SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I5P113

Volume 12 Issue 5, 149-161, May 2025

Original Article

Localized And Global Feature Integration for Efficient Bladder Cancer Detection Using Deep Learning

R. Reena[1], S. Amala Shanthi[2]

[1,2]*Electronics and Communication Engineering, Noorul Islam Centre for Higher Education, Kumaracoil, Kanyakumari, Tamil Nadu, India.*

[1]*Corresponding Author : reena.r.321@outlook.com*

Received: 08 March 2025 Revised: 12 April 2025 Accepted: 12 May 2025 Published: 27 May 2025

Abstract *- Bladder Cancer (BC) is the prevalent urinary system cancer with abnormal cell growth in the bladder lining that, if not detected at an early stage, may cause life-threatening complications. Bladder cancer detection includes identifying and categorizing cancer and non-cancer cases through histopathological imaging methods. Making an accurate diagnosis is essential to initiate therapy on time and enhance patient outcomes. Challenges like the heterogeneity of tumor appearance, overlapping imaging features with benign conditions and imaging artifacts make the diagnosis challenging. Conventional approaches are based on invasive techniques like cystoscopy and manual image analysis, which are time-consuming and unpredictable. The present study focuses on developing a hybrid Deep Learning (DL) model that integrates local and global feature extraction methods for effective bladder cancer detection. The dataset utilized consists of 3045 histopathological images, 1435 of which are classified as healthy and 1610 as Urothelial Cell Carcinoma (UCC). Preprocessing techniques, including normalization and augmentation, are applied to enhance data quality and variability. The hybrid model utilizes Convolutional Neural Networks (CNN) for local feature representation and EfficientNetB0 for global feature representation, combining their outputs using dense layers in order to classify as "healthy" or "UCC". The suggested study makes a high-impact performance with 98.68% accuracy, 98.75% precision, 98.75% recall and an F1-score of 98.75%, outperforming all current approaches to bladder cancer detection. These outcomes denote the success of the hybrid model in overcoming bladder cancer diagnosis challenges, providing a robust and reliable solution to medical practice.*

Keywords *- Bladder cancer, Convolutional Neural Network, Local features, Global features, EfficientNetB0, Deep learning.*

1. Introduction

Bladder cancer, one of the most prevalent malignancies, originates in the urinary tract, characterized by the proliferation of atypical cells in the bladder lining, as illustrated in Figure 1. Bladder cancer is a major health issue globally owing to its high incidence, recurrence and difficulty in diagnosis. Early prediction of bladder cancer is important to improve patient outcomes and lower mortality rates [1]. However, accurate diagnosis is still a challenging process, mainly because of the heterogeneity of cancer presentations, differences in tumor grades and the occurrence of other benign bladder lesions. Recent global health statistics report indicates that BC is the most prevalent cancer worldwide, with tenth position, having greater than 570,000 new cases and as stated by the World Health Organization (WHO), there are more than 200,000 fatalities per year. In the US alone, approximately 17,000 deaths and 82,000 new cases were estimated in 2024, with a rising incidence in aging populations and industrialized regions. The disease predominantly affects men, about four times more frequently than women and is more common among individuals over the age of 55. Smoking remains the leading risk factor, accountable for nearly half of all cases, followed by occupational exposure to carcinogenic chemicals (like aromatic amines in dyes and rubber), chronic bladder infections and prolonged use of certain medications. In addition to its physical cost, bladder cancer places a heavy social and psychological burden on patients, frequently necessitating lifelong surveillance and therapy. Repeated interventions, fear of recurrence, cost, and emotional distress can significantly compromise quality of life, further underscoring the importance of timely and accurate diagnosis.

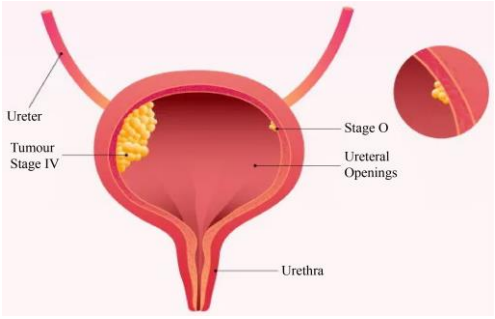

Fig. 1 Visualization of bladder cancer

Bladder cancers are mainly classified into two types: Muscle-Invasive (MI) and Non-Muscle-Invasive (NMI). NMI constitutes the initial phase of the disease, wherein the tumor is limited to the bladder lining, whereas MI refers to deeper penetration into the muscular layers with an increased risk of metastasis. Proper staging and grading of bladder cancer are of utmost importance for the decision-making process regarding appropriate treatment. Traditionally, cystoscopy followed by histopathological examination is the gold standard for assessment [2]. Nonetheless, cystoscopy is invasive, expensive and involves specialized interpretation. In addition, differences in image quality and human skills may lead to diagnostic variability. The heterogeneity of bladder tumours and variability in tumor grades and overlapping features with benign diseases make diagnosis challenging. Imaging studies such as MRI, CT scanning and cystoscopy imaging are of significance in detecting abnormality in the bladder. They are, nonetheless, limited as their ability to diagnose largely falls on the capabilities of radiologists or pathologists [3].

Recent progress in DL has revolutionized the medical image analysis area by enabling the development of precise, robust, non-invasive diagnostic systems on a fully automated basis. Hybrid DL models that leverage the strengths of diverse architectures have been identified as a promising solution to improve diagnostic performance. Although each imaging modality is unique, combining multiple features can enhance diagnostic accuracy. This research introduces a novel hybrid DL approach that integrates EfficientNetB0 and CNN to detect bladder cancer. EfficientNet is proficient in extracting global features, and CNN guarantees that local features are adequately extracted. The main contributions presented by the suggested system are as follows.

- To develop a hybrid DL approach that integrates local and global feature extraction techniques for detecting and classifying bladder cancer from medical imaging data.
- To examine the efficiency of the suggested hybrid model by utilizing the evaluation metrics.
- Comparing the efficiency of the developed hybrid model with existing techniques demonstrating the accuracy and reliability in BC detection.

The remainder of this paper is arranged as follows: Section 2 offers a review of related research focused on bladder cancer detection applications. Section 3 details the suggested methodology, emphasizing the integration of CNN and EfficientNet B0 architectures. In Section 4, the experimental outcomes and performance evaluation of the proposed hybrid framework are presented. Section 5 is concluded by briefing the key findings.

2. Literature Review

Yoo et al. (2023) [4] investigated the diagnostic efficiency of an Artificial Intelligence (AI) approach for bladder cancer detection and tumor grade prediction using 10,991 cystoscopy images. The study utilized a mask region-based CNN with a ResNeXt-101-32 × 8 d-FPN backbone, achieving 95.0% sensitivity, 93.7% specificity, 94.1% diagnostic accuracy and 74.7% DSC values, respectively. The authors noted that the limited number of initially trained images impacts the model performance. Sarkar et al. (2023) [5] proposed a hybrid model that applied statistical Machine Learning (ML) and pre-trained Deep Neural Networks (DNN) to enhance the clinical staging of BC using grayscale CT scans. They analyzed three classification tasks: MIBC from NMIBC, identifying normal tissue vs. BC tissue and post-treatment changes from muscle-invasive ones. The ResNet-50 model showed the best F1 scores, reflecting better recall and precision in all experiments. In spite of the encouraging diagnostic potential and enhanced clinical decision-making, the method was constrained by its invasiveness and high cost, with complications like bladder perforation.

