Pretrained Convolutional Neural Networks for Bladder Cancer Diagnosis via White Light Cystoscopy

Haithem DAHIMI[1], Mhand HIFI[2], Fabien SAINT[3]

***Abstract*— Artificial intelligence enhances diagnostic accuracy and reduces subjectivity in medical imaging, especially in complex tasks like cancer detection. This study assesses several convolutional neural network architectures for classifying bladder cancer in white light cystoscopy images. Using sensitivity, specificity, and the hypervolume indicator, results show that DenseNet consistently outperforms ResNet and VGG16, offering a superior balance of diagnostic performance. These results support DenseNet's potential for clinical use in bladder cancer diagnosis.**

I. INTRODUCTION

Over the past decade, Artificial Intelligence (AI) has significantly transformed the field of oncology [1], [2]. Deep learning algorithms, particularly in image-based diagnostics, have shown accuracy and efficiency, often helping medical professionals in a variety of tasks. These advancements present significant opportunities for integrating AI into the bladder cancer diagnosis process using White Light Cystoscopy (WLC) images, offering the potential to improve diagnostic accuracy, accelerate clinical decision-making, and reduce costs.

Transfer learning is a key innovation in deep learning that facilitates the use of pre-trained Convolutional Neural Networks (CNNs) [14], originally trained on large datasets, for new but related tasks. By utilizing their learned features, it supports efficient adaptation to specific problems, reducing training time [13], lowering computational costs, and minimizing data requirements.

This study conducts a comparative analysis of various state-of-the-art pre-trained CNN architectures for the binary classification of WLC bladder cancer images (cf., Fig 1). The models are evaluated based on their ability to balance sensitivity and specificity—two critical metrics in medical diagnostics. To perform a thorough bi-objective comparison, the hypervolume indicator, a well-established metric in multi-objective optimization, is used to quantify the overall performance of each model. This approach provides a robust framework for identifying the most effective architecture for clinical deployment in bladder cancer diagnostics.

II. BACKGROUND REVIEW

Cystoscopy remains the gold standard for diagnosing bladder cancer. However, its diagnostic accuracy is highly dependent on the clinician's expertise, leading to subjectivity and variability in outcomes [3], [4]. To overcome these limitations, several studies have explored the use of deep learning techniques for cystoscopic image analysis. For instance, Lorencin *et al.* [5] employed Multi-Layer Perceptrons (MLPs) in conjunction with a Laplacian edge detector, achieving an Area Under the Curve (AUC) of 0.99 with a mid-sized MLP architecture.

Shkolyar *et al.* [6] developed a deep CNN designed to detect suspicious papillary lesions in WLC video frames. To enhance the robustness of the model, data augmentation techniques were employed during training. Their approach achieved outstanding diagnostic performance, reaching a sensitivity of 90.9% and a specificity of 98.6%.

On the other hand, Ikeda *et al.* [7] employed the pretrained GoogleNet model for diagnostic classification based on WLC images, reporting a sensitivity of 89.7% and a specificity of 94%. This highlights the powerful potential of transfer learning in medical image analysis, especially in computer-aided diagnostics.

In a large-scale, multicenter diagnostic study, Wu *et al.* [8] analyzed data from 10,729 patients across six hospitals, employing WLC to compile an extensive dataset comprising 69,204 images. To enhance the diversity and robustness of the training data, the researchers applied data augmentation techniques during preprocessing. They employed the Pyramid Scene Parsing Network (PSPNet), a deep learning model built on a pretrained ResNet101 backbone for feature extraction and classification tasks. The model achieved interesting diagnostic performance, reaching a sensitivity of 97.5% and a specificity of 98.3%, surpassing the diagnostic accuracy achieved by the urologists.

In a separate study focusing on a different imaging modality, Nairveen *et al.* [9] conducted a comparative analysis of several pretrained deep learning models using Blue Light Cystoscopy (BLC) images. The study evaluated the performance of VGG16, MobileNetV2, ResNet50, and InceptionV3 for bladder cancer classification. Among the models tested, MobileNetV2 delivered the best results, achieving a sensitivity of 95.77% and a specificity of 87.84%. However, by focusing on BLC, this study does not address the diagnostic challenges associated with WLC, which remains the most widely used modality in clinical practice. While BLC is known to enhance the visualization of certain tumor characteristics, its use is often limited due to higher costs, the need for specialized equipment, and less widespread availability.

