randomly and establish

**Output:** Attain optimal gBest

**Initialize**

For every nest do

Establish random cuckoo;

Compute fitness function;

Select nest randomly;

If  $f_i$  is superior to  $f_j$  then

Substitute successive nest  $X_j$ ;

Else

Generate cuckoo based on  $X_i'$ ;

Substitute  $X_j$ ;

Evaluate worst nest and construct newer one;

Update gBest to successive generation;

**Until** stopping criteria is fulfilled

//Standard PSO

**Input:** Initialize population randomly and establish and p

**Output:** Attain optimal gBest

**Initialize**

```

For every nest do //Xi
    Establish random cuckoo;
    Compute fitness function;
    Select nest randomly;
If  $f_i$  is superior to  $f_j$  then
    Substitute successive nest Xj
Revise the population size;
Revise inertia weight w
    For every particle Xj do
        Revise position and velocity;
        Evaluate fitness function;
Update gBest and pBest
Until stopping criteria is fulfilled

```

**//Hybrid Optimization**

**Input:** Initialize population randomly of particles and cuckoo and establish  $p$  and  $p$

**Output:** Attain optimal gBest

**Initialize**

```

For every solution do //Xi
    Execute initialize phases of standard cuckoo search;
    Sort and hold newer solution;
    Execute initialize phases of standard particle swarm;
    Update pBest and preserve the solution;
    Determine the present best;
Update gBest successively;
Update X based on finest solution;
Until stopping criteria is fulfilled

```

**Searching process**

This study suggests employing the hybrid algorithm as detailed for the optimization phase. Leveraging the strengths of both the classification and finding the best answer to the picture classification problem involves using both the hybrid algorithm and the model. An overview of the approach used to solve this issue may be found in Algorithm 1. As stated earlier, this study is predicated on the assumption that the initiation and bias correction issues have been partially addressed. The objective here is to provide a well-approximated solution termed a starting point enabling the proposed algorithm to converge swiftly and attain the best possible answer worldwide. In this particular paper, we utilize the hybrid-based approach. This decision was mostly made because the approach provides a collection of solutions as opposed to a single answer. As a result, it makes the suggested algorithm's beginning point selection more flexible. The coefficient of joint variation (CJV) between the WM and GM areas is used as a metric to determine this point. Then, solutions are generated at random around this fixed point to produce the starting population. Here is how the CJV is defined:

$$CJV = \frac{\sigma(GM) + \sigma(WM)}{|\mu(GM) - \mu(WM)|} \quad (22)$$

The formula for CJV involves  $\sigma(\cdot)$  representing the standard deviation and  $\mu(\cdot)$  indicating the mean intensity. A lower CJV value signifies superior performance in this context. To further guarantee that all solutions remain inside the search space and avoid diverging behaviour (figure 3), the boundary requirements for the  $i^{\text{th}}$  possible solution are restricted as follows:

$$x_{ij}^k = \begin{cases} x_{max} & \text{if } x_{ij}^k > x_{max} \\ x_{min} & \text{if } x_{ij}^k < x_{min} \\ +x_{ij}^k & \text{else} \end{cases} \quad (23)$$