Barrios et al. (2022) [6] developed Bladder4Net, a DL pipeline for classifying bladder cancer histopathology images as low-risk and high-risk. It consisted of 378 slides and 182 whole-slide images using four CNN-based classifiers to tackle classification issues. The model attained a 0.91-weighted average F1 value from the New Hampshire BC Study (NHBCS) database. However, the research was hindered by insufficient muscle-invasive cases, which stopped further stratification within these subtypes. Du et al. (2021) [7] directed a study to analyze the performance of DL approaches in identifying bladder cancer from cystoscopy images. They used a dataset of 1,002 images of routine bladder tissue and 734 images of bladder tumours from 175 patients and trained CNN with Easy DL and Caffe platform. The framework of Easy DL achieved an accuracy of 96.9%, which is far better than the accuracy of 82.9% of the model based on Caffe. While the study showed the potential of mobile-based diagnostics in clinical settings, it did not elaborate on challenges like scalability or dealing with varied datasets.

Ikeda et al. (2021) [8] studied Transfer Learning (TL) utilizing general and gastroscopic images to improve bladder tumor detection in cystoscopy imaging. The study utilized a CNN trained sequentially with 8728 gastroscopic images and 2102 cystoscopy images of normal and tumor tissues. The framework attained 97.6% specificity and 95.4% sensitivity surpassing other models and performing comparably to expert urologists, particularly for tumors occupying more than 10% of the image. Despite its success, the study was limited by the scarcity of bladder tumor images available for training. Hammouda et al. (2021) [9] created a Multiparametric Computer-Aided Diagnostic (MP-CAD) to improve the accuracy of Bladder Cancer (BC), particularly distinguishing between T1 and T2 stages, using MRI data. The study employed a fully connected CNN for segmentation, followed by feature extraction from functional, texture and morphological data across nested iso-surfaces. These were utilized to train a dataset of 42 cases relying upon a Neural Network (NN) classifier. The framework achieved 95.24%

accuracy with 0.9864 AUC, outperforming other classifiers. However, the study acknowledged its limitation of using a relatively small dataset, affecting the results' generalizability.

Yang et al. (2021) [10] explored the potential of DL frameworks to accurately identify bladder cancer from cystoscopy images. They analyzed 1,150 non-cancer images from 221 affected ones and 1,200 cancer images from 224 affected ones and, using three CNNs and the EasyDL platform, compared their diagnostic efficiency to that of urology experts. The EasyDL platform achieved the maximum accuracy at 96.9%, outperforming Google Net's 92.54%, and demonstrated diagnostic capabilities comparable to those of clinical experts. However, the study faced limitations due to a relatively small image dataset and the compression of neural network weights, which slightly impacted model accuracy. Lorencin et al. (2021) [11] aimed to enhance the accuracy of optical biopsy, a less invasive but lower-accuracy diagnostic method for urinary bladder cancer, by applying DL techniques. The study utilized Deep Convolutional Generative Adversarial Networks (DCGAN) to augment datasets with synthetic images, addressing the challenge of limited patient data. Augmented datasets were used to train CNN-based architectures, including AlexNet and VGG-16, with AlexNet demonstrating significant performance improvements and reduced sensitivity to hyperparameter changes. However, the study highlighted that traditional augmentation methods offered limited insights and data points.

García et al. (2021) [12] conducted research to develop a self-learning framework for grading the severity of MIBC from histological images stained through immunohistochemistry techniques. They employed Deep Convolutional Embedded Attention Clustering (DCEAC), a two-step, fully unsupervised methodology, integrating a convolutional attention block to refine feature extraction and classify samples into non-tumor, mild and infiltrative patterns. The model attained 90.34% average accuracy, outperforming existing clustering-based approaches and demonstrating its ability to autonomously identify clinically relevant patterns without annotated data. Ikeda et al. (2020) [13] conducted research to boost the accuracy of bladder cancer diagnosis by supporting cystoscopy image assessment by AI. They used a dataset of 2102 cystoscopy images and trained a CNN as a tumor classifier, and assessed its performance with 0.98 area under the ROC curve (AUROC), 94.0% specificity and 89.7% sensitivity. The research proved that AI-based assessment accurately classifies the images with tumor lesions. A weakness of this research was the accuracy of the annotation of lesions within the images, which was a function of the quality of learned data.

Zhang et al. (2020) [14] carried out a study to create a CT-based radiomics system for forecasting the grade of bladder cancer based on data from 145 patients who were subjected to CT urography from October 2014 to September 2017. They utilized feature extraction methodologies, removed collinearity, and applied logistic regression to develop the model and assess its performance using ROC curves and diagnostic statistics. The model had 0.860 in the validation group and 0.950 AUC in the training group, with 83.8% diagnostic accuracy, 88.5% sensitivity and 72.7% specificity, respectively, in the validation group. Even with such results, the research was limited by its sample size and choice of bias since it was retrospective in nature. Yin et al. (2020) [15] conducted research to tackle the daunting task of histologically classifying between superficially invasive T1 and non-invasive Ta stages of bladder cancer, which have substantially varied disease progression risks. From a dataset of eosin-stained and 1177 hematoxylin images of bladder tumor tissue, the authors built automatic pipelines to extract close to 700 features from domain knowledge. They examined them with six supervised ML approaches. The research attained a level of 84% accuracy, surpassing convolutional neural networks with features of desmoplastic reaction and tumor cell nuclei identified as major predictors.

Hammouda et al. (2020) [16] proposed a 3D DL-based CNN architecture to segment tumours and pathological bladder walls from T2-weighted MRI data. Their method involved data normalization preprocessing, the use of two networks for segmentation and the soft output of the network is enhanced with a fully connected Conditional Random Field (CRF), utilizing an Adaptive Shape Prior (ASP) model in the second network for better segmentation. The approach registered good performances, confirmed using indices like Hausdorff distance and Dice similarity coefficient, and performed better in precision than currently practised techniques. However, the study acknowledged limitations due to bladder size, shape and pathology variations across patients. Jansen et al. (2020) [17] aimed to enhance reproducibility in grading non-muscle-invasive urothelial cell carcinoma by suggesting a fully automated deep learning-based detection and grading network. Using a dataset of 328 transurethral resection specimens, a U-Net-based segmentation network was employed to identify urothelium, which served as input for a classification network grading tumours per the 2004 WHO system. The automated system achieved better performance with a consensus of three specialized pathologists and properly graded 71% of high-grade cancers and 76% of low-grade. Shkolyar et al. (2019) [18] created a DL algorithm to enhance the cystoscopic detection of bladder cancer. The study used white light video frames with additional validation performed on 54 patients. The convolutional neural network-based CystoNet achieved 90.9% sensitivity and 98.6% specificity in the validation dataset. Despite its success, the study noted the lack of sub-classification of benign and malignant lesions. Despite the significant advancements in ML and DL for detecting and classifying bladder cancer, several research gaps remain. While techniques like TL and CNNs demonstrated promising results, the role of hybrid models combining local and global features remains underexplored, limiting the potential for accurate and robust predictions. Additionally, many existing studies depend

on relatively small and homogenous datasets, raising concerns about the generalizability and scalability of their findings. Furthermore, challenges such as class imbalance, variability in tumor appearance and the lack of standardized evaluation metrics across studies hinder the development of universally applicable models. Despite the emergence of data augmentation methods like GANs to address data scarcity, their application in bladder cancer detection is still in its infancy. Therefore, there is a pressing need for more comprehensive research that integrates hybrid deep learning models utilizing both local and global features. Such studies focus on addressing the limitations of dataset diversity, incorporating AI approaches and developing standardized evaluation protocols to ensure clinical relevance and widespread applicability. Table 1 summarizes the recent studies.