[1]Haithem DAHIMI, EPROAD UR 4669, Université de Picardie Jules Verne, 80000 Amiens, France (haithem.dahimi@u-picardie.fr)

[2]Mhand HIFI EPROAD UR 4669, Université de Picardie Jules Verne, 80000 Amiens, France (hifi@u-picardie.fr)

[3]Fabien SAINT, EPROAD UR 4669, Université de Picardie Jules Verne, 80000 Amiens, France (fabien.saint@u-picardie.fr)

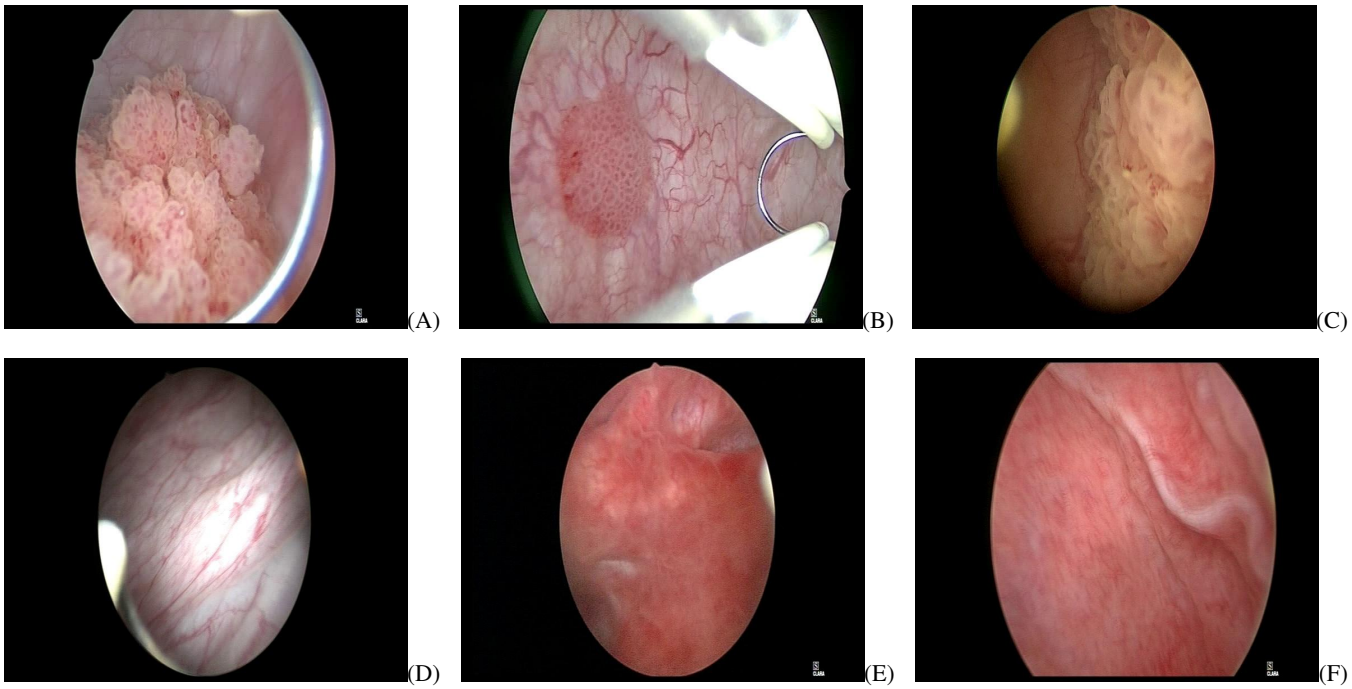

Fig. 1. Representative frames extracted from the proprietary WLC video dataset. Images [A, B, C] depict papillary endothelial bladder cancer tumours, whereas images [D, E, F] illustrate healthy bladders characterized by smooth inner walls.

III. Methodology and approach

A. Dataset description

A total of seventy-one white-light cystoscopy videos were collected from Amiens University Hospital. Among these, 58 videos presented papillary cancerous tumors, while the remaining 13 videos showed no signs of malignancy. To ensure diagnostic accuracy, biopsies were conducted to validate or challenge the results obtained from the cystoscopy examinations. The entire video dataset was subsequently divided into training and testing sets at an 80:20 ratio, ensuring that both sets maintained the same proportion of tumor and control cases, which is important for model generalization.

