

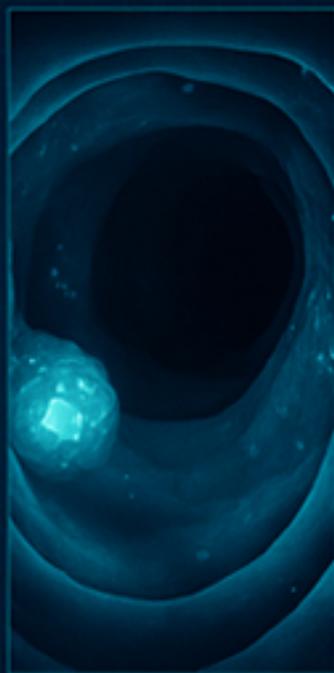
# DIAGNOSTIC ANALYSIS OF VARIOUS CANCER TYPES WITH ARTIFICIAL INTELLIGENCE



**BRAIN  
CANCER**



**BREAST  
CANCER**



**COLON  
CANCER**

**Assoc. Prof. Dr. İshak PAÇAL - Yiğitcan ÇAKMAK**



**DIAGNOSTIC ANALYSIS OF  
VARIOUS CANCER TYPES WITH  
ARTIFICIAL INTELLIGENCE**

**Assoc. Prof. Dr. İshak PAÇAL  
Yiğitcan ÇAKMAK**



***Diagnostic Analysis of Various Cancer Types with Artificial Intelligence***  
***Assoc. Prof. Dr. İshak PAÇAL, Yiğitcan ÇAKMAK***

**Editor in chief:** Berkan Balpetek

**Cover and Page Design:** Duvar Design

**Printing :** June -2025

**Publisher Certificate No:** 49837

**ISBN:** 978-625-5885-52-4

© Duvar Yayınları

853 Sokak No:13 P.10 Kemeraltı-Konak/İzmir

Tel: 0 232 484 88 68

[www.duvar yayinlari.com](http://www.duvar yayinlari.com)

[duvarkitabevi@gmail.com](mailto:duvarkitabevi@gmail.com)

## TABLE OF CONTENTS

<b>Chapter 1</b> .....	<b>5</b>
Optimizing Tumor Classification with Machine Learning and Explainable AI Tools	
<b>Chapter 2</b> .....	<b>27</b>
A Comparative Study of YOLOv9 and YOLOv10 Architectures for Tumor Detection	
<b>Chapter 3</b> .....	<b>50</b>
Evaluating YOLOv10 and YOLOv11 Variants for Efficient Polyp Detection	

# Chapter 1

---

## Optimizing Tumor Classification with Machine Learning and Explainable AI Tools

### ABSTRACT

Timely and accurate diagnosis of breast cancer is key to improving treatment protocols and the survivability of affected patients. In the present study, we investigated the performance of various machine learning (ML) algorithms for classifying tumors as either benign or malignant, using the Wisconsin Breast Cancer dataset, based on the following features: nuclear size, nuclear texture, and nuclear symmetry. We explored ten different ML models. Support Vector Machines (SVM) attained the highest mean classification accuracy of 98.60%, followed by Random Forest (97.90%); both CatBoostClassifier and K-Nearest Neighbors achieved similar mean classification accuracy of (97.20%). Additionally, advanced ensemble methods "like" XGBoost (96.50%) and LightGBM (95.80%), showed similar abilities with superior predictive accuracy. Meanwhile, traditional models, such as Logistic Regression (95.10%), and Gaussian Naive Bayes (91.61%) had respectable predictive performance, while others like Gradient Boosting (89.51%) and Decision Tree (87.41%) produced faster more interpretable results that could be beneficial in the clinic setting if time-crunched meaning they may provide benefit in the clinical scenario. The results of our study indicate that SVM and tree-based ensemble models show promise as ML diagnostic tools. SVM's and tree-based models, if utilized with explainable AI (XAI) tools such as SHAP or LIME, can provide excellent accuracy and the transparency and interpretability required for clinical acceptance. The integration of such techniques may ultimately strengthen trust in AI-assisted diagnostic tools and facilitate their adoption in real-world healthcare settings.

## **1. Introduction**

Breast cancer, which is defined as uncontrolled breast tissue proliferation resulting in lump formation, is the most diagnosed cancer in women around the world. It often manifests as a lump in the breast, but early breast cancer may be asymptomatic and remain unnoticed in many cases [1]. There are significant identified risk factors for breast cancer, including genetic mutations in select genes, family history of breast and/or ovarian cancer, age, hormone replacement therapy, and some lifestyle-related factors. In fact, genetic mutations in BRCA1 and BRCA2 would indicate a substantially higher likelihood of developing breast cancer [2]. Early detection of breast cancer is probably the most important facet of effective management of the disease [3]. Therefore, regular mammographic screening, breast self-examinations, combined with a low-fat diet, should be part of normal health care prevention routines. Maintaining a balanced diet that also promotes healthy body weight, should also assist with risk reduction. Treatment depends on the type and stage of cancer at diagnosis and may include one or more of the following: surgery, chemotherapy, radiotherapy, hormone therapy, targeted agents, etc. [4].

Progress in medical research and greater public awareness has led to advances in early diagnosis, as well as more effective treatment options. A healthy lifestyle, which includes a balanced diet, physical exercise and limited alcohol use has been shown to decrease the potential for illness [5]. The growing awareness of research and public health knowledge has allowed us to better understand breast cancer and how we can manage it. Recently, there has been a substantial increase in breast cancer incidence which is a major public health threat as it takes on many forms. Cancer is still a public health issue globally, affecting one in six people in the world [6]. Although skin cancer is viewed as the most common cancer incidence, breast cancer is the leading cause of cancer mortality in females. These are indications that we need to do more research and public health interventions due to its increasing exposure and severity [7]. Early detection has been shown

to increase treatment successfully. To better understand early detection, we need to identify good profile and get as quick a change in patient symptoms [8].

Numerous machine learning (ML) methods have been utilized to augment detection and prediction for any type of cancer [9]. Research has repeatedly supported the importance of early diagnosis on treatment effectiveness [10,11]. As a result, improving the accuracy and efficiency of breast cancer diagnostic tests are two goals. ML technologies are one of the most promising and can find weak patterns that traditional means do not recognize [12]. Models have large learning capacity with large data, faster diagnosis, and decrease complication risk [13]. The advances in technology have brought an accelerated cancer detection mechanism which allows for earlier and less complex disease detection and reduced invasive procedures. The relationship between personalized treatment and AI and ML has created an essential resource and function in this serious global health dilemma [14].