Table 1. Summary of existing studies

Author (Year)	Data Type	Method	Remarks
Yoo et al. (2023)	Cystoscopy images (10,991)	Mask R-CNN (ResNeXt-101-FPN)	High accuracy; performance limited by initial training data.
Sarkar et al. (2023)	Grayscale CT scans	Hybrid DL + ML (ResNet-50)	Strong F1 scores; high cost and invasive with risk of complications.
Barrios et al. (2022)	Histopathology slides (NHBCS)	Bladder4Net (4 CNNs)	Good risk classification; limited MIBC samples.
Du et al. (2021)	Cystoscopy images (1,736)	EasyDL vs Caffe CNN	EasyDL outperformed; lacked scalability evaluation.
Ikeda et al. (2021)	Cystoscopy + gastroscopy images	Transfer learning CNN	High sensitivity/specificity; limited bladder tumor data.
Hammouda et al. (2021)	MRI data (42 cases)	MP-CAD with CNN + NN	Accurate T1 vs T2 staging; small dataset.
Yang et al. (2021)	Cystoscopy images	CNNs + EasyDL	EasyDL matched expert performance; small dataset, weight compression issues.
Lorencin et al. (2021)	Optical biopsy data	DCGAN + CNNs (AlexNet, VGG-16)	Synthetic data improved performance; limited benefit from traditional augmentation.
García et al. (2021)	IHC-stained histology images	DCEAC (unsupervised)	No labeled data is needed; effective MIBC pattern classification.
Ikeda et al. (2020)	Cystoscopy images (2,102)	CNN classifier	High AUROC; accuracy limited by annotation quality.
Zhang et al. (2020)	CT urography (145 patients)	Radiomics + Logistic Regression	Strong AUC; retrospective design and small sample.
Yin et al. (2020)	H&E-stained tumor images (1,177)	Feature-based ML (6 models)	Outperformed CNNs; domain features like desmoplasia were key.
Hammouda et al. (2020)	T2-weighted MRI	3D CNN + CRF + ASP	Good segmentation metrics; variability in anatomy affected performance.
Jansen et al. (2020)	TUR biopsy specimens (328)	U-Net + classification network	Improved grading reproducibility: 71% HG and 76% LG accuracy.
Shkolyar et al. (2019)	White light cystoscopy video frames	CystoNet (CNN)	High sensitivity/specificity; lacked benign vs malignant lesion classification.

3. Materials and Methods

The detection of bladder cancer using hybrid DL integrates EfficientNetB0 and CNN models to analyze medical images by combining global and local feature extraction. Figure 2 illustrates the workflow, which starts with resizing an input image dataset to 224x224 pixels using RGB channels for consistency. Preprocessing, including normalization, resizing and augmentation, improves data quality and variability. EfficientNetB0, pre-trained on ImageNet, extracts high-level global features, while CNN layers capture finer local details like edges and patterns. These features are fused through concatenation to create a unified representation, processed by dense layers for feature transformation and dimensionality reduction. A dropout layer prevents overfitting, and the refined features are classified using a sigmoid activation function, which categorizes images as UCC or healthy.

3.1. Dataset Description

The samples for this study were derived from 90 hematoxylin-and-eosin-stained histopathology slides of urinary bladder lesions, with images categorized into two primary classes: healthy and UCC, as presented in Figure 3. The dataset focuses on patch-level image annotations, providing more granular detail compared to slide-level annotations. A total of 3045 images were collected, with 1435 images labelled healthy and 1610 as UCC [19]. The images were systematically obtained by taking non-overlapping images of every tissue available in the areas of each slide. These images are then classified manually by a pathologist, ensuring high-quality annotations for training deep-learning models aimed at detecting bladder cancer. The dataset offers both numerical and categorical features, with the numerical aspect comprising pixel values and the categorical aspect distinguishing between UCC and healthy tissue types.

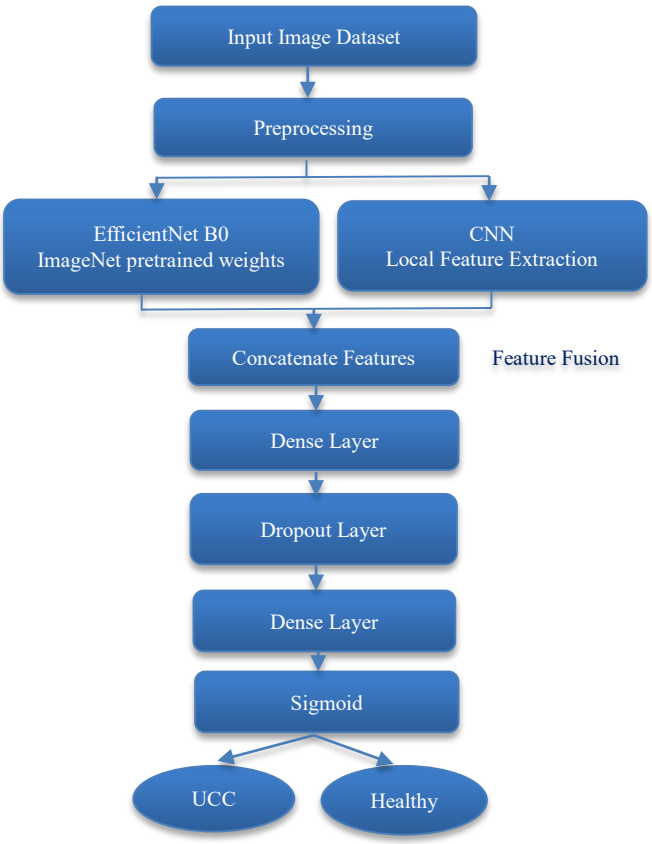

Fig. 2 Proposed model block diagram

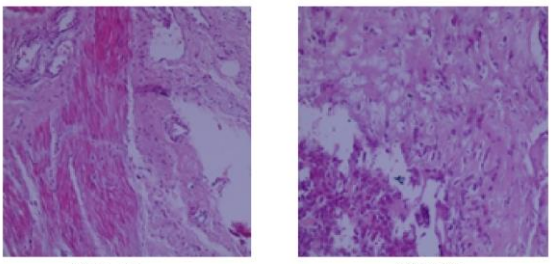

Fig. 3 Sample images from the dataset

3.2. Data Preprocessing and Augmentation

Preprocessing normalizes the input by resizing the images to the same size and changing the color format from BGR to RGB. Convergence of the model is enhanced through normalization of image pixel values through scaling and dividing the pixel intensities to the range 0 to 1. The data is split post-preprocessing in a ratio of 75:15:10 for training, validation, and testing, respectively. Data augmentation also improves the model's capability to generalize by adding variability to the training set. Methods like random rotations of up to 30 degrees, shifts, zooms, brightness changes and horizontal flips are used to expose the model to various image variations.

