Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 8, S. 685-692, 11, 2024
© Telif hakkı IJANSER'e aittir
Araştırma Makalesi

https://as-proceeding.com/index.php/ijanser
ISSN:2980-0811

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 8, pp. 685-692, 11, 2024
Copyright © 2024 IJANSER
Research Article

Deep Learning-Based Classification of Bladder Cancer Using Vision Transformers: A Comparative Study of ViT Models

Merve Parlak Baydoğan[1*], Çağla Danacı [2], Seda Arslan Tuncer [3]

[1]*Bilgisayar Teknolojileri Bölümü / Teknik Bilimler Meslek Yüksekokulu, Fırat Üniversitesi, Türkiye*
[2]*Yazılım Mühendisliği Bölümü / Mühendislik Fakültesi, Fırat Üniversitesi, Türkiye*
[3]*Yazılım Mühendisliği Bölümü / Mühendislik Fakültesi, Fırat Üniversitesi, Türkiye*

[*]*mpbaydogan@firat.edu.tr*

(Received: 11 December 2024, Accepted: 29 December 2024)

(5th International Conference on Scientific and Academic Research ICSAR 2024, December 23-24, 2024)

ATIF/REFERENCE: Parlak Baydoğan, M., Danacı, Ç. & Arslan Tuncer, S. (2024). Deep Learning-Based Classification of Bladder Cancer Using Vision Transformers: A Comparative Study of ViT Models. *International Journal of Advanced Natural Sciences and Engineering Researches*, 8(11), 685-692.

Özet – Bladder cancer is one of the most common cancer types of the urogenital system. Each year, approximately 350,000 new cases are diagnosed, resulting in 150,000 deaths. Early detection of bladder cancer plays a critical role in determining treatment strategies and reducing mortality rates. Therefore, the development of more effective diagnostic and therapeutic approaches for bladder cancer is of significant importance. Based on its invasion of muscle tissue, bladder cancer can develop in two distinct forms: Non-Muscle-Invasive Bladder Cancer (NMIBC) and Muscle-Invasive Bladder Cancer (MIBC). NMIBC is an early-stage cancer type where the cancer is confined to the surface of the bladder without invading the muscle layer. In contrast, MIBC is a more advanced and dangerous type of cancer that invades surrounding tissues. This study proposes an autonomous system based on the deep learning Vision Transformer (ViT) model for the early detection of bladder cancer. Using an open-access, multicenter dataset, the study compares two models to classify magnetic resonance imaging (MRI) scans of bladder cancer. Following preprocessing of the bladder MRI images, model training was conducted to determine the class of the data using the ViT approach. The study evaluates the performance of two ViT models, ViT-Small Patch32 and ViT-Large Patch32, in the task of bladder cancer classification. The results of both models were assessed using the metrics of F1-Score, Recall, Precision, and Accuracy. The study findings reveal that the ViT-Large Patch32 model achieved a performance of 97% across all metrics, providing more accurate and reliable results for bladder cancer classification. The proposed study is expected to serve as a robust tool to assist experts in classifying bladder cancer and optimizing treatment processes through its supportive mechanism during the decision-making phase.ve tedavi süreçlerinin optimize edilmesinde güçlü bir araç sunması beklenmektedir.

Keywords – Vision Transfomer, Deep Learning, Bladder Cancer

I. INTRODUCTİON

Despite numerous innovations in medicine, cancer remains one of the most significant causes of mortality worldwide. Bladder cancer, one of the most common cancer types of the urogenital system, represents a major global health concern. Studies indicate that bladder cancer is diagnosed in approximately 350,000 new cases annually, resulting in around 150,000 deaths. It is observed three times more frequently in men than in women. According to global data, it ranks as the 7th most common cancer in men and the 17th in women. These statistics emphasize the profound impact of bladder cancer on global health and underline the critical importance of early diagnosis and treatment strategies for this disease [1].

Bladder cancer is influenced by both genetic and environmental factors. Smoking is recognized as the most potent and prevalent risk factor for bladder cancer. Additionally, exposure to certain carcinogenic substances associated with specific occupations, such as chemicals and jobs in the dye industry, plays a significant role in the development of this disease. Age and genetic factors also contribute substantially, as the risk increases with age, and specific genetic mutations predispose individuals to bladder cancer [2].