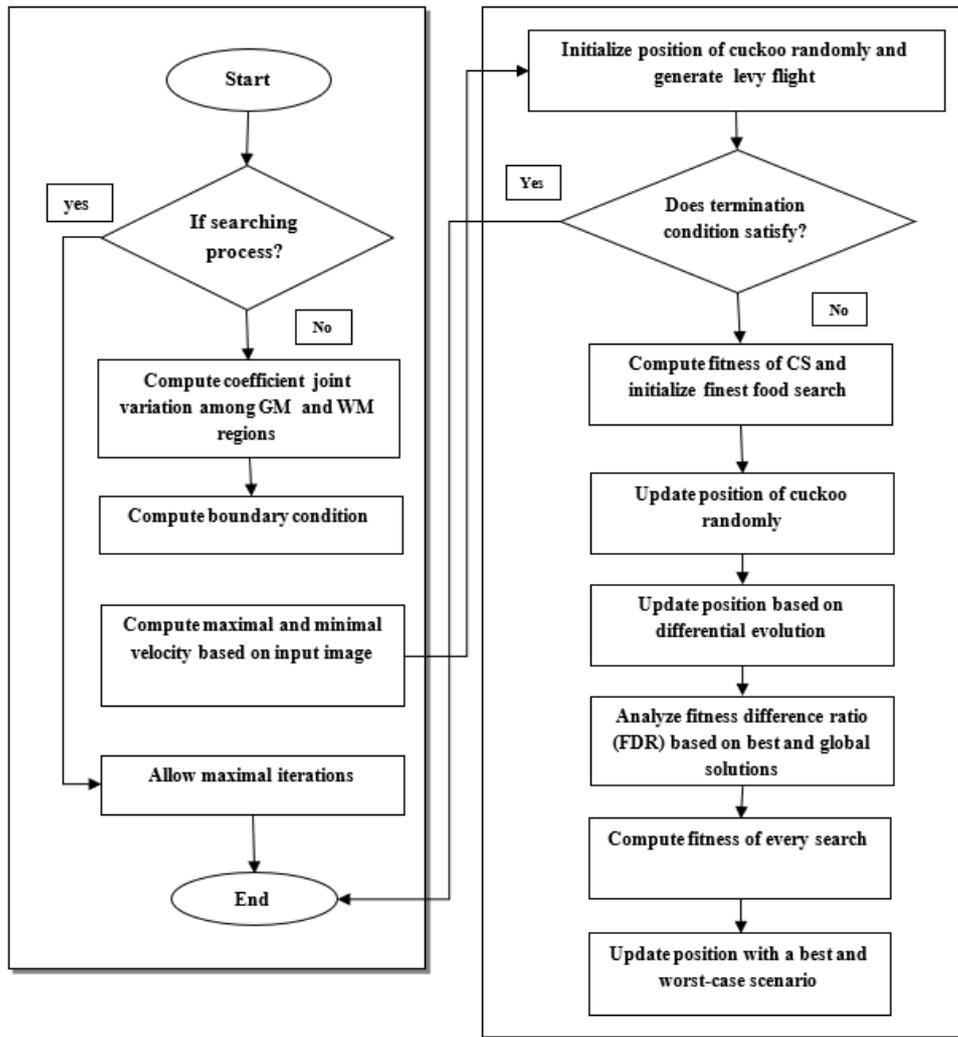


Figure 3. Flow diagram of hybrid model

The boundary conditions for the  $i^{\text{th}}$  potential solution are subject to the following restrictions:  $v_{\min}$  and  $v_{\max}$  indicate the smallest and largest step sizes that are allowed in each dimension. With a specific setting of  $v_{\min} = -v_{\max} = -3$  in this paper.  $\{x_{\min}, x_{\max}\}$  denote the search space’s boundaries in each dimension, which normally match the input image’s minimum and maximum intensity values.

$$x_{ij}^k = \begin{cases} x_{\max} & \text{if } x_{ij}^k > x_{\max} \\ x_{\min} & \text{if } x_{ij}^k < x_{\min} \\ +x_{ij}^k & \text{else} \end{cases} \quad (24)$$

To effectively terminate the hybrid algorithm phase, two standards are set: the highest allowable maximum number of negligible gains in the fitness value and the number of iterations,  $f_{\text{Best}}$ . The phase stops either when  $(\text{iter} > N_{\text{iter}})$  is attained or when  $(|f_{\text{Best}}, \text{new-}f_{\text{Best}}, \text{old}| < 10^{-4}$  occurs  $0, 1 \times N_{\text{iter}}$  times. Lastly, to overcome the challenge of selecting parameter  $\beta$  as described, the CJV criterion is utilized to assess the quality of solutions. This criterion aids in determining the final solution.