For each video, 200 RGB images with a resolution of 576×720 pixels were randomly extracted. The image frames were then manually cleaned to remove low-quality or irrelevant images, such as those with motion blur, poor lighting, or other artifacts that could compromise model performance. From each video containing a tumor, 20 frames were randomly selected, ensuring a balanced representation of tumor characteristics. Similarly, 40 frames were randomly chosen from control videos, aiming to address the inherent data imbalance between the two classes.

	Positive	Negative	Total
Train	900	400	1300
Test	240	120	360
Total	1140	520	1660

TABLE I
CHARACTERISTICS OF THE *Train-Test* DATASET

Note that this dataset preparation helps reduce potential biases in training, ensuring that the model is exposed to a representative variety of images from both tumor and control cases. The final dataset, ready for model training and evaluation, is summarized in Table I.

B. The designed method

In this study, a panel of pretrained CNN architectures was employed, each fine-tuned for the specific task of classifying bladder cancer images. The approach included training each model using different combinations of hyperparameters to identify the most effective configuration. Specifically, three learning rates $\{0.005;\ 0.001;\ 0.0005\}$ and two weight decay values $\{1e-4; 5e-4\}$ were tested, resulting in six distinct hyperparameter configurations per model. Each configuration was trained over 20 epochs, producing a total of 120 model instances for each CNN architecture. These configurations allowed for a detailed analysis of the model's performance across different training conditions and hyperparameter settings.

1) Pretrained VGG16:

VGG16, proposed by Simonyan and Zisserman in 2014 [10], was developed to explore the impact of network depth on image classification performance. It is distinguished by its use of small 3×3 convolutional kernels with a stride of 1 and same padding, stacked to effectively capture hierarchical image features.

The architecture comprises 16 weight layers, including 13 convolutional layers grouped into five blocks, each followed by max-pooling layers for spatial downsampling. These are

followed by three fully connected layers responsible for high-level feature interpretation and classification. Initially trained on the large-scale *ImageNet* dataset—containing over 1.3 million labeled images across 1,000 categories—VGG16 has become a benchmark model in computer vision. Its relatively simple yet deep structure, along with strong empirical performance, has made it widely adopted as both a feature extractor and a base model for transfer learning, particularly in applications such as medical image analysis.

2) Pretrained ResNet:

ResNet, introduced by He et al. in 2015 [11], is a deep convolutional neural network (CNN) architecture developed to address the degradation problem observed in very deep networks. Traditionally, increasing the depth of a neural network was expected to improve performance; however, beyond a certain point, deeper models began to suffer from higher training error and poor generalization. ResNet overcomes this limitation by introducing the concept of *residual learning*, where shortcut (or skip) connections allow the input of a given layer to bypass one or more intermediate layers and be directly added to the output. This approach enables the network to learn residual functions relative to the input, rather than attempting to learn complex unreferenced mappings directly.

The primary advantage of skip connections is their ability to improve gradient propagation, effectively addressing the issue of gradient attenuation that often hinders the training of deep networks. This mechanism allows ResNet architectures to be trained successfully at greater depths, enabling them to capture more complex, hierarchical features while preserving training stability and performance.

In this study, two ResNet variants—ResNet-50 and ResNet-152—were employed to investigate the impact of architectural depth on classification performance for bladder cancer detection. The comparison of these models explores:

1) how increasing network depth influences the extraction of discriminative features, and
2) how it affects key diagnostic metrics such as sensitivity and specificity.

Such an evaluation offers valuable insights into the suitability of deep residual networks for complex medical imaging tasks, particularly in contexts where diagnostic precision is paramount.

As highlighted above, a panel of CNN architectures was employed in this study, each fine-tuned to address the specific task of bladder cancer classification. For each architecture, models were trained using six distinct hyperparameter configurations, obtained by combining three learning rates $\{0.005,\ 0.001,\ 0.0005\}$ with two weight decay values $\{1\mathrm{e}{-4},\ 5\mathrm{e}{-4}\}$. Each configuration was trained for 20 epochs, resulting in a population of 120 model instances per architecture. Hence, these models were subsequently evaluated to assess their diagnostic performance and robustness.