3.3. Feature Extraction Phase

The procedure of identifying and removing relevant features and patterns from raw data, specifically images, to make the input easier for an ML model is referred to as feature extraction. It assists in reducing the complexity of the data without losing crucial information for classification or prediction tasks. In this study, feature extraction was done by combining CNN layers and the EfficientNetB0 model, which assists in capturing local and global features in the images. The CNN layers emphasize the fine-grained information such as textures and edges, whereas EfficientNetB0 offers a higher-level insight into larger patterns and structures. Both features are concatenated to enhance classification accuracy. CNN is a DL model commonly employed in image and video processing, replicating weights and biases between layers. Such a design simplifies complexity and improves generalization for computer vision tasks. CNN includes fully connected layers for classification, pooling layers for down-sampling feature maps and convolutional layers for feature extraction and combining low-level features into high-level representations [20]. Integrating feature extraction and classification, a CNN efficiently processes data for accurate predictions, as illustrated in Figure 4.

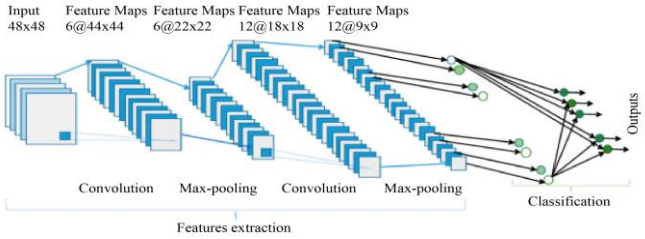

Fig. 4 Basic architecture of CNN

In a CNN, feature extraction occurs within the convolution and pooling layers, while the classification layer handles prediction tasks. The input image is initially processed through convolution operations using shared weights and multiple trained kernels. The architecture determines the feature maps' count in a convolutional layer, and a deep CNN typically comprises multiple stacked convolutional layers. These layers apply various filters to the raw image data, extracting critical features essential for classification. The

value of a neuron or node at a specific position within the feature map of a layer is represented in Equation (1) as follows.

$$v_{pq}^{ij} = g(b_{pq} + \sum_m \sum_{x=0}^{X_i} \sum_{y=0}^{Y_i} w_{pqm}^{xy} v_{(p-1)m}^{(i+x)(j+y)}) \quad (1)$$

In this case, m represents the index of the feature map in the layer connected to the current feature map. The term w denotes the weight corresponding to a specific position linked to the m^{th} feature map, while height is denoted by h and width of the spatial convolutional kernel is signified by w. Additionally, b represents the bias associated with the feature map in the layer represented in Equation (2).

$$F(i) = max(0, i) \quad (2)$$

Usually, a nonlinear activation function is done after the convolution layer. While functions like sigmoid or tanh are for this purpose, the ReLU is preferred as it significantly enhances training speed, making the network more efficient. Pooling layers introduce invariance by reducing the resolution of feature maps. Each pooling layer is associated with the preceding convolutional layer, which forms the feature extraction section of the network. The feature maps are transferred to the Fully Connected (FC) layers after being flattened into a one-dimensional vector. The final classification is handled by these layers using the features that the convolutional and pooling layers have extracted. Further hidden layers and output layers that predict the class are included in the FC layer. It initially obtains the flattened vector as input and processes it through the hidden layers, which can be mathematically expressed in Equation (3).

$$h_p(i) = w_p \cdot i + b_p \quad (3)$$

Here, w represents the weight and b denotes the bias. Equation (4) provides the mathematical expression for the activation function employed for the hidden layers output.

$$a_p = activation\ (h_p(i)) \quad (4)$$

The FC layer output can be further processed through additional hidden layers if required. In a binary classification task, the output layer usually includes one neuron with a sigmoid function, as shown in Equation (5). The training phase of the model employs this output to compute the training loss. There are various optimization techniques used to adjust the weight of the network during training in an attempt to improve the model's generalization.

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (5)$$

Local feature extraction with CNNs is a sequence of layers created to recognize and extract localized patterns from images. Convolutional layers first utilize tiny filters to convolve over the image by sliding over it to identify simple patterns such as edges, corners and texture. These filters are used as the basis for low-level feature extraction, which has a fundamental role in studying the image's structure. A Non-Linear Activation Function (ReLU) is placed after every convolutional layer in the network. This function allows the network to learn complex relationships and enables the model to capture complex patterns. Following the convolutional layers, max-pooling layers are used on the feature maps to downsample the spatial dimensions, taking the maximum value from small regions, thus preserving the most important features while minimizing computational complexity. This process not only lessens the dimensions of the feature maps but also avoids overfitting, improving the effectiveness and robustness of the approach. The output from these layers is flattened into a one-dimensional vector, ready to process the extracted features in fully connected layers. The model makes better predictions by feeding the flattened vector into a dense layer of 256 units, which is responsible for learning complex, high-level abstractions by assigning weights and biases to the features.

EfficientNet introduces a compound scaling method to efficiently scale ConvNets while optimizing resource usage. The EfficientNet model ranges from B0 to B7, with each variant offering increased parameters and accuracy [21]. EfficientNetB0, a pre-trained model, was employed to obtain comprehensive, high-level features that encapsulate the global characteristics of an input image. Figure 5 illustrates EfficientNetB0, which processes a 224x224x3 input image through a 3x3 convolution layer using batch normalization and Swish activation. Its core component, the Mobile Inverted Bottleneck Convolution (MBConv) block, utilizes depth-wise separable convolutions and an expansion step to reduce parameters and computational complexity. Integrated Squeeze-and-Excitation (SE) blocks adjust the feature maps to focus on key features, enhancing representational capacity. Following several MBConv layers, Global Average Pooling (GAP) aggregates spatial information, and the output passes through a dense layer and softmax for final classification. Equations (6) to (10) illustrate the scaling of width, depth and resolution.

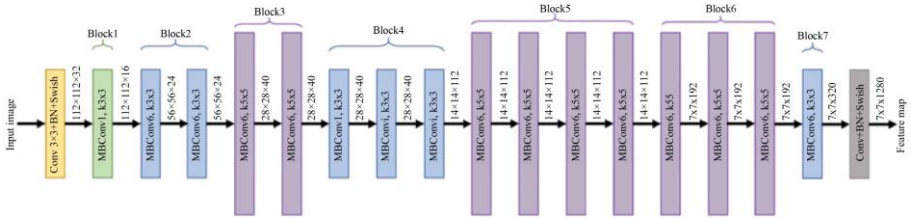

Fig. 5 Basic architecture of EfficientNet B0

$$x = \beta^{\emptyset} \quad (6)$$

$$h = \alpha^{\emptyset} \quad (7)$$

$$s = \gamma^{\emptyset} \quad (8)$$

Shows that $$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \quad (9)$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1 \quad (10)$$

Where the width is denoted as x, the height is represented by h and the resolution of the model is represented as s, where $\alpha, \beta and\ \gamma$ were constant coefficients realized through a small grid search on the original small model. Efficient Net utilizes a compound scaling approach that modifies the depth, width and resolution of the system based on a scaling factor $\emptyset$.

The model first identifies simple attributes, such as edges and textures and then continues towards more intricate features, such as shapes and objects, that are necessary in order to make sense of the image. Following feature extraction, GAP is employed in the model to take the average of each and every feature map along the spatial dimensions to generate a single representative value per feature map. This processing is a dimensionality reduction of the feature maps, giving a sparse and computationally inexpensive representation of the important features. The outcome is a set of high-level features that capture the general content in the image, such as shapes, patterns and structures, which are important for numerous tasks like image classification, object detection, or recognition. EfficientNetB0's feature extraction enables a very useful way of extracting global information from images while keeping the model efficient in handling large data but retaining high levels of accuracy when detecting complex patterns. The concatenation layer combines the features extracted by the convolutional layers with those obtained from EfficientNetB0. The purpose of this concatenation is to increase the framework's ability to understand the image by utilizing both localized and global information. The CNN features focus on detecting fine-grained details such as textures, edges and small regions within the image, which are crucial for understanding intricate patterns. On the other hand, the EfficientNetB0 features capture a broader, more holistic view of the image, emphasizing larger patterns and complex structures.