### RESULTS AND DISCUSSION

The computational system utilized for this study featured a single GPU system with a P100 GPU, furnished with 128 GB of storage space and 25 GB of RAM. The research was conducted using a MATLAB 2020a environment for code development. To expedite the model development process, several libraries were employed. These included libraries enabling the utilization of built-in layers like Dense and Convolution. Additionally, matplotlib was utilized for visualizing graphs depicting accuracy versus epochs, while sklearn was used for the train-test-split function facilitating the division of the dataset into the requisite sets for training, testing, and

validation. These libraries were instrumental in accelerating the development and analysis of the proposed model, enabling efficient experimentation and faster results retrieval.

### Metrics evaluation

Certainly, accuracy, in the context of evaluating the way a categorization model performs, shows the proportion of successfully predicted labels to the dataset's overall size. The following equation is frequently used to define this metric:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} * 100\% \quad (25)$$

In simpler terms, accuracy assesses the capacity of the model to accurately categorize instances across all classes in the dataset. It's a fundamental metric used to gauge the overall correctness of predictions made by the model and has a percentage attached to it. When assessing a classification model, the percentage of correctly predicted samples in a given class compared to the total number of samples categorized in that class is known as precision. Precision may be computed mathematically using the following formula:

$$Precision = \frac{TP}{TP + FP} * 100\% \quad (26)$$

Put differently, precision expresses how well a model predicts positive outcomes for a given class and demonstrates the model's capacity to misclassify negative instances as positive within that class. Recall, in the context of evaluating a classification model, is the ratio of correctly predicted positive instances to all positive cases in a class that happen. Mathematically, recall is calculated using the following formula:

$$Recall = \frac{TP}{TP + FN} * 100\% \quad (27)$$

Recall, to put it simply, measures how well the model captures all pertinent instances of a given class by accurately identifying all positive occurrences inside that class. The F1 score is a statistic that integrates recall and accuracy into one assessment. It is expressed as follows and is computed as the harmonic mean of recall and precision:

$$F1 - score = \frac{2 * TP}{2 * TP + FP + FN} * 100\% \quad (28)$$

The F1 score offers an aggregate metric that takes into account both false positives and false negatives by striking a balance between recall and accuracy. where there is an unequal class distribution or where recall and precision are equally crucial for assessing a model's performance, it is very helpful. The following definitions apply when assessing a classification model for true positive (TP), true negative (TN), false positive (FP), and false negative (FN):

True Positive (TP): the quantity of cases that are accurately assigned to a specific class.

True Negative (TN): the number of cases that were correctly classified as not fitting into a certain category.

False Positive (FP): the number of cases that are incorrectly allocated to a particular class when they don't belong there.

False Negative (FN): the number of examples that are wrongly classified as not belonging to a particular class when they do.

These metrics are fundamental in evaluating the performance of classification models, helping to assess the model's accuracy in predicting different classes within a dataset.

### Experimental outcomes

The experimentation with transfer learning models involved fine-tuning various pre-trained models on the BACH dataset and the results were assessed based on accuracy. Here's a breakdown of the experimentation process and observations: Initially, several learning models were experimented with and their accuracy on the BACH dataset was evaluated. Models like VGG-16 and VGG-19 despite being relatively smaller in size compared

to deeper models demonstrated comparatively better results in terms of accuracy. The fine-tuning process involved freezing all layers except the last layer in every model. To get class predictions, a thick layer with softmax activation was then implemented. Next, each model's final two layers were unsealed, and the accuracy was determined as more layers were unfrozen progressively. The number of unfrozen layers was optimized based on observations. It was noticed that after unfreezing more than four layers, the model's accuracy didn't improve but rather began to decrease. This trend was attributed to the relationship between the available training data and the increasing number of parameters to train. Typically, the training model utilized around 50-60 epochs to avoid over-fitting after which showed a decline in validation accuracy. Specific parameter settings were used with the Adam optimizer, such as setting  $\beta_1$  to 0,8 and  $\beta_2$  to 0,99, while the other parameters remained unaltered. Among the models experimented with, the fine-tuned G-CNN model showcased the best performance achieving a validation accuracy of 95 %. This outcome surpassed the performance of Inception-Resnet v2 and VGG-19. Notably, it was shown that while processing huge amounts of data, lighter models like results with VGG-16 or VGG-19 tended to be better as in figure 4 and figure 5.