3) Pretrained VGG16:

VGG16, introduced by Simonyan and Zisserman in 2014 [10], is a deep CNN designed to investigate the effect of network depth on large-scale image classification performance. Such an architecture is distinctive for its use of small 3×3 convolutional filters, each with a stride of 1 and same padding, arranged in a series. This sequential stacking allows the network to efficiently capture and learn intricate patterns and hierarchical structures in images.

The network consists of 16 weight layers, including 13 convolutional layers grouped into five convolutional blocks. Each block is followed by max-pooling layers, which perform spatial downsampling. The model concludes with 3 fully connected layers that are responsible for high-level reasoning and classification. VGG16 was trained on the large-scale ImageNet dataset, which includes approximately 1.3 million labeled images across 1,000 object categories. Thanks to its simple yet effective architecture, VGG16 has become a cornerstone in computer vision, often used as a feature extractor and benchmark in transfer learning for various downstream tasks.

4) Pretrained ResNet:

ResNet, a deep convolutional neural network (CNN) architecture, was introduced by He et al. in 2015 [11]. The primary objective of ResNet is to address the degradation problem observed in very deep networks, where, contrary to theoretical expectations, increasing depth results in higher training error. This issue arises because the network becomes more challenging to optimize as the number of layers increases, which leads to performance degradation.

The fundamental advancement of ResNet lies in the introduction of residual learning, facilitated by skip connections. These connections allow the input of a layer to bypass one or more intermediate layers and be directly added to the output, creating shortcut paths for gradient flow. This allows the network to learn residual functions, i.e., the difference between the desired output and the input, rather than attempting to learn the direct mapping. By focusing on residuals instead of direct mappings, the network more effectively handles deeper architectures.

A key advantage of residual learning is its ability to reduce the vanishing gradient problem, a common issue in deep networks. Skip connections allow gradients to propagate more effectively throughout the network, preserving important information during backpropagation and enabling the training of significantly deeper architectures. Consequently, ResNet models are able to capture more complex features while maintaining stability during training.

In this study, two variants of the ResNet architecture are employed: ResNet-50 and ResNet-152. The difference between these two models lies in the number of layers, with ResNet-50 containing 50 layers and ResNet-152 consisting of 152 layers. This comparison aims to investigate the impact of network depth on the performance of the model in bladder cancer detection. By evaluating these two variants, the study seeks to explore how increasing architectural depth influences the extraction of discriminative features and overall model effectiveness in terms of key diagnostic metrics, particularly *sensitivity* and *specificity*.

This analysis provides valuable results into the applicability of deep residual networks in medical image classification, where precision in feature extraction is important.

5) Pretrained DenseNet:

Densely connected convolutional neural networks (namely DenseNets), introduced by Huang et al. [12], extend the principles of residual learning by establishing direct connectivity between all layers within a dense block. Building upon ResNet's skip connections, DenseNet architectures allow each layer to receive, as input, the concatenated feature maps from all preceding layers. It promotes extensive feature reuse and enhances gradient flow throughout the network. In deeper DenseNet models, layers are organized into modular components known as dense blocks. These dense blocks are interconnected by transition layers, which consist of batch normalization, a 1×1 convolution for feature compression, and a pooling operation for spatial downsampling.

Two variants of DenseNet are explored, each with different network depths: DenseNet-121 and DenseNet-201.

C. Assessment Metrics

A comparative assessment of the models was conducted using relevant performance metrics, with a focus on *"sensitivity"* and *"specificity"* due to their critical role in bladder cancer diagnostics. The sensitivity (also known as *"recall"*) reflects the model's ability to correctly identify all true cancer cases, which is important for ensuring timely intervention and potentially life-saving results. The specificity, on the other hand, measures the model's ability to correctly identify non-cancer cases, minimizing the risk of unnecessary and often invasive follow-up procedures for patients.

Additionally, the overall *"accuracy"* of the models was assessed to provide a broader understanding of their effectiveness in the context of this task. Such a metric evaluates the proportion of correct predictions across all cases, giving results into the general performance of the model.