3.4. Classification Phase

Classification utilizes the extracted features from the input data to predict a label or category for the input. The features are examined after feature extraction using FC layers to learn the relationships and carry out the ultimate classification, predicting whether the image is a healthy or UCC case. The FC layers in the model are essential in distinguishing the features obtained from earlier layers of convolutional and EfficientNetB0. The concatenated features are used as input to a dense layer with ReLU and 256 neurons, learning intricate relationships between global and local features. To decrease overfitting and improve the model's generalization abilities, the dropout layer at a 20% rate is presented after the initial FC layer, randomly disabling 20% of the neurons in every pass through the training. The second FC layer, containing 128 neurons, is employed to further represent the features by tuning and optimizing the learned feature relations. Lastly, the output layer consists of one neuron with a sigmoid, perfect for binary classification problems.

The sigmoid function gets an output value of 0 or 1, which is the probability that the input image belongs to "healthy" or "UCC" classes. This structure guarantees that the model makes precise and consistent predictions by utilizing the features extracted and learned from the input image. Figure 6 displays the proposed study architecture. The algorithm for the suggested system is given below.

Algorithm: Bladder Cancer Detection Using CNN-EffecientNetB0 Model
Input: Histopathological images of bladder
Output: Bladder cancer detection model (healthy or UCC)
Begin:
Load and preprocess data:

1. Collect dataset: $C = \{(L_i, M_i)\}_{i=0}^{N-1}$, where L_i is a histopathological and $M_i \in \{0, 1\}$ (1: UCC, 0: healthy).
2. Preprocess:
   - Resize: $L_i \rightarrow L_i' \in R^{224\times224}$
   - Normalize: $L_i' \rightarrow \frac{L_i' - \mu}{\sigma}$
   - Data Augmentation: $L_i' \rightarrow \{L_i''\}$(Shear, Zoom, Flip (horizontal and vertical), Rotation)
3. Define CNN-EffecientNetB0 Model:
   Input: $224 \times 224 \times 3$
   CNN Branch (local feature extraction)
   Conv2D (32, (3,3), activation='relu')
   MaxPooling2D (pool size= (2, 2))
   Conv2D (64, (3, 3), activation='relu')
   MaxPooling2D (pool size= (2, 2))
   Flatten ()
   Dense (256, activation='relu')
   EfficientNetB0 Branch (Global Feature Extraction)
   GlobalAveragePooling2D ()
   Concatenate ()
   Dense (256, activation='relu')
   Dropout (0.2)
   Dense (128, activation='relu')
   Dense (1, activation='sigmoid')
4. Compile the model P:
   optimizer=Adam ()
   learning rate=0.000001
   loss function=binary _crossentropy
5. Train the model P:
   Fit the model: P. fit (L_{train} ,M_{train}validation_data= (L_{val},M_{val}), batch size=32, epochs=50).
6. Evaluate the model P:
   Evaluate: P. eval ((L_{test} , M_{test}),

Save the Model
End

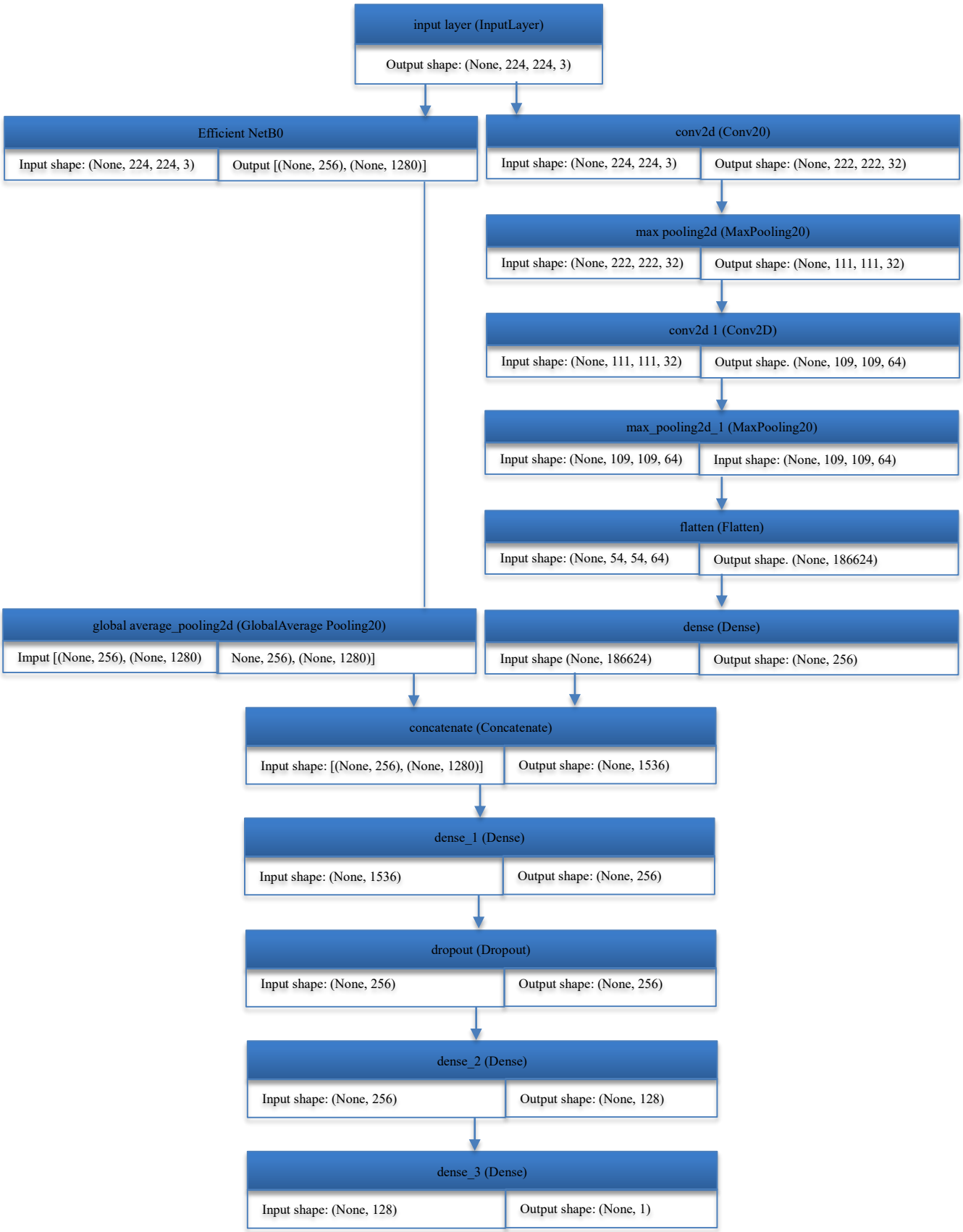

Fig. 6 Proposed model architecture

3.5. Software and Hardware Setup

The proposed system is implemented on the Google Colaboratory platform, using Python and the Keras framework for the entire development progression. Google Colab provides pre-installed Tensor Flow and access to high-performance resources, including a Graphics Processing Unit (GPU), 68.50 GB of storage and 12.75 GB of RAM, all operating in a 64-bit Windows 10 environment. Python, renowned for its flexibility, offers an intuitive syntax along with robust support through various libraries and frameworks. The system's effectiveness was assessed by evaluating its performance on a test dataset. Hyperparameters, predefined settings that influence how a deep learning model learns from data, were optimized using empirical methods, as detailed in Table 2.