Various methods are employed in the diagnosis of bladder cancer, aiming to detect its presence, determine its stage, and assist in treatment planning. Common techniques include ultrasonography, urine tests, biopsy, computed tomography (CT), and magnetic resonance imaging (MRI). MRI is more widely used for evaluating local invasion of bladder cancer and for detailed examination of soft tissue structures in the pelvic region. It provides superior resolution and contrast for soft tissue imaging, enabling detailed visualization of the bladder’s internal structure. In the early stages of bladder cancer, cancer cells are typically confined to the bladder wall and often exhibit low contrast characteristics. MRI offers significant advantages in detecting tumor size, shape, location, and subtle changes in the bladder wall and surrounding tissues [3].

Bladder cancer develops in two forms based on its invasion of muscle tissue: Non-Muscle-Invasive Bladder Cancer (NMIBC) and Muscle-Invasive Bladder Cancer (MIBC) [4]. NMIBC, an earlier stage cancer type, is confined to the surface of the bladder without invading the muscle layer. If detected early, NMIBC is typically easier and more successful to treat [5]. On the other hand, MIBC is a more advanced and dangerous cancer type, spreading to deeper layers of the bladder and surrounding tissues. Once it invades the muscle layer, it can rapidly metastasize to other organs in the body [6].

This disease often goes undiagnosed in its early stages due to the absence of distinct symptoms or the mildness of symptoms. Particularly with low-risk and small lesions, an inexperienced clinician might overlook the presence of cancer during clinical evaluations [7]. As a result, cancer is often not detected until it has reached advanced stages. Early diagnosis allows for treatment before the cancer penetrates deeper layers of the bladder, significantly improving treatment outcomes. Early-stage bladder cancer can be effectively treated with surgical intervention, intravesical therapies, and other conservative approaches. Additionally, early diagnosis reduces the risk of recurrence after treatment. However, due to the vague or mild clinical manifestations of bladder cancer in its early stages, detection is often challenging [8].

To overcome these limitations in early-stage diagnosis, this study proposes an autonomous system using bladder MRI images trained with deep learning-based methods. The system aims to diagnose bladder cancer in its NMIBC stage.

Deep learning models have the ability to automatically extract features from MRI images, determining tumor size, location, and other important parameters [9]. This capability supports clinical decision-making processes, contributing to the rapid and accurate diagnosis of bladder cancer. Small lesions, tumors, or microscopic changes in MRI images can be more accurately identified through the high-resolution analyses of deep learning models [10]. In medical image analysis, convolutional neural networks (CNNs) are commonly used, along with the Vision Transformer (ViT) model, which has recently gained prominence. ViT demonstrates superior accuracy in disease detection compared to traditional methods due to its ability to comprehensively analyze details in medical images [11].

There is existing literature on computer-aided diagnosis systems for bladder cancer. For example, a study conducted by Fuster et al. proposed a CNN-based deep learning method to detect invasive cancerous regions in NMIBC patients, utilizing the VGG16 architecture for feature extraction. The study achieved an F1 score of 71.9% [12]. Another study by Khosravi et al. introduced a CNN-based method called CNN_Smoothie to classify different cancer types using histopathology images from public databases. The dataset included images of lung, breast, and bladder cancer, and the study reported accuracies of 91% for breast cancer, an average of 99% for bladder cancer, and 92% for lung cancer subtypes [13]. Yin et al. conducted a study using hematoxylin and eosin (H&E)-stained bladder tumor tissue images to classify non-invasive (stage Ta) and invasive (stage T1) bladder cancer. Six machine learning algorithms, including Adaptive Boosting (Adaboost), Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Probabilistic Neural Network (PNN), and Multilayer Perceptron (MLP), were used, achieving accuracies ranging from 91% to 96% [14]. Jansen et al. proposed a neural network-based approach to detect and grade NMIBC in H&E-stained images. Compared to the opinions of three pathologists, the proposed approach achieved accuracies of 71% for high-grade cancers and 76% for low-grade cancers [15].