Model	Training accuracy	Validation accuracy
G-CNN	99 %	95 %
VGG-16	97,1 %	81 %
Inception V3	97 %	80 %
VGG-19	96 %	81 %
Xception	93 %	71 %
ResNet 110	94 %	73 %

Model	Training accuracy	Validation accuracy
50-50	99 %	95 %
60-40	98 %	94 %
75-25	98,5 %	96 %
80-20	98 %	96,5 %
85-15	97 %	97 %
90-10	99 %	98 %

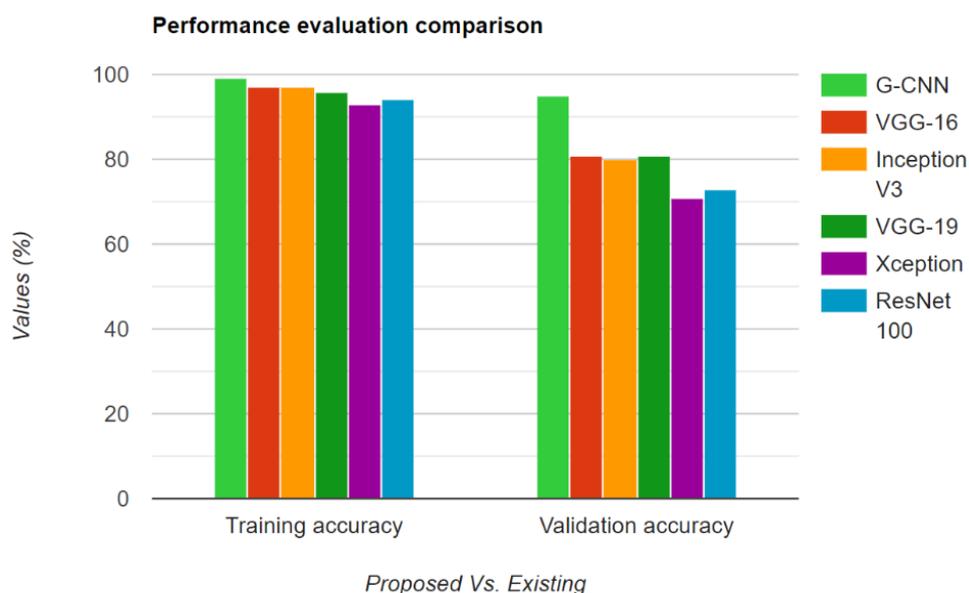