The formulas for these metrics are as follows:

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{1}$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \tag{2}$$

$$\text{Accuracy} = \frac{\text{True Negatives} + \text{True Positives}}{\text{All predictions}} \tag{3}$$

To offer a more thorough evaluation of model performance, the "Hypervolume Indicator" (HI) was incorporated. This metric quantifies the trade-offs between *"sensitivity"* and *"specificity"* by identifying the set of non-dominated solutions for each model. The HI is computed relative to a reference point at (sensitivity = 0, specificity = 0), representing the area covered by the non-dominated solutions within the objective space. By doing so, it provides a more complete assessment of model performance, capturing the interplay between multiple objectives and allowing for a nuanced comparison across different models.

D. Resources

The implementation and training of all models in this study were performed using the PyTorch framework. To accelerate the computational process, all experiments were conducted on an NVIDIA Tesla V100 GPU with 32 GB of memory, provided by the Matrics computing platform at the University of Amiens [15]. This high-performance infrastructure enabled efficient training and evaluation of deep learning models.

IV. Preliminary experimental results

In this section, the experimental results of the pretrained models are presented for the prediction of bladder cancer using the proprietary dataset. The evaluation focuses on examining how different architectural choices, variations in hyperparameter configurations, and model depth impact overall predictive performance. A comparative analysis is conducted to highlight the trade-offs between sensitivity and specificity, ultimately identifying the most effective configurations for this specific medical diagnosis task.

It is important to note that this work represents an initial experimental study, where the main objective was to establish a baseline comparison between several pretrained CNN architectures for the classification of white light cystoscopy images. The selection of hyperparameters was deliberately limited to a small grid of commonly used values from the literature. This design choice aimed to ensure a coherent and reproducible experimental setup, while also keeping the computational cost reasonable at this early stage. Future work will aim to expand this analysis by exploring a broader hyperparameter space, potentially through automated tuning methods such as cross-validation or Bayesian optimization.

Now, the results detailed in Table II are commented on:

1) VGG16 displayed the poorest performance across all hyperparameter configurations, with significant evidence of overfitting. This is particularly evident in the extremely high sensitivity paired with low specificity, suggesting that the model predominantly predicted positive cases while failing to correctly identify negative ones. Interestingly, it was observed that lower learning rates and smaller weight decay values contributed to an improvement in VGG16's performance. The configuration with $lr = 0.0005$ and $weight\_decay = 1e-4$ resulted in higher mean accuracy, more balanced sensitivity and specificity, and reduced standard deviation, highlighting more stable model behavior compared to the other configurations.
2) The ResNet architectures outperformed VGG16 by a significant margin, demonstrating more stable and reliable performance across various hyperparameter settings. However, similar to VGG16, ResNet models showed sensitivity to higher learning rates, which often resulted in suboptimal accuracy and an imbalance between sensitivity and specificity. This suggests that, while the residual connections in ResNet enable deeper learning and improved gradient flow, they do not fully