ML-based diagnostic systems can be useful to at-risk patients through screening programs that enroll them when they are identified in national diagnostic records, and hence doctors can provide well timed treatment interventions and or referrals to specialists when making a diagnosis, in the case of a life-threatening disease such as breast cancer [15]. Traditional methods of diagnostics in breast cancer screening are recognized as inadequate with some studies indicating that the historical model of diagnosis through imaging modalities, such as dynamic MRI and X-ray have a profound influence on survival, yet (are) often incompetent in their ability to restrict the identification of early-stage breast cancer, thereby allowing a rapid disease progression. Akhil et al. [16] explained that traditional techniques are limited not only because of their methodological incompetence, but because of class imbalance, poor pre-processing, and inefficient feature selection. Similarly, Law et al. [17] developed & focused on meta-heuristic algorithms Gravitational Search Optimization (GSA), Emperor Penguin Optimization (EPO) and a hybrid hGSAEPO for classification to determine the best differentiator to only select relevant features,

discard irrelevant information, whilst minimizing model complexity. Likewise, Timothy et al. [18] reclaimed early diagnosis on patients with dense breast tissue by demonstrating the model performance benefit of implementing Linear Discriminant Analysis (LDA) with the outmost consideration on the model reliability indices; accuracy, precision, & F1 score. Using the UCI dataset, Aman et al. [19] developed an ensemble model that combined the decision tree, AdaBoost, Gaussian Naive Bayes, and multilayer perceptron such that this method surpassed previous studies at an accuracy of 97.66% only further positioning the accuracy in an ensemble model method.

Moreover, Iman et al. [20] examined the effectiveness of Extreme Learning Machine (ELM) models, noting their capability to avoid overfitting and achieve classification performance comparable to SVMs in both binary and multiclass problems. Mohammad et al. [21], through a meta-analysis of 310 studies across 30 datasets, revealed that deep learning models particularly RNNs achieved accuracy rates as high as 98.58%, outperforming conventional methods. Serhat et al. [22] demonstrated the robust performance of LightGBM and ensemble models on clinical and genetic datasets, emphasizing the importance of texture and concavity-related features. Mustafa et al. [23] proposed the OSELM algorithm, which showed reliable performance on the WBCD and WDBC datasets and demonstrated strong potential as a clinical decision support tool. Shazzad et al. [24] significantly improved diagnostic accuracy up to 99.82% through SHAP-based feature selection, highlighting the importance of selecting the right features. Finally, Gani et al. [25] demonstrated that integrating dimensionality reduction techniques such as PCA and LDA led to substantial improvements in model accuracy. These findings collectively underscore the efficacy of ML approaches in diagnosing and classifying breast cancer.

Conventional diagnostic procedures for breast cancer are commonly inefficient and expensive, making treatment more difficult, in comparison, AI algorithms are frequently applied in medical data processing and cancer detection [26,27]. Stemming from big data and ML-based systems, can now shift the focus

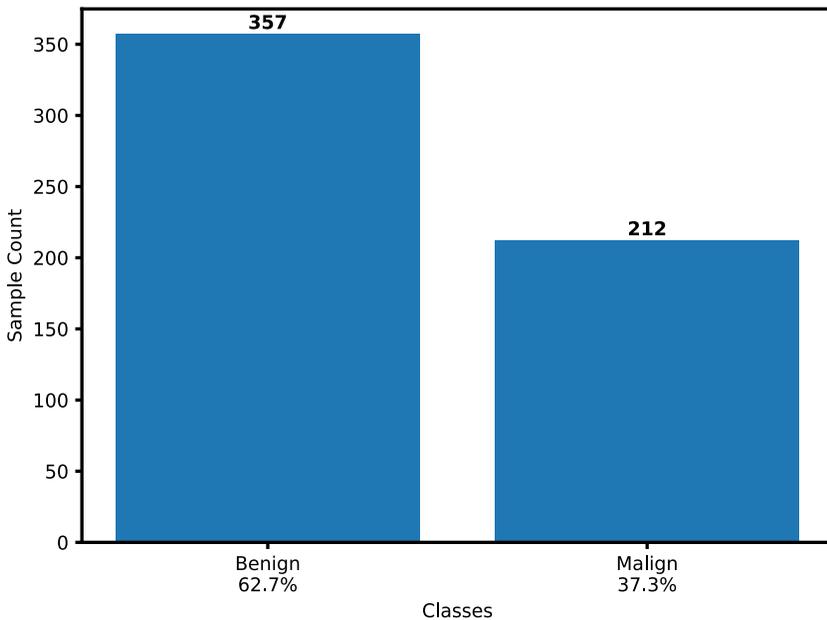
from quantity of data to quality of data [28,29]. With widespread accessibility to data, ML algorithms are regularly implemented in the healthcare sector to carry out several activities, such as predicting disease, diagnosing disease, treatment planning, cost-effectiveness, and real-time decisions. In evaluating ten distinct classification algorithms Support Vector Machine (SVM), Random Forest, CatBoostClassifier, K-Nearest Neighbors (KNN), eXtreme Gradient Boosting (XGBoost), LightGBM, Logistic Regression, Gaussian Naive Bayes, Gradient Boosting, and Decision Tree the priority is to leverage the innovation that investment in cancer research and treatment has yielded. Implementing Artificial Neural Network (ANN)-based systems further offers a path towards increasing both the effectiveness and efficiency of targeted information and communication technologies supporting healthcare innovation [30]. AI's ability to use large volumes of patient data and interpret patterns that were never foreseen by the responsible clinician, enables a more accurate detection of potential cancers and personalized treatment [31,32].

The research presents various ML algorithms for breast cancer detection. In the introduction, the importance of early diagnosis of breast cancer for a better prognosis is discussed [33]. Every algorithm is systematically evaluated using the breast cancer dataset. The methodology involved a series of activities with regards to data; the data preprocessing, model training/testing, and determining the most accurate classifier. The results show how beneficial it is for clinics to utilize ML as an early diagnosis/identification of breast cancer which allows clinicians to earlier diagnose, recommend less invasive treatments, and help with improvement in oncology. The results also showed the great promise of integrating AI and ML practice in the clinic ultimately changing healthcare to achieve more personalized healthcare [34].