Figure 4. Performance evaluation comparison

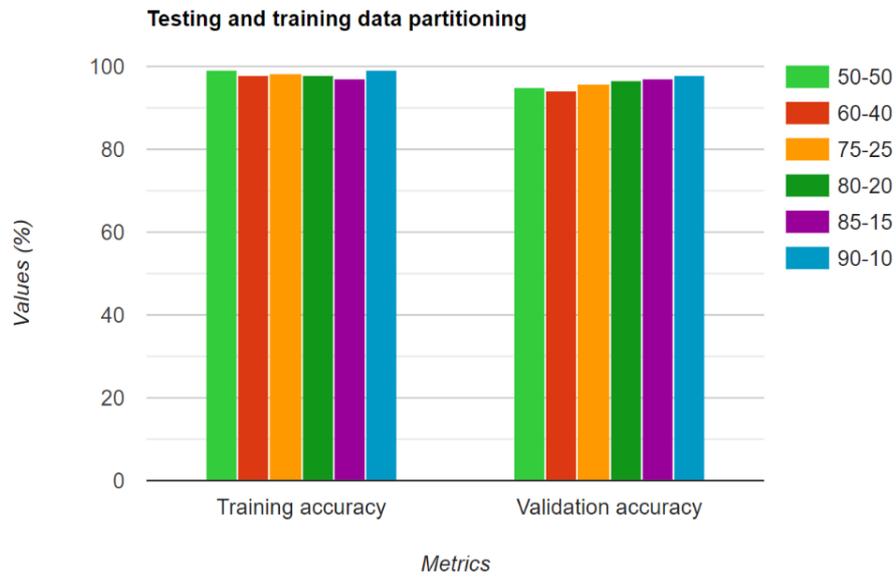


Figure 5. Testing and training data partitioning

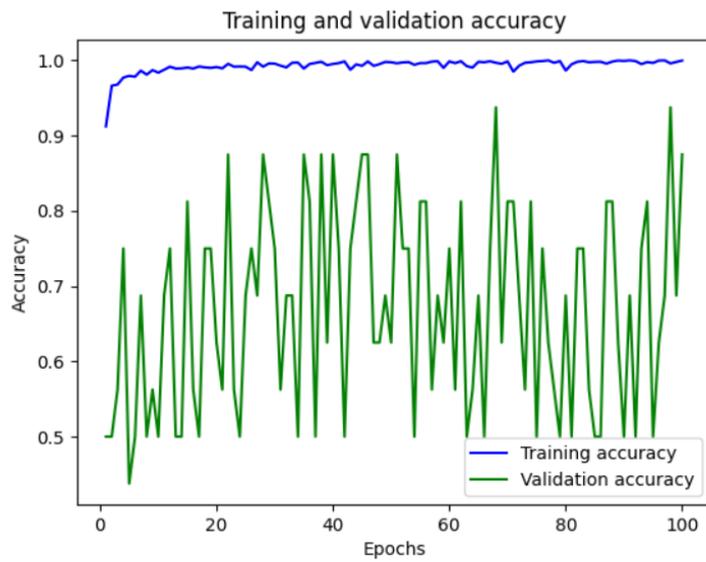


Figure 6. Training and validation accuracy

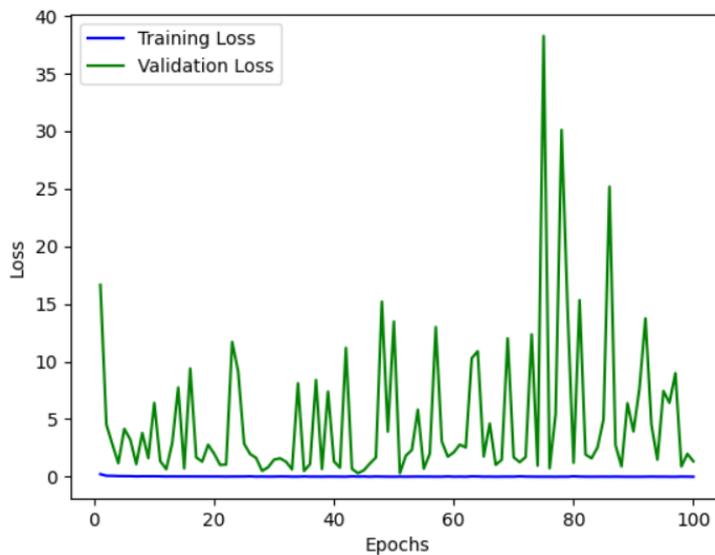
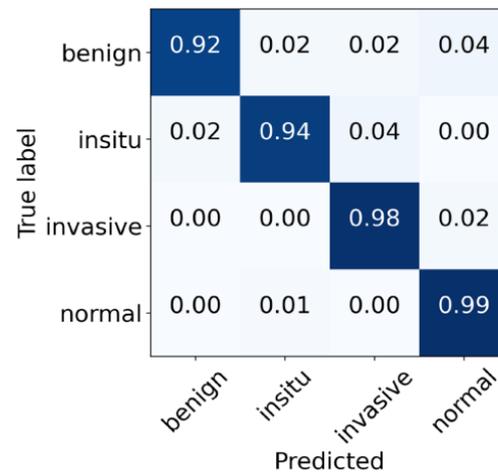