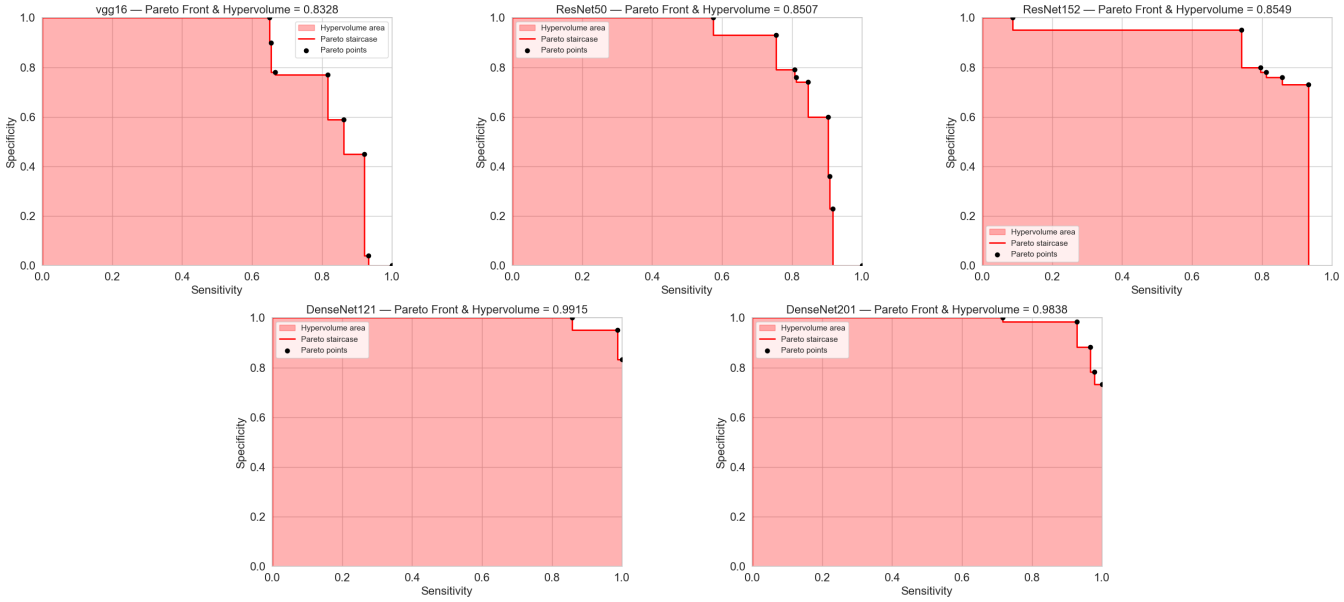

Fig. 2. Illustration of Pareto fronts and hypervolume indicator values for deep learning architectures: VGG16, ResNet50, ResNet152, DenseNet121, and DenseNet201.

Model	Mean Hyperparameter	Mean Accuracy	Mean Sensitivity	Mean Specificity
VGG16	lr=0.005, weight_decay=1e-4	0.0500 ± 0.2179	0.9500 ± 0.2179	0.3147 ± 0.0897
	lr=0.005, weight_decay=5e-4	0.1000 ± 0.3000	0.9000 ± 0.3000	0.3353 ± 0.1235
	lr=0.001, weight_decay=1e-4	0.0542 ± 0.2177	0.9500 ± 0.2179	0.3176 ± 0.0900
	lr=0.001, weight_decay=5e-4	0.0000 ± 0.0000	1.0000 ± 0.0000	0.2941 ± 0.0000
	lr=0.0005, weight_decay=1e-4	0.6354 ± 0.1551	0.7130 ± 0.2381	0.6582 ± 0.0936
	lr=0.0005, weight_decay=5e-4	0.5013 ± 0.3101	0.6185 ± 0.3056	0.5357 ± 0.1422
ResNet50	lr=0.005, weight_decay=1e-4	0.0250 ± 0.1090	0.9910 ± 0.0392	0.3091 ± 0.0654
	lr=0.005, weight_decay=5e-4	0.0500 ± 0.2179	0.9500 ± 0.2179	0.3147 ± 0.0897
	lr=0.001, weight_decay=1e-4	0.6212 ± 0.2733	0.7700 ± 0.1574	0.6650 ± 0.1625
	lr=0.001, weight_decay=5e-4	0.4687 ± 0.3184	0.6650 ± 0.2766	0.5265 ± 0.1541
	lr=0.0005, weight_decay=1e-4	0.7458 ± 0.1385	0.6385 ± 0.1787	0.7143 ± 0.0615
	lr=0.0005, weight_decay=5e-4	0.7215 ± 0.1124	0.7730 ± 0.1299	0.7366 ± 0.0576
ResNet152	lr=0.005, weight_decay=1e-4	0.0000 ± 0.0000	1.0000 ± 0.0000	0.2941 ± 0.0000
	lr=0.005, weight_decay=5e-4	0.0125 ± 0.0000	0.9900 ± 0.0000	0.3000 ± 0.0000
	lr=0.001, weight_decay=1e-4	0.5296 ± 0.3513	0.8310 ± 0.1430	0.6182 ± 0.2165
	lr=0.001, weight_decay=5e-4	0.3000 ± 0.2967	0.8440 ± 0.1778	0.4600 ± 0.1655
	lr=0.0005, weight_decay=1e-4	0.5775 ± 0.2999	0.7755 ± 0.1576	0.6357 ± 0.1776
	lr=0.0005, weight_decay=5e-4	0.6556 ± 0.2888	0.6930 ± 0.1778	0.6666 ± 0.1674
DenseNet121	lr=0.005, weight_decay=1e-4	0.6996 ± 0.1978	0.8842 ± 0.1352	0.7365 ± 0.1489
	lr=0.005, weight_decay=5e-4	0.0000 ± 0.0000	0.9992 ± 0.0036	0.1998 ± 0.0007
	lr=0.001, weight_decay=1e-4	0.7537 ± 0.1917	0.8833 ± 0.0719	0.7797 ± 0.1452
	lr=0.001, weight_decay=5e-4	0.7258 ± 0.1320	0.9192 ± 0.0692	0.7645 ± 0.0941
	lr=0.0005, weight_decay=1e-4	0.8065 ± 0.1267	0.8642 ± 0.1223	0.8180 ± 0.0852
	lr=0.0005, weight_decay=5e-4	0.8446 ± 0.1280	0.8475 ± 0.1410	0.8452 ± 0.0879
DenseNet201	lr=0.005, weight_decay=1e-4	0.0000 ± 0.0000	1.0000 ± 0.0000	0.2000 ± 0.0000
	lr=0.005, weight_decay=5e-4	0.4477 ± 0.4369	0.5842 ± 0.4245	0.4750 ± 0.2673
	lr=0.001, weight_decay=1e-4	0.7175 ± 0.2262	0.8567 ± 0.1864	0.7453 ± 0.1662
	lr=0.001, weight_decay=5e-4	0.7208 ± 0.2187	0.8975 ± 0.1098	0.7562 ± 0.1647
	lr=0.0005, weight_decay=1e-4	0.8017 ± 0.2643	0.8858 ± 0.1132	0.8185 ± 0.2009
	lr=0.0005, weight_decay=5e-4	0.7431 ± 0.2239	0.9250 ± 0.0526	0.7795 ± 0.1715