This study introduces an autonomous system based on the deep learning ViT model for bladder cancer detection. The study utilizes MRI data for two types of bladder cancer, NMIBC and MIBC. The system was developed using deep learning techniques to provide fast and accurate results in clinical applications. Furthermore, it aims to minimize subjective errors, reduce physicians' workloads, and improve diagnostic accuracy. The remainder of this paper is organized as follows: Section 2 presents the dataset and methods used in the proposed system. Section 3 discusses the experimental results with comparative analysis in tables. Finally, Section 4 provides the general conclusions of the study.

II. MATERIAL VE METHODS

In this section of the study, detailed information about the materials used and the methods applied is presented.

A. *Material*

In this study, the publicly available dataset named "Bladder Cancer Classification," published on the Kaggle platform, was utilized to perform bladder cancer classification [16,17]. The dataset contains T2-weighted MRI images from a total of 279 patients. Upon examining the distribution of the 279 patients, it was found that the images originated from four different centers: 160 patients from Center C_1, 48 patients from Center C_2, 32 patients from Center C_3, and 35 patients from Center C_4. The data for each center was acquired using different MRI devices. The dataset was labeled into two classes: NMIBC and MIBC. Figure 1 presents sample images from the classes included in the dataset.

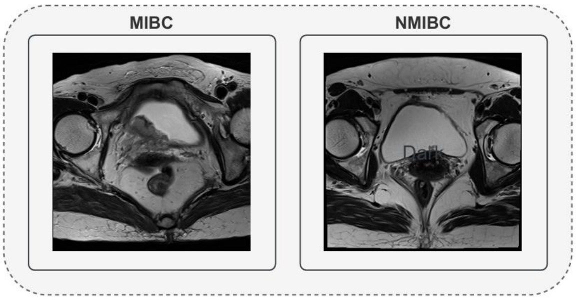

Figure 1. Sample Images of MIBC and NMIBC Classes

B. *Methods*

In this study, the classification of bladder cancer was performed using deep learning-based methods, with a focus on comparison and evaluation. Vision Transformer (ViT) models, which have recently gained popularity, were employed, specifically the ViT-Small Patch32 and ViT-Large Patch32 models. In the initial step, the data underwent preprocessing, followed by model training to determine the class of the data using the ViT method. The results of the two models were evaluated using F1-Score, Recall, Precision, and Accuracy metrics. Figure 2 presents the system diagram of the conducted study.

As observed in Figure 2, the images in the dataset were preprocessed in the first step. At this stage, lighting adjustments were made to eliminate potential contrast differences among the images. The data were resized to 384×384 dimensions and normalized to suit the input requirements of the selected models. These preprocessing steps established the foundation for faster and more accurate learning by the models.

The dataset was then split into 80% training and 20% validation data to proceed with the model training phase.

In the model training step, the Vision Transformer (ViT) architecture was employed to examine the performance of the ViT-Small Patch32 and ViT-Large Patch32 models. Both ViT models classify input images by dividing them into fixed patches of 32×32 dimensions and converting these patches into feature vectors. The primary differences between the ViT-Small Patch32 and ViT-Large Patch32 models arise from their parameter count and complexity. While the ViT-Large Patch32 model is more complex and capable of learning deeper features more effectively, the ViT-Small Patch32 model has fewer parameters, enabling faster training.

The AdamW optimization algorithm was used during model training to balance learning speed and prevent overfitting. For both models, the batch size was set to 16, and the number of epochs was set to 50. The learning rates were determined as 1.3e-3 for the ViT-Small Patch32 model and 1.2e-3 for the ViT-Large Patch32 model. While batch size and epoch count were determined through manual combinations, the learning rate was optimized using the learn_find function from the FastAI library. The ViT-Large Patch32 model exhibited higher generalization capabilities and lower computational requirements, whereas the ViT-Small Patch32 model offered a faster training process.