## 2. Materials and Methods

### 2.1 Dataset

The present study utilizes a breast cancer dataset obtained from the University of Wisconsin Hospital, comprising 569 instances with 31 features. Each instance represents detailed information regarding a patient's tumor characteristics. The primary objective is to classify these tumors as either benign (0) or malignant (1). Among the cases, 357 are benign and 212 are malignant. The dataset provides a rich and high-dimensional structure, enabling thorough exploration of patterns associated with tumor classification. Given that breast cancer remains the second leading cause of cancer-related deaths among women, early and accurate diagnosis is of critical importance. Effective classification not only enhances treatment outcomes but also reduces unnecessary medical interventions. The numerical and percentage distribution of benign and malignant cases in the dataset is presented in Figure 1.1 [35].



**Figure 1.1** Distribution of benign and malign cases in the dataset, represented both as percentages and numerical values.

## 2.2 Characteristics and Classes of The Dataset

This research delves into key tumor features that play a pivotal role in identifying breast cancer. Each characteristic offers insight into specific, clinically significant behaviors of tumor cells, helping to form a clearer picture of their nature. A detailed breakdown of these features and their associated attributes can be found in Table 1.1, serving as a foundation for the subsequent analysis and interpretation.

**Table 1.1** Dataset features and descriptions.

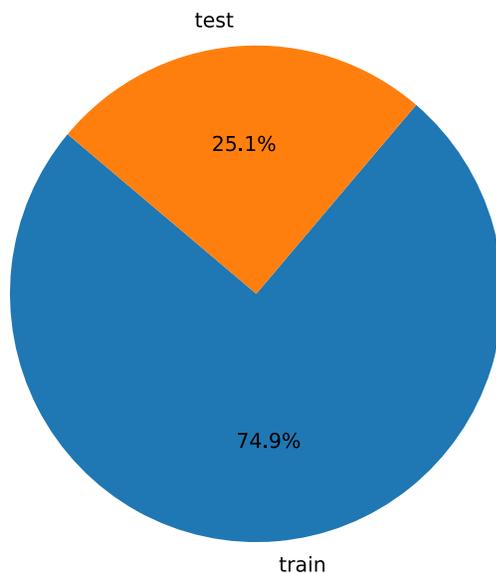
<b>Features</b>	<b>Description</b>
x.radius_mean	Mean radius of the tumor cells
x.texture_mean	Mean texture of the tumor cells
x.perimeter_mean	Mean perimeter of the tumor cells
x.area_mean	Mean area of the tumor cells
x.smoothness_mean	Mean smoothness of the tumor cells
x.compactness_mean	Mean compactness of the tumor cells
x.concavity_mean	Mean concavity of the tumor cells
x.concave_points_mean	Mean number of concave portions of the contour of the tumor cells
x.symmetry_mean	Mean symmetry of the tumor cells
x.fractal_dimension_mean	Mean "coastline approximation" of the tumor cells
x.radius_se	Standard error of the radius of the tumor cells
x.texture_se	Standard error of the texture of the tumor cells
x.perimeter_se	Standard error of the perimeter of the tumor cells
x.area_se	Standard error of the area of the tumor cells
x.smoothness_se	Standard error of the smoothness of the tumor cells
x.compactness_se	Standard error of the compactness of the tumor cells
x.concavity_se	Standard error of the concavity of the tumor cells
x.concave_points_se	Standard error of the number of concave portions of the contour of the tumor cells
x.symmetry_se	Standard error of the symmetry of the tumor cells
x.fractal_dimension_se	Standard error of the "coastline approximation" of the tumor cells
x.radius_worst	Worst (largest) radius of the tumor cells
x.texture_worst	Worst (most severe) texture of the tumor cells

x.perimeter_worst	Worst (largest) perimeter of the tumor cells
x.area_worst	Worst (largest) area of the tumor cells
x.smoothness_worst	Worst (most severe) smoothness of the tumor cells
x.compactness_worst	Worst (most severe) compactness of the tumor cells
x.concavity_worst	Worst (most severe) concavity of the tumor cells
x.concave_points_worst	Worst (most severe) number of concave portions of the contour of the tumor cells
x.symmetry_worst	Worst (most severe) symmetry of the tumor cells
x.fractal_dimension_worst	Worst (most severe) "coastline approximation" of the tumor cells
Y	target

---

### 2.3 Data preprocessing steps

To enhance the quality and efficiency of the training process, the dataset underwent several preprocessing steps. Initially, the ID column was removed, as it served merely as a unique identifier for each sample and provided no predictive value. Retaining this column could have led the model to associate specific IDs with outcomes, thereby increasing the risk of overfitting. Next, missing values in the dataset were addressed: rows containing null entries were either eliminated or imputed using statistical techniques such as mean or median replacement. Feature scaling was applied to standardize the range of all variables, which is essential for maintaining model performance, particularly in distance-based algorithms. If categorical variables were present, they were converted into numerical format using one-hot or label encoding methods. No logarithmic transformation was applied to the target variable, as it was already in a classification-appropriate format. Finally, the dataset was split into 75% for training and 25% for testing, as illustrated in Figure 1.2.



**Figure 1.2** Division of the dataset into 75% for training and 25% for testing, illustrating the allocation used for model training and evaluation.

## 2.4 Classification Models

Table 1.2 provides an overview of the machine learning algorithms employed in this study, along with brief explanations of their fundamental working principles. These algorithms were selected based on their proven effectiveness in prior classification tasks and their ability to handle complex, high-dimensional data structures. Given that the Wisconsin Breast Cancer Dataset contains 30 features describing tumor cell characteristics, each algorithm was chosen for its potential to capture meaningful patterns within such data. The selection process prioritized criteria such as predictive accuracy, computational efficiency, and the capacity to detect subtle patterns within the dataset. The strengths of each algorithm were evaluated in alignment with the overarching goal of improving breast cancer diagnosis. The aggregate descriptions in Table 1.2 highlight how each method contributes to achieving the objectives of this research.