Table 2. Hyperparameters of the proposed model

Hyperparameters	Values
Learning rate	0.000001
Dropout	0.2
Optimizer	Adam
Number of Epochs	25
Activation Function	ReLU, Sigmoid
Batch Size	32
Loss Function	Binary crossentropy

4. Results and Discussion

The model's performance at each epoch of training and validation is graphically represented by the accuracy plot. It helps to identify trends such as improvement, decreasing, or overfitting. By comparing training and validation accuracy, the plot delivers information about the system's generalisation ability and highlights the learning process's effectiveness. A loss plot visualizes the training and validation loss values of a framework throughout each epoch. It helps evaluate the model's learning process by showing its ability to minimize error during training and generalize to unseen data. A properly decreasing training loss combined with a validation loss that stabilizes or slightly decreases indicates effective training and good generalization. Figure 7 displays the accuracy and loss plots for the suggested framework.

In the initial epoch, the training accuracy starts at 50.60%and the validation accuracy is at 49.45%, indicating the difficulty of the model to generalize and capture meaningful patterns from the data. The training loss is high, reflecting the model's early stage of learning, while the validation loss shows that it is still far from optimal performance. As training progresses, accuracy improves significantly, and by the final epoch (Epoch 25), training accuracy reaches 97.04%. This shows that the model has learned to extract features effectively and generalize well to unseen data. When considering the system loss, the training loss is relatively high at the initial epoch at 0.6964, and the validation loss is slightly lower at 0.6923, indicating that the model is still in its early learning phase and has not yet optimized its parameters effectively. Over the epochs, the training loss gradually reduces as the model learns to minimize errors, reaching a significantly lower value of 0.1507 by epoch 25. As the training progresses, the model stabilizes, with both loss and accuracy values aligning better between the training and validation datasets. By the later epochs, the consistent reduction in both losses and the increase in accuracy shows that the system has effectively learned strong patterns and attained better classification performance.

Evaluation parameters play a central role in assessing the efficacy of the suggested DL framework, offering information about its predictive accuracy and classification efficiency. Metrics like accuracy, recall, precision and F1-score offer complementary perspectives on the model's effectiveness, highlighting its strengths and limitations. These metrics are particularly helpful in identifying challenges like overfitting, underfitting, or the effects of class imbalance, ensuring the system's robustness and reliability. The mathematical definitions of these metrics are presented in Equations (11) to (14), serving as the foundation for the computation and analysis.

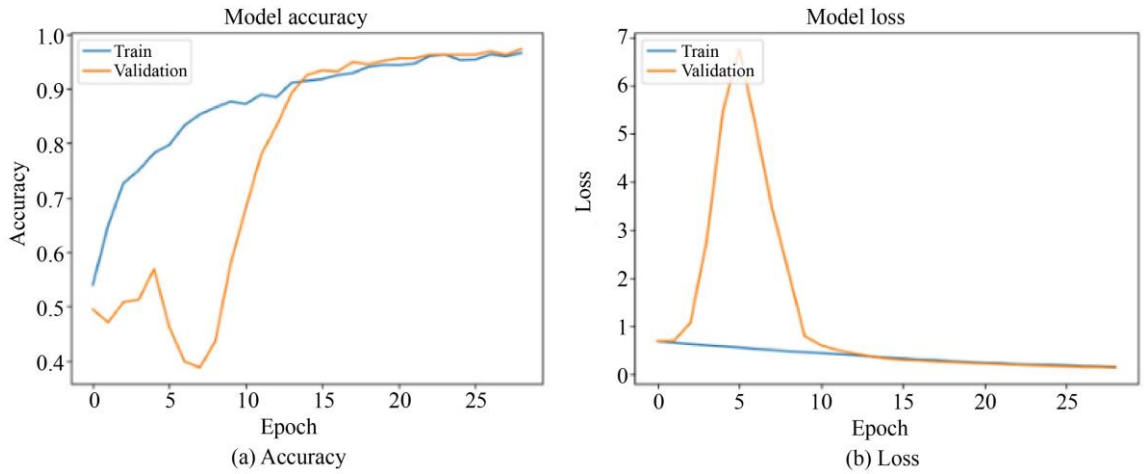

Fig. 7 Accuracy and loss plot of the suggested model

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

$$F1\ Score = 2 * \frac{(Recall*Precision)}{(Recall+Precision)} \quad (14)$$

Where TN = True Negative, TP = True Positive, FP = False Positive, FN = False Negative.

Figure 8 displays a detailed graphical investigation of the suggested model in the BC diagnosis. The approach attains a high accuracy of 98.68%, which shows that the model is highly operative in classifying between healthy and cancerous cases. Precision, at 98.75%, emphasizes the model's ability to identify correctly positive cases such that most predicted positives are actually positive. Likewise, the 98.75% recall illustrates how the model performs well in classifying true positive instances correctly, with very few occurrences of false negatives. The F1 score of 98.75% highlights the model's overall efficiency, effectively capturing both precision and recall to minimize errors in classification. These high values across the evaluation metrics suggest that the proposed framework efficiently detects bladder cancer, offering high reliability in clinical applications.

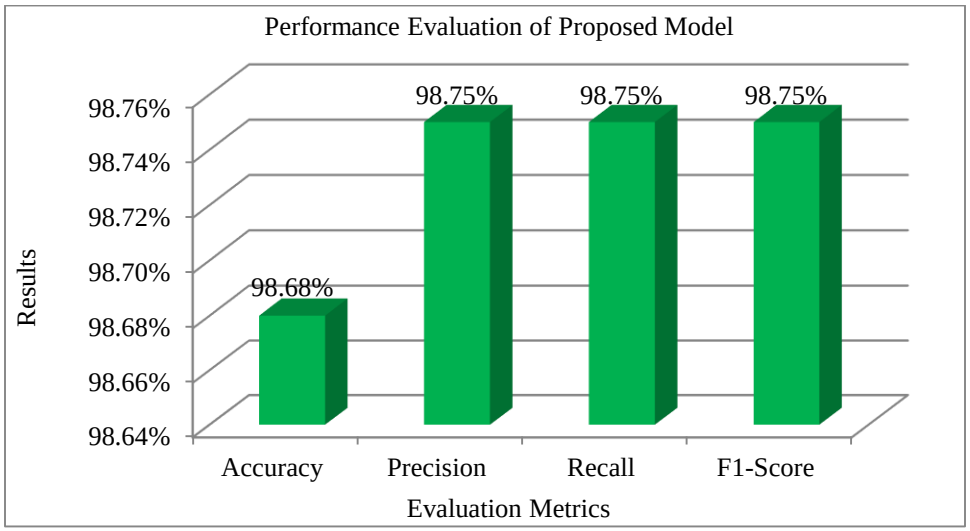

Fig. 8 Performance evaluation of the suggested model

Table 3. Classification report of the proposed model

	Precision	Recall	F1-score
Healthy	0.99	0.99	0.99
UCC	0.99	0.99	0.99
Accuracy		0.99	
macro avg	0.99	0.99	0.99
weighted avg	0.99	0.99	0.99

Table 3 presents the classification report of the proposed framework, showing impressive performance across all evaluation metrics. The model attains a precision, recall, and F1-score of 0.99 for both the cases, "Healthy" and "UCC", demonstrating a strong ability to accurately identify both healthy and cancerous cases. The 0.99 overall accuracy reflects the approach's effectiveness in correctly predicting both classes. The weighted average and macro average scores, both at 0.99, reinforce the consistency and robustness of the framework in balancing precision and recall across different classes.