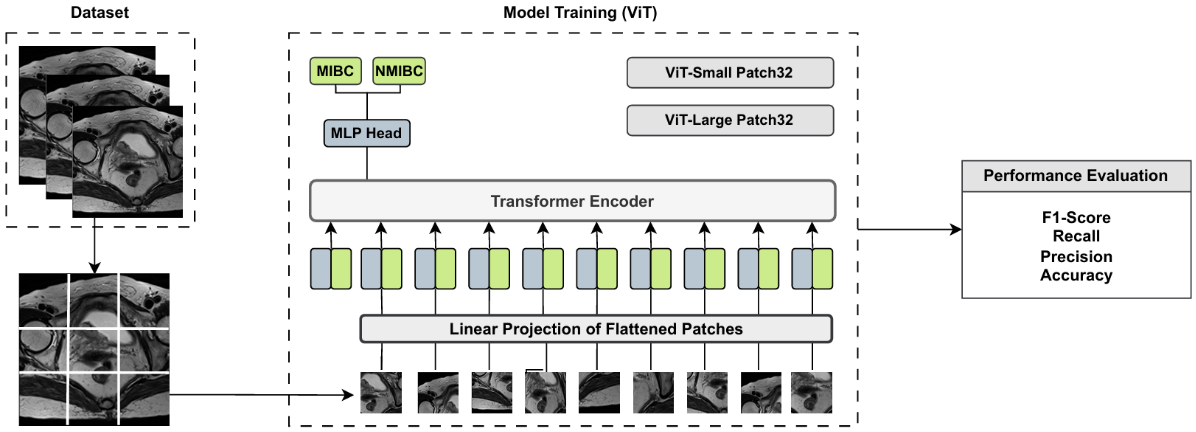

Figure 2. Pipeline of the Proposed System

The performance of both models was compared during training and validation, evaluating classification success based on training duration, accuracy, recall, precision, and F1-Score metrics. Training duration represents the total time spent by a model during the training process and varies depending on parameters such as model architecture, hardware used, dataset size, epoch count, and batch size. In this study, epoch count, batch size, and hardware parameters were kept constant for both models to examine differences in training duration arising from architectural variations. The mathematical expressions for accuracy, recall, precision, and F1-Score metrics are provided in Equations 1–4, respectively [18].

Accuracy: Accuracy represents the proportion of correct predictions made by the model out of all predictions.

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Number\ of\ Predictions} \qquad (1)$$

Precision: Precision measures the accuracy of the model's positive predictions, indicating the proportion of true positive predictions out of all positive predictions made by the model. It is particularly important when minimizing false positives is critical.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \qquad (2)$$

Recall: Recall measures how accurately the model identifies all actual positive cases. It is calculated as the proportion of true positive predictions out of all actual positive cases. Recall is a critical metric when minimizing false negatives is important.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \qquad (3)$$

F1-Score: The F1-Score is used to measure the balance between precision and recall.

$$F1 - Score = 2\frac{Precision * Recall}{Precision + Recall} \qquad (4)$$

III. EXPERIMENTAL FINDINGS AND DISCUSSION

In this study, the performance of ViT models (ViT-Large Patch32 and ViT-Small Patch32) was evaluated for bladder cancer classification. Various metrics (accuracy, precision, recall, and F1-Score) were calculated for both models to compare their classification accuracy and overall performance. The complexity matrices and learning curves illustrating the training losses for both models were analyzed alongside the performance metrics to identify the strengths and weaknesses of each model. Table 1 summarizes the performance metrics for the ViT-Large and ViT-Small models.

Table 1. Performance Evaluation Metrics Obtained from Classification Results

Metrics	**ViT-Small Patch32**	**ViT-Large Patch32**
Precision (%)	95	97
Recall (%)	95	97
F1-Score (%)	95	97
Accuracy (%)	95	97
Training Time(second)	569.80	4082.36

The ViT-Large Patch32 model outperformed the ViT-Small Patch32 model across all performance metrics. With a 97% performance in each metric, the ViT-Large Patch32 model demonstrated more accurate and reliable results for bladder cancer classification. This superior performance is attributed to the model's higher parametric capacity and ability to learn more complex features. The ViT-Small Patch32 model, on the other hand, achieved a 95% performance in the metrics, showcasing its effectiveness with lower computational requirements. While it performs slightly lower compared to the ViT-Large Patch32 model, it remains a viable option, particularly in scenarios with limited hardware resources. In terms of training time, the ViT-Large Patch32 model required approximately seven times longer than the ViT-Small Patch32 model (4082.36 seconds vs. 569.80 seconds). The longer training time for the ViT-Large model is due to its higher parametric capacity and ability to learn more complex features, which demand greater computational resources. Figure 2 presents the confusion matrices and learning curves to facilitate a detailed examination of the performance metrics alongside the models training dynamics.