**Table 1.2** Details of algorithms used in this study.

<b>Algorithms</b>	<b>Description</b>
<b>Support Vector Machines (SVM)</b>	SVM are supervised learning models that work well in high-dimensional spaces by finding the optimal hyperplane that separates classes. They are effective in complex classification tasks but can be computationally intensive, especially with large datasets [36].
<b>Random Forest (RF)</b>	Random Forest enhances decision trees by building multiple trees using random subsets of data and features [37]. This reduces variance and overfitting while improving generalization. Interpretability, however, can decrease with model complexity [38].
<b>CatBoostClassifier</b>	CatBoost handles categorical variables natively, which reduces the need for heavy preprocessing. Its ordered boosting method minimizes overfitting while maintaining high predictive performance, particularly in datasets with many categorical features [39].
<b>K-Nearest Neighbors (KNN)</b>	KNN is a simple [40], instance-based learning method that classifies samples based on the majority class of their k nearest neighbors. While intuitive and non-parametric, its performance can degrade with high-dimensional or imbalanced data [41].
<b>Extreme Gradient Boosting (XGBoost)</b>	XGBoost is an advanced form of gradient boosting that offers high performance through regularization, parallel processing, and second-order optimization [42]. It is particularly effective in large-scale and structured datasets [43].
<b>Light Gradient Boosting Machine (LightGBM)</b>	LightGBM uses a histogram-based method and grows trees leaf-wise rather than level-wise, which enhances both speed and memory efficiency. Though faster, it can overfit without appropriate hyperparameter tuning [44].

### **Logistic Regression (LR)**

Logistic Regression is a foundational classification algorithm that models the probability of a binary outcome using a logistic function. Despite its simplicity, it is widely used due to its interpretability and speed, though it may underperform on non-linear data .

### **Gaussian Naive Bayes (GNB)**

GNB is a probabilistic model based on Bayes' theorem, assuming feature independence and Gaussian distribution. It is especially useful for high-dimensional data and performs well even with small sample sizes .

### **Gradient Boosting (GB)**

Gradient Boosting sequentially fits new models to the residuals of prior models [45], improving predictive performance [46]. It is powerful but sensitive to overfitting unless properly regularized using techniques like shrinkage or subsampling [47]. Decision Trees classify data by splitting it based on feature values, forming a hierarchical tree [48].

### **Decision Tree (DT)**

Their strengths are simplicity and interpretability, though they are prone to overfitting without pruning or depth control [49].

---

## **3. Results and Discussion**

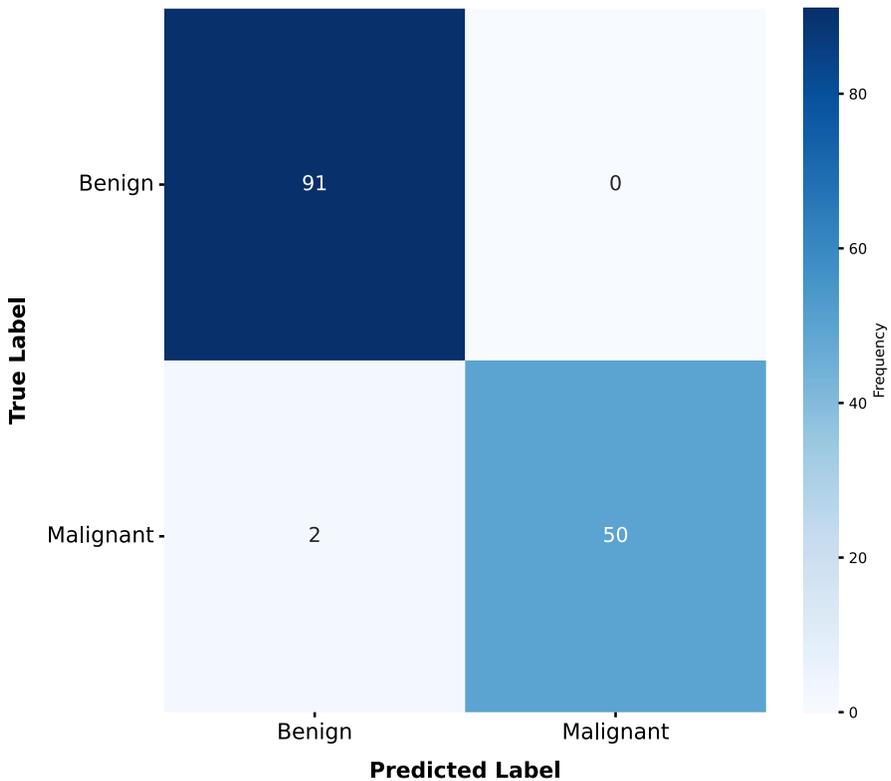
### **3.1 Support Vector Machines (SVM)**

The Support Vector Machine (SVM) classifier exhibited excellent and consistent performance in distinguishing between benign and malignant breast tumors, evidenced by an in general accuracy of 98.6%. The effectiveness of the SVM in solving binary classification problems in the medical domain has been compellingly illustrated by these findings. The SVM achieved a recall of 100% perfectly identifying all benign (non-cancerous) tumors, which resulted in no false negatives. Likewise, the SVM achieved precision of 100% for malignant locations and an F1-score of 98% for malignant locations, which created a

scenario with no false positives. The F1-scores by class were equally impressive, achieving 99% for benign and 98% for malignant tumors. Although a slight decline in recall (96%) was observed for malignant cases, the model’s aggregate performance remains highly valuable from a clinical perspective. The macro and weighted average F1-scores were 98% and 99%, respectively, indicating the classifier’s robust generalization capability even in the presence of class imbalance (91 benign and 52 malignant instances). These findings collectively emphasize that the SVM algorithm is a reliable and powerful tool for binary classification tasks, particularly in sensitive domains such as medical diagnostics. A detailed evaluation of the model’s performance is presented in Table 1.3, while a visual representation of its classification behavior is illustrated in Figure 1.3.

**Table 1.3** Observed data for the SVM algorithm.

<b>Class</b>	<b>Precision %</b>	<b>Recall %</b>	<b>F1-Score %</b>	<b>Support</b>
Benign	98.0	100.0	99.0	91
Malign	100.0	96.0	98.0	52
Macro avg	99.0	98.0	98.0	143
Weighted avg	99.0	99.0	99.0	143