A confusion matrix is utilized to assess the efficiency of a classification method by visualizing the counts of correct and incorrect predictions. It organizes these results into a matrix that compares the actual labels with the predicted ones. Figure 9 provides a thorough assessment of the model's performance in classifying bladder cancer. Out of 144 samples, 142 healthy samples were correctly classified as healthy, indicating that the model properly identifies healthy cases with high reliability and 2 healthy case samples were incorrectly classified as UCC with a low false positive rate. Likewise, out of 161 samples for cancerous cases, 2 samples were incorrectly classified as healthy, and 159 samples were properly predicted as cancerous. In general, this confusion matrix shows a very efficient model with minimal misclassification, leading to a strong diagnosis of BC with high reliability.

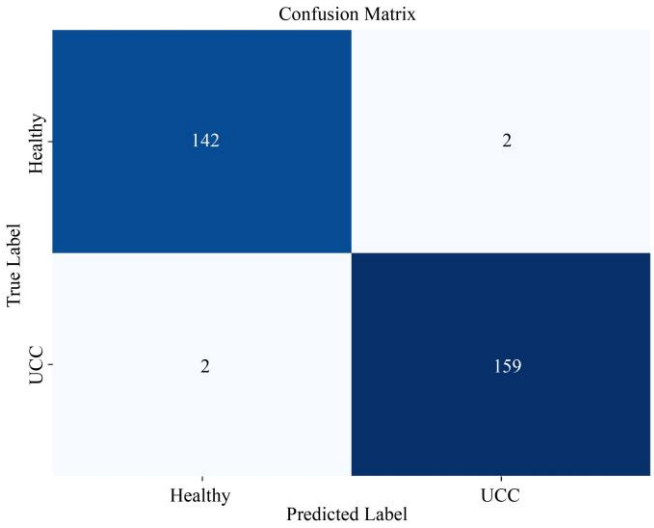

Fig. 9 Confusion matrix of the proposed system

The Receiver Operating Characteristic (ROC) curve represents a graphical tool employed to evaluate the performance of a classification approach at varying threshold levels. The TP rate is plotted against the FP rate in the ROC curve, representing the trade-off between the two measures. The ROC curve presented in Figure 10 displays the overall performance of the CNN-Efficient Net-hybrid model discriminating healthy and UCC cases.

The curve went in the direction of the top-left corner, indicating good performance in classification with a high rate of TP and low rates of FP. In the case of the value of AUC being almost 1, it means the model's near-perfect ability to predict bladder cancer. The result points out the performance and reliability of the suggested model in delivering a precise diagnosis of bladder cancer.

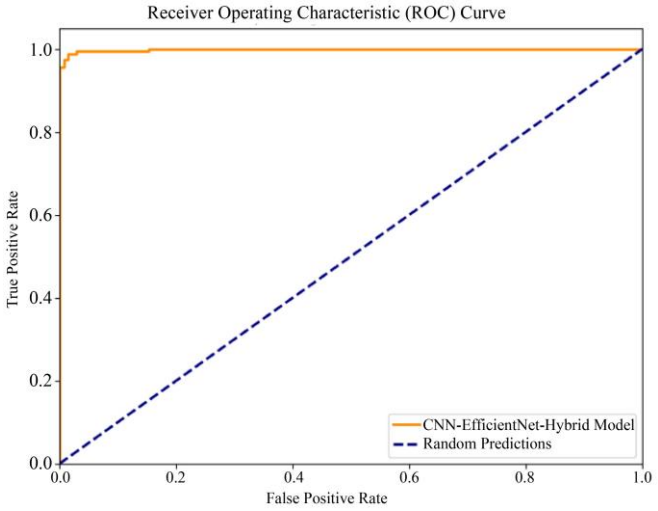

Fig. 10 ROC curve of the suggested model

The model correctly identified a randomly chosen image from the dataset as either "healthy" or "UCC," as demonstrated in Figure 11. This accurate categorization demonstrates how well the model can detect and group images. The results further demonstrate the system's reliability and strong performance in handling the dataset.

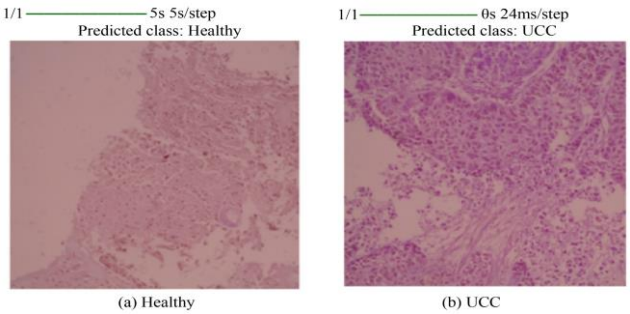

Fig. 11 Predicted output of bladder cancer (a)Healthy, and (b)UCC.

Table 4. Accuracy comparison of the suggested model with existing approaches

Authors & Ref	Method	Dataset	Accuracy
Du et al. [7]	CNN	734 bladder tumor images and 1,002 normal bladder tissue images	82.9%
García et al. [12]	DCEAC	Histological images	90.34%
Yoo et al. [4]	ResNeXt-101-32	10,991 cystoscopy images	94.1%
Hammouda et al. [9]	CNN	MRI data (42 cases)	95.24%
Yang et al. [10]	EasyDL	1,200 cancer images &1,150 non-cancer images	96.9%,
Proposed Model		**1435 healthy images and 1610 UCC images**	**98.68%**

Table 4 provides a comparison of the accuracy of the suggested CNN-EfficientNet approach with that of existing methods for diagnosing bladder cancer. The proposed model attains 98.68% accuracy, surpassing the next-best approach, the EasyDL framework, which achieves 96.9%. Other methods, such as the CNN-based approach using MRI data (95.24%) andResNeXt-101-32 backbone model trained on a large cystoscopy image dataset (94.1%), also demonstrate high accuracy but still fall short of the suggested model's performance. Models such as DCEAC (90.34%) and CNN trained using the Caffe framework (82.9%) further illustrate the varying levels of diagnostic effectiveness across different methods and datasets. Figure 12 graphically illustrates the accuracy of the comparison between the proposed model and

the existing approaches. The superior performance of the CNN-Efficient Net model is attributed to its advanced architectural design, which combines efficiency and accuracy by optimizing feature extraction and representation. Unlike traditional CNN architectures, Efficient Net employs a compound scaling strategy that balances network depth, resolution and width, allowing it to process complex features more effectively. Moreover, the proposed model is trained on a heterogeneous database of 1,435 normal and 1,610 bladder cancer images, thus increasing its generalizability and robustness. This synergistic combination of novel architecture and high-quality dataset makes the proposed model superior, increasing its accuracy and reliability as a tool for bladder cancer diagnosis.