Figure 7. Training and validation loss



**Figure 8.** Confusion matrix

Then, the experimentation revealed that smaller models, particularly VGG-16 outperformed deeper models like Inception-Resnet v2 and VGG-19 when handling substantial amounts of data in the context of the BACH dataset. The fine-tuning strategy involving layer unfreezing up to a certain extent was critical in optimizing the models' performance. Certainly, in the process of finding the optimal solution after fine-tuning the pre-trained model, various train-test splitting combinations were experimented with to strike a balance between training data sufficiency and preventing over-fitting. Over-fitting, a prevalent challenge during the training phase was addressed by introducing the validation dataset idea. After every epoch, the validation set played a crucial role in tracking cross-entropy loss and categorical accuracy. This approach helped in assessing whether the model was improving its accuracy on unseen data or if it was beginning to overfit the training data. Consequently, when the validation accuracy reached a plateau or showed no further improvement, indicating potential over-fitting, the training process was halted at that juncture. Table 2 indicates that the 80/20 train-test split configuration yielded the best results among the splitting combinations tested. As the number of epochs rose, the accuracy of the training and validation processes improved as seen in figure 6. It's clear that the prototype experienced over-fitting, but training was stopped when the validation accuracy plateaued, indicating a diminishing return in accuracy improvement with additional epochs. Moreover, the VGG-16, VGG-19, and Inception-ResNet v2 models' validation loss curves are shown in figure 7 providing insights into their performance in terms of minimizing loss during the training process. These visualizations were instrumental in understanding and controlling over-fitting tendencies while training the models on the dataset, ultimately guiding the decision-making process for terminating training to prevent over-fitting. Figure 8 shows the confusion matrix.

### Classification outcomes

In a binary classification scenario, the model distinguishes between malignant (normal and benign) and benign (invasive and in situ) classifications with an astounding 98,6 % accuracy. Upon closer examination, however, we find that we just utilize the posterior probability to adjust the frequency array and use these probabilities as weighting elements to enhance the input of the final picture classification model. The benign class is less accurate than the normal, in situ, and invasive classes. We significantly beat several cutting-edge models and obtained an outstanding overall accuracy of 97,50 % in all four classes.

The performance metrics for patches inside each of the four classes. Notably, the accuracy of in situ patches is the lowest whereas invasive patches exhibit the best precision. This variance can be attributed to the distinct structural characteristics of invasive images in contrast to the other courses. In situ, photos tend to exhibit localized infections, which in benign and normal circumstances may resemble lobules. Moreover, during the process of converting images into patches, all patches were labelled identically to their respective images. This methodology might make it challenging to accurately identify in situ images, leading to lower precision scores. However, it's important to note that in real scenarios, within in situ imaging, a few normal or benign areas may be seen. Designating these spots as being in situ could potentially misinform the model, impacting the accuracy of patch predictions adversely. In essence, while the precision of in situ patches may appear lower due to the challenges associated with their identification in the patch-level classification model, it's crucial to consider the intricacies involved in accurately labelling patches within the in-situ category, as mislabeling could adversely affect the model's predictive accuracy.

## CONCLUSIONS

Using histological pictures to classify breast cancer tissues is a challenging undertaking. In the present study, we suggested a deep learning (DL)-based technique designed to use histopathological images for breast cancer picture categorization. Since the great resolution of these pictures makes direct processing inside DL-based models' resource-intensive, we first devised a model for classifying image patches. This work derived image-level accuracy using this patch-level data. Rather than using conventional methods, our strategy was to create a new G-CNN. To produce image-level predictions for the four groups of normal, benign, invasive, and in situ, this network mapped the patch-level data. When our suggested method's findings were compared to those of earlier techniques, encouraging results were found, suggesting that our method has the potential to be a reliable diagnostic tool for the categorization of breast cancer using histological pictures. Our suggested paradigm does, however, have certain drawbacks. Notably, training numerous models is necessary to achieve validation accuracy, which adds complexity to the overall model architecture and takes time. We propose that future work should focus on creating an end-to-end deep learning model that can directly analyze high-resolution photos as input and provide optimum outcomes. Moreover, instead of giving all of the patches in an image the same label, to increase patch-level accuracy, a model that is capable of accurately assigning patch-level labels may be developed. Additionally, investigating other histopathology imaging datasets can increase the model's resilience and breadth of use.