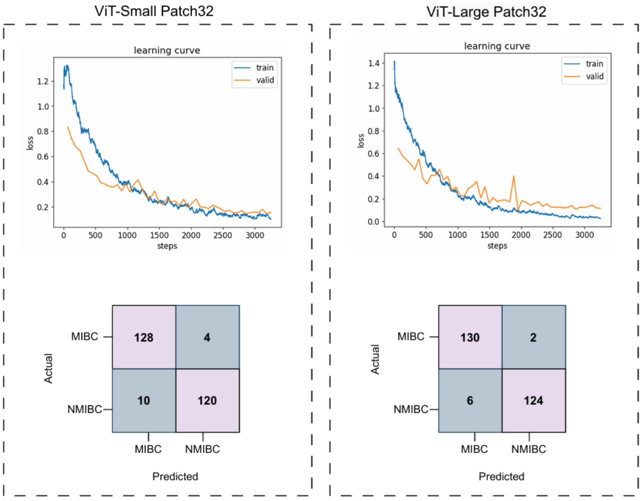

Figure 3. Confusion Matrices and Loss Graphs of Trained Models

The confusion matrix for the ViT-Large Patch32 model demonstrates its high performance in classification. Specifically, in the MIBC class, 130 out of 132 samples were correctly classified, with only 2 samples misclassified as false positives. These findings indicate that the model has a very low error rate for the MIBC class and reliable true positive predictions. When analyzing the learning curve of the ViT-Large Patch32 model, it is evident that both validation and training losses decrease rapidly. The training loss shows a steady downward trend and approaches a minimal value. Meanwhile, the validation loss generally decreases in parallel with the training loss, with some fluctuations occurring at specific intervals. These fluctuations suggest that the model does not overfit the validation data. As the training progresses, the stabilization of validation loss further improves the model's overall performance, indicating consistent performance on both validation and training data.

For the ViT-Small Patch32 model, the learning curve also shows a steady decrease. However, compared to the ViT-Large Patch32 model, it exhibits more fluctuations, indicating that the ViT-Small model has a lower learning capacity and less consistent generalization performance on validation data. These fluctuations may result from the model's limited parameter count, which restricts its ability to fully analyze complex data distributions. Nevertheless, the validation and training losses remain closely aligned, and no overfitting is observed during the training process. The lower false positive and false negative rates in the ViT-Large Patch32 model, along with fewer fluctuations in the loss curves, suggest that this model is a more reliable classification tool for the problem at hand. In contrast, the ViT-Small model can be considered a viable alternative in scenarios with resource and hardware constraints. These results clearly indicate that larger and more complex models like the ViT-Large Patch32 offer higher accuracy and consistency, albeit at the cost of longer training times and increased computational requirements.

This study evaluated the performance of ViT models, specifically ViT-Small Patch32 and ViT-Large Patch32, for bladder cancer classification. The results show that the ViT-Large model achieved higher accuracy, recall, precision, and F1-Score values compared to the ViT-Small model. The superior performance of the ViT-Large model can be attributed to its larger parametric capacity and ability to learn more complex features. The findings indicate that the ViT-Large model is more reliable for critical classification tasks. However, the training time for the ViT-Large model was approximately seven times longer than that of the ViT-Small model. This limitation in computational cost makes the ViT-Small model a more practical option in scenarios with restricted resources, highlighting the need for model selection based on specific problem requirements. The proposed study is expected to serve as a supportive system for experts in classifying bladder cancer and optimizing treatment processes during the decision-making phase.

IV. RESULTS

The study presented an effective deep learning-based solution for bladder cancer classification using Vision Transformer (ViT) models, specifically ViT-Small Patch32 and ViT-Large Patch32. Using an open-access, multicenter dataset of MRI images, the performance of both models was compared. The evaluations showed that the ViT-Large model achieved higher accuracy, precision, recall, and F1 scores, demonstrating superior performance. However, the ViT-Small model emerged as a viable alternative in resource-constrained environments due to its shorter training time and lower computational costs.