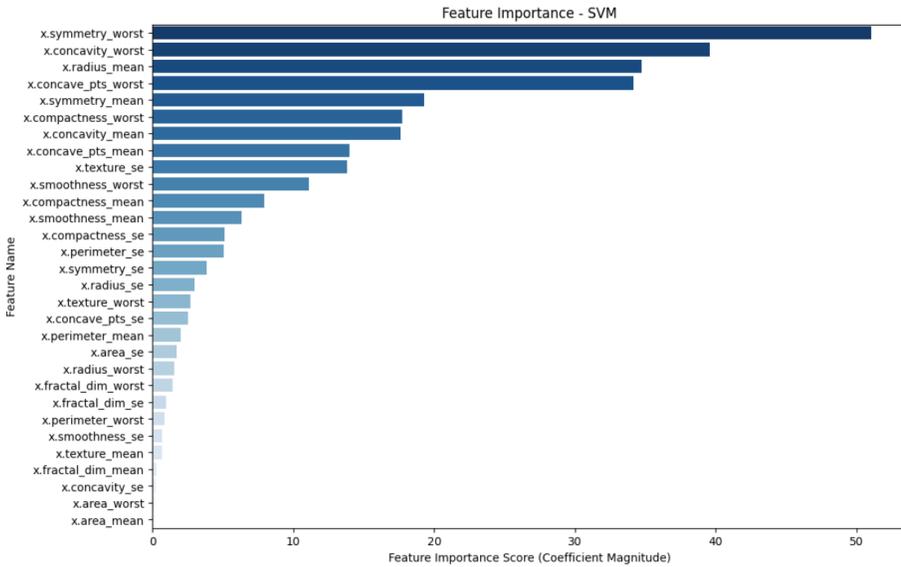


**Figure 1.3** The Confusion Matrix of SVM model used in this study.

### 3.2 Feature Importance

The generated bar chart illustrates feature importance derived from the SVM model, measured by the magnitude of the coefficients. The most influential feature is `x.symmetry_worst`, with an importance score of approximately 53 [50]. This is followed by `x.concavity_worst` (~39), `x.radius_mean` (~36), `x.concave_pts_worst` (~30), and `x.symmetry_mean` (~25), all of which significantly contribute to the model's predictions. After these top-ranking features, a sharp decline in importance scores is observed. In contrast, features such as `x.area_mean`, `x.area_worst`, `x.concavity_se`, and `x.fractal_dim_mean` exhibit near-zero importance scores, indicating a minimal impact on the model's decision-making process. These results suggest that, for this linear SVM model,

a smaller subset of features carries the majority of the predictive power. The graphical representation of these findings is provided in Figure 1.4 [51].



**Figure 1.4** Illustrates the feature importance plot for the SVM algorithm, highlighting the relative significance of each feature in the model's classification process.

#### 4. Conclusion

**Table 1.4** Experimental results of the ML algorithms.

Rank	Models	Accuracy%
1	Support Vector Machines	98.60
2	Random Forest	97.90
3	CatBoostClassifier	97.20
4	K Nearest Neighbors	97.20
5	XGradient Boosting	96.50
6	LightGBM	95.80
7	Logistic Regression	95.10
8	Gaussian Naive Bayes	91.61
9	Gradient Boosting	89.51
10	Decision Tree	87.41

This study demonstrates the potential benefits of machine learning (ML) algorithms for the early detection of life-threatening diseases such as breast cancer using the Wisconsin Breast Cancer Dataset, with the results presented in Table 1.4. Several machine learning models were applied based on the measurement of cellular nucleus features such as size, texture, and symmetry to classify tumors as benign or malignant. Support Vector Machine (SVM) performed the best aggregate with an accuracy of 98.6%, followed by Random Forest (97.9%), CatBoostClassifier, and K-Nearest Neighbors (KNN) which both had the same accuracy of 97.2%. In particular, SVM performed well with the complex decision boundaries provided in the data, which suggests great utility as a reliable machine learning tool towards reinforcing medical diagnosis tasks. Importantly, advanced ensemble methods like XGBoost (96.5%) and LightGBM (95.8%) also reported comparable accuracy. On the other hand, traditional models, like logistic regression (95.1%) and Gaussian naive bayes (91.61%), still provided satisfactory classification accuracy. Although some algorithms, like gradient boosting (89.51%) and decision trees (87.41%), demonstrated less accuracy, they also add benefits in speed and interpretability in many clinical cases. In general, this study demonstrates that ML models can aid in the early identification of critical illness, which could fast track early clinical decision making, help reduce diagnostic errors, and help support personalized treatment plans. Further, if utilized with explainable AI (XAI), the models can further support interpretability and provide increased trust and comfort level for practitioners.

## References

- [1] American Cancer Society: Breast Cancer, (n.d.).
- [2] A. Zambelli, L. Cortesi, M. Gaudio, G. Arpino, G. Bianchini, F. Caruso, S. Cinieri, G. Curigliano, L. Del Mastro, S. De Placido, A. Fabi, L. Fortunato, D. Generali, A. Gennari, S. Gori, G. Grandi, V. Guarneri, M. Klinger, L. Livi, C. Marchiò, I. Palumbo, P. Panizza, G. Pravettoni, G. Pruneri, F. Puglisi, A. Sapino, C. Tinterri, D. Turchetti, M. De Laurentiis, Parp-inhibitors in the therapeutic landscape of breast cancer patients with BRCA1 and BRCA2 pathogenic germline variants: An Italian consensus paper and critical review, *Cancer Treat Rev* 130 (2024). <https://doi.org/10.1016/j.ctrv.2024.102815>.
- [3] R.L. Siegel Mph, A.N. Giaquinto, | Ahmedin, J. Dvm, R.L. Siegel, *Cancer statistics, 2024*, (2024). <https://doi.org/10.3322/caac.21820>.
- [4] H. Ahmad, A. Ali, R. Ali, A.T. Khalil, I. Khan, M.M. Khan, M. Alorini, Preliminary insights on the mutational spectrum of BRCA1 and BRCA2 genes in Pakhtun ethnicity breast cancer patients from Khyber Pakhtunkhwa (KP), Pakistan, *Neoplasia (United States)* 51 (2024). <https://doi.org/10.1016/j.neo.2024.100989>.
- [5] T. Maramara, M.C. Matos, S. Ardila, A. Phantana-angkool, D. Henry, Minimizing Breast Cancer Risk with Diet and Exercise, *Curr Breast Cancer Rep* 16 (2024) 45–52. <https://doi.org/10.1007/s12609-024-00524-7>.
- [6] K. Schmidt, A. Thatcher, A. Grobe, P. Broussard, L. Hicks, H. Gu, L.G. Ellies, D.D. Sears, L. Kalachev, E. Kroll, The combined treatment with ketogenic diet and metformin slows tumor growth in two mouse models of triple negative breast cancer, *Transl Med Commun* 9 (2024). <https://doi.org/10.1186/s41231-024-00178-8>.
- [7] Y.J. Lee, S.D. Park, S.H. Park, H.J. Kwon, S.H. Lee, Y. Kim, J.H. Kim, Exercise affects high-fat diet-stimulated breast cancer metastasis through