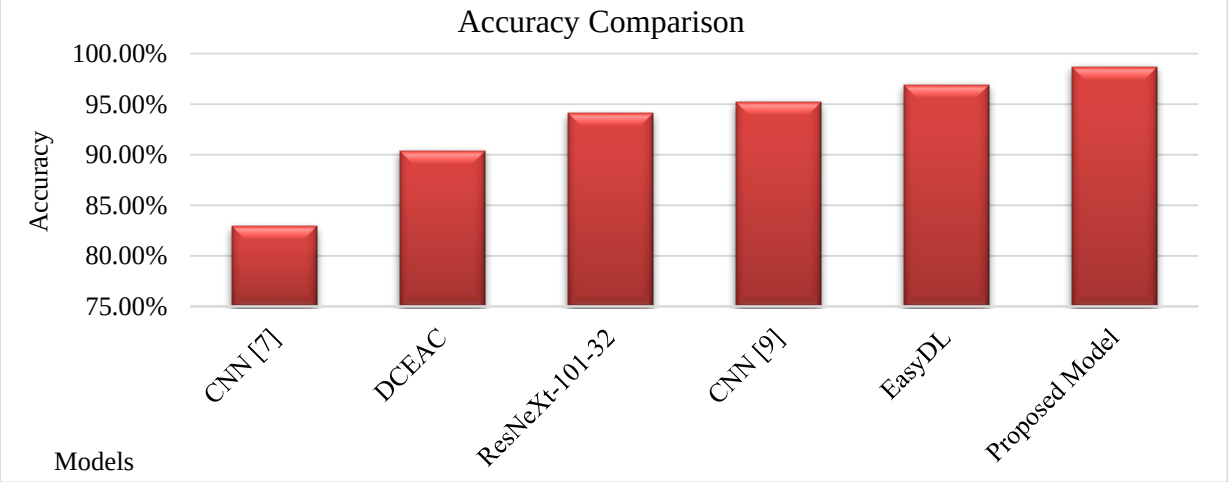

Fig. 12 Accuracy comparison of the proposed model with existing approaches

5. Conclusion

Bladder cancer, a common malignancy characterized by abnormal cell growth in the bladder wall, is dangerous to health if not detected early. Conventional diagnostic techniques tend to be challenged with low accuracy, delayed diagnosis and poor feature extraction, which can hamper early intervention. To overcome these issues, this research suggested a hybrid DL model that integrates CNNs for local feature analysis and EfficientNetB0 for global feature extraction from histopathological images. The model was learned and tested using a collection of 3,045 histopathological images, with 1,435 classified as "healthy" and 1,610 classified as "UCC". The framework proved exceptional performance, achieving an F1 score of 98.75%, 98.68% accuracy, 98.75% precision and 98.75% recall, surpassing the traditional models considerably. These outcomes show the model's consistency in accurately classifying histopathological images. However, the study would benefit from a more detailed discussion of its limitations, such as potential overfitting due to dataset size and lack of external validation. Additionally, future work could enhance generalizability by incorporating larger and more diverse datasets, integrating multimodal imaging such as CT and MRI and adopting explainable AI techniques to improve interpretability and clinical adoption. Real-time processing capabilities could also be explored to facilitate practical deployment in clinical settings.

Acknowledgements

I would like to express my sincere gratitude to all those who contributed to completing this research paper. I extend my heartfelt thanks to my supervisor, my family, my colleagues and fellow researchers for their encouragement and understanding during the demanding phases of this work.

References

[1] Chao-Zhe Zhu et al., "A Review on the Accuracy of Bladder Cancer Detection Methods," *Journal of Cancer*, vol. 10, no. 17, pp. 4038-4044, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[2] Milena Taskovska, Mateja Erdani Kreft, and Tomaž Smrkolj, "Current and Innovative Approaches in the Treatment of Non-Muscle Invasive Bladder Cancer: The Role of Transurethral Resection of Bladder Tumor and Organoids," *Radology and Oncology*, vol. 54, no. 2, pp. 135-143, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Vincenzo K. Wong et al., "Imaging and Management of Bladder Cancer," *Cancers*, vol. 13, no. 6, pp. 1-22, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Jeong Woo Yoo et al., "Deep Learning Diagnostics for Bladder Tumor Identification and Grade Prediction using RGB Method," *Scientific Reports*, vol. 12, pp. 1-8, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Suryadipto Sarkar et al., "Performing Automatic Identification and Staging of Urothelial Carcinoma in Bladder Cancer Patients Using a Hybrid Deep-Machine Learning Approach," *Cancers*, vol. 15, no. 6, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Wayner Barrios et al., "Bladder Cancer Prognosis using Deep Neural Networks and Histopathology Images," *Journal of Pathology Informatics*, vol. 13, pp. 1-9, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Yang Du et al., "A Deep Learning Network-Assisted Bladder Tumour Recognition Under Cystoscopy Based on Caffe Deep Learning Framework and EasyDL Platform," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 17, no. 1, pp. 1-8, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Atsushi Ikeda et al., "Cystoscopic Imaging for Bladder Cancer Detection Based on Stepwise Organic Transfer Learning with a Pretrained Convolutional Neural Network," *Journal of Endourology*, vol. 35, no. 7, pp. 1030-1035, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] K. Hammouda et al., "A Multiparametric MRI-based CAD System for Accurate Diagnosis of Bladder Cancer Staging," *Computerized Medical Imaging and Graphics*, vol. 90, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Rui Yang et al., "Automatic Recognition of Bladder Tumours using Deep Learning Technology and its Clinical Application," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 17, no. 2, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Ivan Lorencin et al., "On Urinary Bladder Cancer Diagnosis: Utilization of Deep Convolutional Generative Adversarial Networks for Data Augmentation," *Biology*, vol. 10, no. 3, pp. 1-27, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Gabriel García et al., "A Novel Self-Learning Framework for Bladder Cancer Grading using Histopathological Images," *Computers in Biology and Medicine*, vol. 138, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Atsushi Ikeda et al., "Support System of Cystoscopic Diagnosis for Bladder Cancer Based on Artificial Intelligence," *Journal of Endourology*, vol. 34, no. 3, pp. 352-358, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Gumuyang Zhang et al., "CT-based Radiomics to Predict the Pathological Grade of Bladder Cancer," *European Radiology*, vol. 30, pp. 6749-6756, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Peng-Nien Yin et al., "Histopathological Distinction of Non-Invasive and Invasive Bladder Cancers using Machine Learning Approaches," *BMC Medical Informatics and Decision Making*, vol. 20, pp. 1-11, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] K. Hammouda et al., "A 3D CNN with a Learnable Adaptive Shape Prior for Accurate Segmentation of Bladder Wall Using MR Images," *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Iowa City, IA, USA, pp. 935-938, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[17] Ilaria Jansen et al., "Automated Detection and Grading of Non–Muscle-Invasive Urothelial Cell Carcinoma of the Bladder," *The American Journal of Pathology*, vol. 190, no. 7, 1483-1490, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[18] Eugene Shkolyar et al., "Augmented Bladder Tumor Detection using Deep Learning," *European Urology*, vol. 76, no. 6, pp. 714-718, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Yusra Ameen et al., *Hematoxylin-and-Eosin-Stained Bladder Urothelial Cell Carcinoma Versus Inflammation Digital Histopathology Image Dataset*, Data Sheet, 2023. [Publisher Link]

[20] Tahsin Istiyaq et al., "*Polycystic Ovary Syndrome Detection Using Neural Network*," Thesis, Department of Computer Science and Engineering, Brac University, pp. 1-30, 2023. [Google Scholar] [Publisher Link]

[21] Adnan Hussain et al., "An Efficient and Robust Hand Gesture Recognition System of Sign Language Employing Finetuned Inception-V3 and Efficientnet-B0 Network," *Computer Systems Science & Engineering*, vol. 46, no. 3, pp. 3509-3525, 2023. [CrossRef] [Google Scholar] [Publisher Link]