## BIBLIOGRAPHIC REFERENCES

1. Sung, H. et al. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209-249 (2021).
2. Elfgren, C. et al. Comparative analysis of confocal microscopy on fresh breast core needle biopsies and conventional histology. *Diagnost. Pathol.* 14, 1-8 (2019).
3. Hameed, Z., Zahia, S., Garcia-Zapirain, B., Javier Aguirre, J. & María Vanegas, A. Breast cancer histopathology image classification using an ensemble of deep learning models. *Sensors* 20, 4373 (2020).
4. Aresta, G. et al. Bach: Grand challenge on breast cancer histology images. *Med. Image Anal.* 56, 122-139 (2019)
5. 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, e0214587 (2019)
6. Elmannai, H., Hamdi, M. & AlGarni, A. Deep learning models combining for breast cancer histopathology image classification. *Int. J. Comput. Intell. Syst.* 14, 1003-1013 (2021)
7. Sharma, S. & Kumar, S. Te xception model: A potential feature extractor in breast cancer histology images classification. *ICT Express* 8, 101-108 (2022).
8. Bianconi, F., Kather, J. N. & Reyes-Aldasoro, C. C. Experimental assessment of color deconvolution and color normalization for automated classification of histology images stained with hematoxylin and eosin. *Cancers* 12, 3337 (2020).
9. Salvi, M., Acharya, U. R., Molinari, F. & Meiburger, K. M. Te impact of pre-and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. *Comput. Biol. Med.* 128, 104129 (2021).
10. Zhu, C. et al. Breast cancer histopathology image classification through assembling multiple compact cnns. *BMC Med. Inform. Decis. Mak.* 19, 1-17 (2019).
11. Szegedy, C., Iofe, S., Vanhoucke, V. & Alemi, A. A. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Tirty-First AAAI Conference on Artificial Intelligence* (2017).
12. Hao, Y. et al. Breast cancer histopathological images classification based on deep semantic features and gray level co-occurrence matrix. *PLoS ONE* 17, e0267955 (2022).
13. Turgut S, Dağtekin M, Ensari T. Microarray breast cancer data classification using machine learning methods. In: 2018 electric electronics, computer science, biomedical engineerings' meeting (EBBT). IEEE; 2018. p. 1-3.

14. Magna AAR, Allende-Cid H, Taramasco C, Becerra C, Figueroa RL. Application of machine learning and word embeddings in the classification of cancer diagnosis using patient anamnesis. *IEEE Access*. 2020;8:106198-213

15. Kayikci S, Khoshgoftaar T. A stack based multimodal machine learning model for breast cancer diagnosis. In: 2022 international congress on human-computer interaction, optimization and robotic applications (HORA). IEEE; 2022. p. 1-5.

16. Kayikci S. A deep learning method for passing completely automated public turing test. In: 2018 3rd international conference on computer science and engineering (UBMK). IEEE; 2018. p. 41-44

17. Abbasniya MR, Sheikholeslamzadeh SA, Nasiri H, Emami S (2022) Classification of breast tumors based on histopathology images using deep features and ensemble of gradient boosting methods. *Comput Electr Eng* 103:108382

18. Barsha NA, Rahman A, Mahdy M (2021) Automated detection and grading of invasive ductal carcinoma breast cancer using ensemble of deep learning models. *Comput Biol Med* 139:104931.

19. Chen S, Wu J, Liu X (2021) EMORL: effective multi-objective reinforcement learning method for hyperparameter optimization. *Eng Appl Artif Intell* 104:104315.

20. Choudhary T, Mishra V, Goswami A, Sarangapani J (2021) A transfer learning with structured filter pruning approach for improved breast cancer classification on point-of-care devices. *Comput Biol Med* 134:104432.

21. Ezhilraman SV, Srinivasan S, Suseendran G (2019) Breast cancer detection using gradient boost ensemble decision tree classifier. *Int J Eng Adv Technol* 9(2):2169-2173

22. Fondón I, Sarmiento A, García AI, Silvestre M, Eloy C, Polónia A, Aguiar P (2018) Automatic classification of tissue malignancy for breast carcinoma diagnosis. *Comput Biol Med* 96:41-51.