TABLE II

MEAN AND STANDARD DEVIATION OF PERFORMANCE METRICS (ACCURACY, SENSITIVITY, AND SPECIFICITY) ACROSS TRAINING EPOCHS FOR EACH COMBINATION OF HYPERPARAMETER SETTINGS AND PRETRAINED MODEL ARCHITECTURE.

compensate for the destabilizing effects caused by aggressive learning rates.

3) In comparing ResNet50 and ResNet152, the results were mixed. Increasing the depth of the architecture

from 50 to 152 layers did not always lead to performance gains. In fact, ResNet152 showed higher variability and, in some cases, even underperformed relative to the shallower ResNet50 model. This observation challenges the assumption that deeper architectures are inherently superior, particularly in the context of smaller, domain-specific datasets such as those used in medical imaging. The challenges of overfitting and convergence in deeper architectures become more pronounced when working with such datasets.

4) Among the ResNet models, ResNet50 with a learning rate of $lr = 0.0005$ produced the best overall performance, achieving a strong balance between sensitivity and specificity while minimizing the variance. This configuration proved to be the most reliable among all ResNet models tested.
5) DenseNet architectures consistently delivered the best performance across all evaluation metrics, particularly when trained with lower learning rates. The key factor behind DenseNet's superior performance is its densely connected convolutional structure, which facilitates efficient feature reuse throughout the network. Similar to the ResNet models, DenseNet121 outperformed DenseNet201, underscoring the notion that increasing network depth does not always lead to better performance. In fact, DenseNet121 achieved the highest accuracy, sensitivity, and specificity, with lower variance across runs. These results solidify DenseNet121 as the top-performing architecture in this study.

An analysis of the Pareto front and the corresponding hypervolume indicator, constructed using the sensitivity and specificity of each model (Fig.2), further supports the findings presented in TableII. DenseNet architectures achieved the highest hypervolume values, indicating a strong and consistent trade-off between sensitivity and specificity. Importantly, DenseNet121 slightly outperformed DenseNet201, reinforcing its ability to maintain a more favorable balance across these critical performance metrics. The provided result confirms DenseNet121's superiority in medical diagnostic tasks, where achieving both high sensitivity and specificity is essential for reliable clinical decision-making.

V. Conclusion

The proposed preliminary study presents a comparative evaluation of pre-trained CNN architectures for binary classification of bladder cancer using white light cystoscopy images. Sensitivity and specificity were the primary diagnostic metrics, and the hypervolume indicator was used to assess overall performance. With the limited computational results, DenseNet121 emerged as the most promising architecture, offering the best trade-off between sensitivity and specificity.