- irisin secretion by altering cancer stem cell properties, *Biochem Biophys Rep* 38 (2024). <https://doi.org/10.1016/j.bbrep.2024.101684>.
- [8] S. Gori, F. De Rose, A. Ferro, A. Fabi, C. Angiolini, G. Azzarello, M. Cancian, M. Cinquini, L. Arecco, C. Aristei, D. Bernardi, L. Biganzoli, A. Cariello, L. Cortesi, E. Cretella, C. Criscitiello, U. De Giorgi, M. Carmen De Santis, G. Deledda, M. Dessena, S. Donati, A. Dri, G. Ferretti, J. Foglietta, D. Franceschini, P. Franco, A. Schirone, D. Generali, L. Gianni, S. Giordani, G. Grandi, M. Cristina Leonardi, S. Magno, L. Malorni, C. Mantoan, F. Martorana, I. Meattini, B. Meduri, L. Merlini, F. Miglietta, A. Modena, F. Nicolis, I. Palumbo, P. Panizza, F. Angela Rovera, P. Salvini, A. Santoro, M. Taffurelli, A. Toss, P. Tralongo, M. Turazza, M. Valerio, M. Verzè, P. Vici, C. Zamagni, G. Curigliano, G. Pappagallo, A. Zambelli, Follow-up of early breast cancer in a public health system: A 2024 AIGOM consensus project, *Cancer Treat Rev* 131 (2024). <https://doi.org/10.1016/j.ctrv.2024.102832>.
- [9] I. Pacal, O. Attallah, Hybrid deep learning model for automated colorectal cancer detection using local and global feature extraction, *Knowl Based Syst* 319 (2025) 113625. <https://doi.org/10.1016/J.KNOSYS.2025.113625>.
- [10] B. Ozdemir, E. Aslan, I. Pacal, Attention Enhanced InceptionNeXt Based Hybrid Deep Learning Model for Lung Cancer Detection, *IEEE Access* (2025). <https://doi.org/10.1109/ACCESS.2025.3539122>.
- [11] I. Aruk, I. Pacal, A.N. Toprak, A novel hybrid ConvNeXt-based approach for enhanced skin lesion classification, *Expert Syst Appl* 283 (2025) 127721. <https://doi.org/10.1016/J.ESWA.2025.127721>.
- [12] H. Duan, Y. Zhang, H. Qiu, X. Fu, C. Liu, X. Zang, A. Xu, Z. Wu, X. Li, Q. Zhang, Z. Zhang, F. Cui, Machine learning-based prediction model for distant metastasis of breast cancer, *Comput Biol Med* 169 (2024). <https://doi.org/10.1016/j.combiomed.2024.107943>.

- [13] J.F. AlSamhori, A.R.F. AlSamhori, L.A. Duncan, A. Qalajo, H.F. Alshahwan, M. Al-abbadi, M. Al Soudi, R. Zakraoui, A.F. AlSamhori, S.A. Alryalat, A.J. Nashwan, Artificial intelligence for breast cancer: Implications for diagnosis and management, *Journal of Medicine, Surgery, and Public Health* 3 (2024) 100120. <https://doi.org/10.1016/j.glmedi.2024.100120>.
- [14] S.M. Malakouti, M.B. Menhaj, A.A. Suratgar, ML: Early Breast Cancer Diagnosis, *Current Problems in Cancer: Case Reports* 13 (2024). <https://doi.org/10.1016/j.cpcpr.2024.100278>.
- [15] I. Pacal, O. Attallah, InceptionNeXt-Transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis, *Biomed Signal Process Control* 110 (2025) 108116. <https://doi.org/10.1016/J.BSPC.2025.108116>.
- [16] A.K. Das, S.K. Biswas, A. Mandal, A. Bhattacharya, S. Sanyal, Machine Learning based Intelligent System for Breast Cancer Prediction (MLISBCP), *Expert Syst Appl* 242 (2024). <https://doi.org/10.1016/j.eswa.2023.122673>.
- [17] L.K. Singh, M. Khanna, R. Singh, An enhanced soft-computing based strategy for efficient feature selection for timely breast cancer prediction: Wisconsin Diagnostic Breast Cancer dataset case, *Multimed Tools Appl* 83 (2024) 76607–76672. <https://doi.org/10.1007/s11042-024-18473-9>.
- [18] T. Stockton, B. Peddle, A. Gaulin, E. Wiechert, W. Lu, Comparative Analysis of Data Preprocessing Methods in Machine Learning for Breast Cancer Classification, (2024) 268–279. [https://doi.org/10.1007/978-3-031-57870-0\\_24](https://doi.org/10.1007/978-3-031-57870-0_24).
- [19] A. Sharma, D. Goyal, R. Mohana, An ensemble learning-based framework for breast cancer prediction, *Decision Analytics Journal* 10 (2024). <https://doi.org/10.1016/j.dajour.2023.100372>.

- [20] I. kadhim ajlan, H. Murad, A.A. Salim, A. fadhil bin yousif, Extreme Learning machine algorithm for breast Cancer diagnosis, *Multimed Tools Appl* (2024). <https://doi.org/10.1007/s11042-024-19515-y>.
- [21] M.R. Darbandi, M. Darbandi, S. Darbandi, I. Bado, M. Hadizadeh, H.R. Khorram Khorshid, Artificial intelligence breakthroughs in pioneering early diagnosis and precision treatment of breast cancer: A multimethod study, *Eur J Cancer* 209 (2024). <https://doi.org/10.1016/j.ejca.2024.114227>.
- [22] S. Ünalán, O. Günay, I. Akkurt, K. Gunoglu, H.O. Tekin, A comparative study on breast cancer classification with stratified shuffle split and K-fold cross validation via ensembled machine learning, *J Radiat Res Appl Sci* 17 (2024) 101080. <https://doi.org/10.1016/j.jrras.2024.101080>.
- [23] M.A.A. Albadr, F.T. AL-Dhief, L. Man, A. Arram, A.H. Abbas, R.Z. Homod, Online sequential extreme learning machine approach for breast cancer diagnosis, *Neural Comput Appl* 36 (2024) 10413–10429. <https://doi.org/10.1007/s00521-024-09617-x>.
- [24] M.S.H. Shaon, T. Karim, M.S. Shakil, M.Z. Hasan, A comparative study of machine learning models with LASSO and SHAP feature selection for breast cancer prediction, *Healthcare Analytics* 6 (2024). <https://doi.org/10.1016/j.health.2024.100353>.
- [25] G. Esen, A. Altaibek, J. Amankulov, B. Matkerim, M. Nurtas, Enhancing Breast Cancer Detection with Dimensionality Reduction Techniques: A Study Using PCA and LDA on Wisconsin Breast Cancer Data, in: *Procedia Comput Sci*, Elsevier B.V., 2024: pp. 414–421. <https://doi.org/10.1016/j.procs.2024.11.128>.
- [26] I. Pacal, B. Ozdemir, J. Zeynalov, H. Gasimov, N. Pacal, A novel CNN-ViT-based deep learning model for early skin cancer diagnosis, *Biomed Signal Process Control* 104 (2025). <https://doi.org/10.1016/j.bspc.2025.107627>.