23. Goel T, Murugan R, Mirjalili S, Chakrabartty DK (2022) Multicovid-net: multi-objective optimized network for covid-19 diagnosis from chest x-ray images. *Appl Soft Comput* 115:108250.

24. Liew XY, Hameed N, Clos J (2021) An investigation of XGBoostbased algorithm for breast cancer classification. *Mach Learn Appl* 6:100154

25. Mostafa SS, Mendonca F, Ravelo-Garcia AG, Juliá-Serdá GG, Morgado-Dias F (2020) Multi-objective hyperparameter optimization of convolutional neural network for obstructive sleep apnea detection. *IEEE Access* 8:129586-12959

26. Ozaki Y, Tanigaki Y, Watanabe S, Nomura M, Onishi M (2022) Multiobjective tree-structured Parzen estimator. *J Artif Intell Res* 73:1209-1250

27. Spanhol FA, Oliveira LS, Petitjean C, Heutte L (2015) A dataset for breast cancer histopathological image classification. *IEEE Trans Biomed Eng* 63:1455-1462

28. Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S.S., Safwan, M., Alex, V., Marami, B., Prastawa, M., Chan, M., Donovan, M., et al.: BACH: Grand challenge on breast cancer histology images. *Medical Image Analysis* 56, 122-139 (2019)

29. He, L., Long, L.R., Antani, S., Thoma, G.R.: Histology image analysis for carcinoma detection and grading. *Computer Methods and Programs in Biomedicine* 107(3), 538-556 (2012)

30. Holzinger, A., Goebel, R., Mengel, M., Müller, H.: *Artificial Intelligence and Machine Learning for Digital Pathology: State-of-the-art and Future Challenges*, vol. 12090. Springer Nature (2020)

31. Aksac, A., Demetrick, D.J., Ozyer, T., Alhajj, R.: BreCaHAD: a dataset for breast cancer histopathological annotation and diagnosis. *BMC Research Notes* 12(1), 1-3 (2021)

32. Janowczyk, A., Madabhushi, A.: Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics* 7 (2016)
33. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. *Medical Image Analysis* 42, 60-88 (2017)
34. Steiner, D.F., MacDonald, R., Liu, Y., Truszkowski, P., Hipp, J.D., Gammage, C., Thng, F., Peng, L., Stumpe, M.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 (2018)
35. Chugh, G., Kumar, S., Singh, N.: Survey on machine learning and deep learning applications in breast cancer diagnosis. *Cognitive Computation* pp. 1-20 (2021)
36. Tosta, T.A.A., de Faria, P.R., Neves, L.A., do Nascimento, M.Z.: Computational normalization of h&e-stained histological images: Progress, challenges and future potential. *Artificial Intelligence in Medicine* 95, 118-132 (2019)
37. Roy, S., Lal, S., Kini, J.R.: Novel color normalization method for Hematoxylin & Eosin stained histopathology images. *IEEE Access* 7, 28982-28998 (2019)
38. Tosta, T.A.A., de Faria, P.R., Neves, L.A., do Nascimento, M.Z.: Color normalization of faded H&E-stained histological images using spectral matching. *Computers in Biology and Medicine* 111, 103344 (2019)
39. Bukenya, F.: A hybrid approach for stain normalisation in digital histopathological images. *Multimedia Tools and Applications* 79(3), 2339-2362 (2020)
40. Zanjani, F.G., Zinger, S., Bejnordi, B.E., van der Laak, J.A., de With, P.H.: Stain normalization of histopathology images using generative adversarial networks. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 573-577. IEEE (2018)

#### **FINANCING**

No financing.

#### **CONFLICT OF INTEREST**

None.

#### **AUTHORSHIP CONTRIBUTION**

*Conceptualization:* N Hari Babu, Enireddy Vamsidhar.

*Data curation:* N Hari Babu, Enireddy Vamsidhar.

*Formal analysis:* N Hari Babu, Enireddy Vamsidhar.

*Drafting - original draft:* N Hari Babu, Enireddy Vamsidhar.

*Writing - proofreading and editing:* N Hari Babu, Enireddy Vamsidhar.