This study demonstrates that deep learning-based models can be effectively used in challenging tasks such as medical image analysis. The results indicate that it is possible to develop a robust decision-making mechanism to assist in bladder cancer classification. Considering the study's limitations, analyzing the models using only a single dataset may restrict the understanding of their generalization capabilities. Therefore, testing the models on various datasets and different medical scenarios would provide a better understanding of their generalization potential. The findings highlight the potential for artificial intelligence-based systems to see broader use in clinical applications.

REFERENCES

[1] Jemal A, Bray F, Center MM, et al. Global cancer statstcs. CA Cancer J Cln 61: 69-90, 2011

[2] Jung, I., & Messing, E. (2000). Molecular mechanisms and pathways in bladder cancer development and progression. Cancer control, 7(4), 325-334.

[3] Xu, X., Zhang, X., Tian, Q., Wang, H., Cui, L. B., Li, S., ... & Liu, Y. (2019). Quantitative identification of nonmuscle-invasive and muscle-invasive bladder carcinomas: a multiparametric MRI radiomics analysis. Journal of Magnetic Resonance Imaging, 49(5), 1489-1498.

[4] Deep learning on enhanced CT images can predict the muscular invasiveness of bladder cancer. Frontiers in Oncology.

[5] Sylvester, R. J., Van Der Meijden, A. P., Oosterlinck, W., Witjes, J. A., Bouffioux, C., Denis, L., ... & Kurth, K. (2006). Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. European urology, 49(3), 466-477.

[6] Sherif, A., Jonsson, M. N., & Wiklund, N. P. (2007). Treatment of muscle-invasive bladder cancer. Expert Review of Anticancer Therapy, 7(9), 1279-1283.

[7] Lucca, I., de Martino, M., Klatte, T., & Shariat, S. F. (2015). Novel biomarkers to predict response and prognosis in localized bladder cancer. Urologic Clinics, 42(2), 225-233.

[8] Burger, M., Catto, J. W., Dalbagni, G., Grossman, H. B., Herr, H., Karakiewicz, P., ... & Lotan, Y. (2013). Epidemiology and risk factors of urothelial bladder cancer. European urology, 63(2), 234-241.

[9] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88.

[10] Schlemper, J., Caballero, J., Hajnal, J. V., Price, A., & Rueckert, D. (2017). A deep cascade of convolutional neural networks for MR image reconstruction. In Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings 25 (pp. 647-658). Springer International Publishing.

[11] Zuo, S., Xiao, Y., Chang, X., & Wang, X. (2022). Vision transformers for dense prediction: A survey. Knowledge-Based Systems, 253, 109552.

[12] Fuster, S., Khoraminia, F., Kiraz, U., Kanwal, N., Kvikstad, V., Eftestøl, T., ... & Engan, K. (2022, June). Invasive cancerous area detection in Non-Muscle invasive bladder cancer whole slide images. In 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP) (pp. 1-5). IEEE

[13] Khosravi, P., Kazemi, E., Imielinski, M., Elemento, O., & Hajirasouliha, I. (2018). Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. EBioMedicine, 27, 317-328.

[14] Yin PN, Kc K, Wei S, Yu Q, Li R, Haake A.R, ... Cui F. Histopathological distinction of noninvasive and invasive bladder cancers using machine learning approaches. BMC medical informatics and decision making. 2020; 20: 111

[15] Jansen, I., Lucas, M., Bosschieter, J., de Boer, O. J., Meijer, S. L., van Leeuwen, T. G., ... & Savci-Heijink, C. D. (2020). Automated detection and grading of non–muscle-invasive urothelial cell carcinoma of the bladder. The American journal of pathology, 190(7), 1483-1490.

[16] Cao, K., Zou, Y., Zhang, C., Zhang, W., Zhang, J., Wang, G., ... & Huang, B. (2024). A multicenter bladder cancer MRI dataset and baseline evaluation of federated learning in clinical application. Scientific Data, 11(1), 1147.

[17] [https://www.kaggle.com/datasets/shirtgm/bladder-cancer-classification

[18] Vujović, Ž. (2021). Classification model evaluation metrics. International Journal of Advanced Computer Science and Applications, 12(6), 599-606.