- [27] B. Bayram, I. Kunduracioglu, S. Ince, I. Pacal, A systematic review of deep learning in MRI-based cerebral vascular occlusion-based brain diseases, *Neuroscience* 568 (2025) 76–94. <https://doi.org/10.1016/j.neuroscience.2025.01.020>.
- [28] B. Ozdemir, E. Aslan, Attention Enhanced InceptionNeXt-Based Hybrid Deep Learning Model for Lung Cancer Detection, (n.d.). <https://doi.org/10.1109/ACCESS.2025.3539122>.
- [29] B. Ozdemir, I. Pacal, A robust deep learning framework for multiclass skin cancer classification, *Sci Rep* 15 (2025). <https://doi.org/10.1038/s41598-025-89230-7>.
- [30] T. Mahmood, T. Saba, A. Rehman, F.S. Alamri, Harnessing the power of radiomics and deep learning for improved breast cancer diagnosis with multiparametric breast mammography, *Expert Syst Appl* 249 (2024). <https://doi.org/10.1016/j.eswa.2024.123747>.
- [31] Y. Cakmak, S. Safak, M.A. Bayram, I. Pacal, Comprehensive Evaluation of Machine Learning and ANN Models for Breast Cancer Detection, *Computer and Decision Making: An International Journal* 1 (2024) 84–102. <https://doi.org/10.59543/COMDEM.V1I.10349>.
- [32] Y. Cakmak, I. Pacal, Enhancing Breast Cancer Diagnosis: A Comparative Evaluation of Machine Learning Algorithms Using the Wisconsin Dataset, *Journal of Operations Intelligence* 3 (2025) 175–196. <https://doi.org/10.31181/JOPI31202539>.
- [33] S. Benghazouani, S. Nouh, A. Zakrani, Optimizing breast cancer diagnosis: Harnessing the power of nature-inspired metaheuristics for feature selection with soft voting classifiers, *International Journal of Cognitive Computing in Engineering* 6 (2025) 1–20. <https://doi.org/10.1016/j.ijcce.2024.09.005>.
- [34] H.E. Saroğlu, I. Shayea, B. Saoud, M.H. Azmi, A.A. El-Saleh, S.A. Saad, M. Alnakhli, Machine learning, IoT and 5G technologies for breast cancer

- studies: A review, *Alexandria Engineering Journal* 89 (2024) 210–223.  
<https://doi.org/10.1016/j.aej.2024.01.043>.
- [35] Breast Cancer Wisconsin Diagnostic Dataset, (n.d.).  
<https://www.kaggle.com/datasets/utkarshx27/breast-cancer-wisconsin-diagnostic-dataset> (accessed July 17, 2024).
- [36] W.S. Noble, What is a support vector machine?, 2006.  
<http://www.nature.com/naturebiotechnology>.
- [37] G. Biau, E. Scornet, A Random Forest Guided Tour, (2015).  
<http://arxiv.org/abs/1511.05741>.
- [38] L. Breiman, Random Forests, 2001.
- [39] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin, CatBoost: unbiased boosting with categorical features, (n.d.).  
<https://github.com/catboost/catboost> (accessed March 4, 2025).
- [40] G. Guo, H. Wang, D.A. Bell, S. Profile, Y. Bi, D. Bell, K. Greer, KNN Model-Based Approach in Classification, 2004.  
<https://www.researchgate.net/publication/2948052>.
- [41] G. Guo, H. Wang, D. Bell, Y. Bi, K. Greer, KNN Model-Based Approach in Classification, 2003.
- [42] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 13-17-August-2016* (2016) 785–794. <https://doi.org/10.1145/2939672.2939785>.
- [43] Z. Arif Ali, Z. H. Abduljabbar, H. A. Tahir, A. Bibo Sallow, S.M. Almufti, eXtreme Gradient Boosting Algorithm with Machine Learning: a Review, *Academic Journal of Nawroz University* 12 (2023) 320–334.  
<https://doi.org/10.25007/ajnu.v12n2a1612>.
- [44] S.O. Olatunji, M.A.A. Khan, F. Alanazi, R. Yaanallah, S. Alghamdi, R. Alshammari, F. Alkhatim, M. Farooqui, M.I.B. Ahmed, Machine Learning-Based Models for the Preemptive Diagnosis of Sickle Cell

- Anemia Using Clinical Data, (2024) 101–112.  
[https://doi.org/10.1007/978-3-031-67547-8\\_9](https://doi.org/10.1007/978-3-031-67547-8_9).
- [45] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, *Front Neurorobot* 7 (2013). <https://doi.org/10.3389/fnbot.2013.00021>.
- [46] Z. He, D. Lin, T. Lau, M. Wu, Gradient Boosting Machine: A Survey Point Zero One Technology, (2019).
- [47] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, *Front Neurorobot* 7 (2013) 63623.  
<https://doi.org/10.3389/FNBOT.2013.00021/BIBTEX>.
- [48] J.R. Quinlan, *Induction of Decision Trees*, 1986.
- [49] C. Kingsford, S.L. Salzberg, What are decision trees?, *Nat Biotechnol* 26 (2008) 1011. <https://doi.org/10.1038/NBT0908-1011>.
- [50] A. Got, D. Zouache, A. Moussaoui, L. Abualigah, A. Alsayat, Improved Manta Ray Foraging Optimizer-based SVM for Feature Selection Problems: A Medical Case Study, *J Bionic Eng* 21 (2024) 409–425.  
<https://doi.org/10.1007/s42235-023-00436-9>.
- [51] J. Azimjonov, T. Kim, Designing accurate lightweight intrusion detection systems for IoT networks using fine-tuned linear SVM and feature selectors, *Comput Secur* 137 (2024).  
<https://doi.org/10.1016/j.cose.2023.103598>.