Hyperparameter tuning had a significant impact on performance, with lower learning rates and weight decay values yielding greater model stability. However, deeper architectures did not consistently improve results, highlighting the need to align model complexity with task constraints. To ensure reproducibility and manage computational cost, hyperparameters were selected from a small, literature-informed grid.

As a first step, this study establishes a baseline for future work, which will include advanced hyperparameter optimization methods (e.g., cross-validation, Bayesian optimization), expansion of the dataset to cover a wider range of histological subtypes, and the development of custom DenseNet-inspired models adapted to real-world clinical scenarios. Future studies should also incorporate complementary evaluation tools—such as confusion matrices—to provide a more intuitive and detailed visualization of model performance across prediction classes.

References

[1] Z. H. Chen, L. Lin, C. F. Wu, C. F. Li, R. H. Xu, and Y. Sun, "Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine," *Cancer Commun. (Lond.)*, vol. 41, no. 11, pp. 1100–1115, Nov. 2021, doi: 10.1002/cac2.12215.

[2] H. Shimizu and K. I. Nakayama, "Artificial intelligence in oncology," *Cancer Sci.*, vol. 111, no. 5, pp. 1452–1460, May 2020, doi: 10.1111/cas.14377.

[3] C. S. Guldhammer, J. L. Vásquez, V. M. Kristensen, T. Norus, N. Nadler, J. B. Jensen, and N. Azawi, "Cystoscopy accuracy in detecting bladder tumors: A prospective video-confirmed study," *Cancers (Basel)*, vol. 16, no. 1, p. 160, Dec. 2023, doi: 10.3390/cancers16010160.

[4] C. Z. Zhu, H. N. Ting, K. H. Ng, and T. A. Ong, "A review on the accuracy of bladder cancer detection methods," *J. Cancer*, vol. 10, no. 17, pp. 4038–4044, Jul. 2019, doi: 10.7150/jca.28989.

[5] I. Lorencin, N. Andjelić, J. Španjol, and Z. Car, "Using multi-layer perceptron with Laplacian edge detector for bladder cancer diagnosis," *Artif. Intell. Med.*, vol. 102, p. 101746, Jan. 2020, doi: 10.1016/j.artmed.2019.101746.

[6] E. Shkolyar, X. Jia, T. C. Chang, D. Trivedi, K. E. Mach, M. Q. Meng, L. Xing, and J. C. Liao, "Augmented bladder tumor detection using deep learning," *Eur. Urol.*, vol. 76, no. 6, pp. 714–718, Dec. 2019, doi: 10.1016/j.eururo.2019.08.032.

[7] A. Ikeda, H. Nosato, Y. Kochi, T. Kojima, K. Kawai, H. Sakanashi, M. Murakawa, and H. Nishiyama, "Support system of cystoscopic diagnosis for bladder cancer based on artificial intelligence," *J. Endourol.*, vol. 34, no. 3, pp. 352–358, Mar. 2020, doi: 10.1089/end.2019.0509.

[8] S. Wu, X. Chen, J. Pan, W. Dong, X. Diao, R. Zhang, Y. Zhang, Y. Zhang, G. Qian, H. Chen, H. Lin, S. Xu, Z. Chen, X. Zhou, H. Mei, C. Wu, Q. Lv, B. Yuan, Z. Chen, W. Liao, X. Yang, H. Chen, J. Huang, and T. Lin, "An artificial intelligence system for the detection of bladder cancer via cystoscopy: A multicenter diagnostic study," *J. Natl. Cancer Inst.*, vol. 114, no. 2, pp. 220–227, Feb. 2022, doi: 10.1093/jnci/djab179.

[9] N. Ali, C. Bolenz, T. Todenhöfer, A. Stenzel, P. Deetmar, M. Kriegmair, T. Knoll, S. Porubsky, A. Hartmann, J. Popp, M. C. Kriegmair, and T. Bocklitz, "Deep learning-based classification of blue light cystoscopy imaging during transurethral resection of bladder tumors," *Sci. Rep.*, vol. 11, no. 1, p. 11629, Jun. 2021, doi: 10.1038/s41598-021-91081-x.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[12] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 4700–4708.

[13] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On large-batch training for deep learning: Generalization gap and sharp minima," *arXiv preprint arXiv:1609.04836*, 2016.

[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[15] Matrics Platform, University of Amiens. [Online]. Available: `https://www.matrics.u-picardie.fr`