# Chapter 2

---

## A Comparative Study of YOLOv9 and YOLOv10 Architectures for Tumor Detection

### Abstract

The human brain, a highly complex organ, is critical for all bodily functions; neoplastic lesions therein, such as brain tumors, represent a significant source of global morbidity and mortality with challenging prognoses. Magnetic Resonance Imaging (MRI) is the established gold standard for non-invasive brain tumor detection and characterization. Recently, artificial intelligence (AI), particularly deep learning (DL) algorithms like Convolutional Neural Networks (CNNs) and "You Only Look Once" (YOLO) models, has emerged to augment MRI analysis, aiming to improve diagnostic accuracy and efficiency. This study presents a comparative performance analysis of contemporary YOLOv9 and YOLOv10 models for automated brain tumor detection from MRI scans. Evaluation encompassed standard metrics (precision, recall, mAP50, mAP50-95) and considered model architecture, parameter count, and computational cost (FLOPs). Our findings indicate that while larger model variants generally yield higher precision and recall, performance gains exhibit diminishing returns beyond certain architectural complexity. Notably, the YOLOv10x model demonstrated a superior balance between high detection accuracy excelling in precision and mAP50-95 metrics and model efficiency. Furthermore, cost-performance trade-off analysis revealed that computationally intensive models (e.g., YOLOv9e) lacked proportional performance advantages over streamlined architectures (e.g., YOLOv9m, YOLOv10x). This suggests carefully selected, smaller, optimized models can achieve comparable, or more favorable, in total performance when factoring in efficiency. Significantly, this study demonstrates that appropriately optimized YOLO-based architectures can achieve performance levels in brain tumor detection that show considerable promise for clinical applicability.

## 1. Introduction

Brain tumors, resulting from abnormal cell proliferation in the brain, represent a significant group of neurological diseases that can severely impact both physiological and cognitive functions [1,2]. Tumors are generally classified as benign or malignant, with early diagnosis being a critical factor for treatment success. Radiologists largely rely on magnetic resonance imaging (MRI) for brain tumor diagnosis, which provides information on spatial localization, size, and type of tumor [3]. MRI remains the preferred imaging modality for brain tumor diagnosis due to its ability to provide high-quality imaging of brain tissue [4,5].

However, despite advanced imaging technologies, the manual analysis of brain tumors is a time-consuming, laborious, inherently subjective, and error-prone process. Difficulties in interpretation can arise, particularly due to irregular tumor shapes, indistinct and ill-defined edges, and challenges in distinguishing tumor tissue from surrounding healthy tissue [6,7]. With the advent of large datasets, the limitations of manual analysis have become even more apparent, thereby increasing expectations for automated detection and classification systems and intensifying interest in deep learning strategies [8].

In recent years, artificial intelligence has been successfully applied to various types of cancer and other fields, and it is increasingly becoming a critical area of importance [9–14]. Rapid advancements in deep learning have revealed the potential of object detection architectures, such as "You Only Look Once" (YOLO), for use in medical image analysis. While previous YOLO versions and other deep learning models have shown promising results in various medical imaging tasks, comprehensive comparative analyses of the latest generation architectures, especially in complex scenarios like multi-class brain tumor detection and for real-world clinical applications, are limited. Previous work has shown the potential for YOLO-based algorithms to succeed in medical imaging tasks. For instance, Kang et al. created the BGF-YOLO model based on YOLOv8, which they improved to detect brain tumors significantly better than before. They stated that YOLOv9 was successful in detecting and localizing

tumors with greater accuracy. To improve the model's focus on significant features, the BGF-YOLO model introduced Bi-level Routing Attention, Generalized Feature Pyramid Networks (GFPN) and a fourth detection head. They successfully improved their model over the YOLOv8x model and reported a difference of 4.7% on mAP on the Br35H dataset [15]. Elhanashi et al. explored brain tumors detection and localization using YOLOv9. They reported better accuracy and finding speed compared to previous models. YOLOv9 made significant improvement in tumor detection, and localization when trained on a brain MRI dataset [16].

Pandey and Bhandari combined a YOLOv5-based model with transfer learning for early-stage brain tumor detection. The results from the testing showed the first proper application for clinical settings. YOLOv5's accuracy was superior to existing alternative deep neural networks such as AlexNet, ResNet-50, GoogleNet, MobileNet, VGG-16, YOLOv3 and YOLOv4. When this study performed their experiments on the BraTS21 dataset, YOLOv5 exhibited an mAP@0.5 score of 94.7%, which was then improved with morphological filtering to 97.2%, confirming the success of this model to detect brain tumors [17]. Du and colleagues developed a new model called YOLO-CPC; an extension of YOLOv7, that was used to detect and to segment breast tumors. With the incorporation of CBAM attention mechanism, PConv, and coordinate convolutions, YOLO-CPC improved performance by reaching identification and segmentation accuracy, recall, and mean accuracy with rates of 97.01%, 97.98%, and 90.78% respectively and exceeded the performance of Faster R-CNN, YOLOv3, YOLOv5, and YOLOv6. The advances in accuracy promote the potential use of YOLO-CPC for other medical imaging applications [18].

Addressing the gap, this study aims to comprehensively compare the capabilities of state-of-the-art YOLOv9 and YOLOv10 variants in automatically detecting and classifying brain tumors on an MRI dataset comprising four distinct classes: Glioma, Meningioma, Pituitary, and No Tumor. The research seeks to reveal the strengths and weaknesses of these two leading models, particularly by

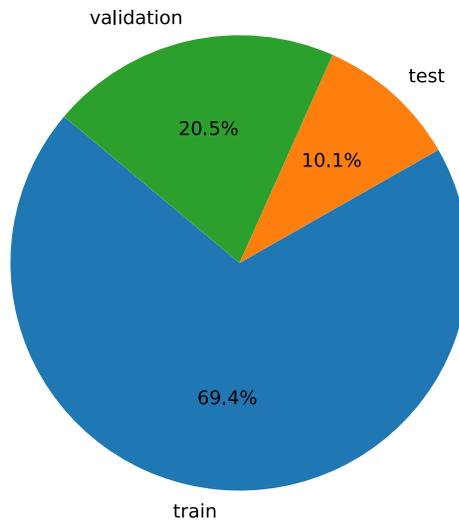
incorporating the "No Tumor" class often overlooked in the literature to holistically evaluate model performance on both positive and negative examples [19,20]. The anticipated findings are expected to provide significant insights into the efficacy of YOLO-based models for brain tumor detection and classification within medical image analysis, and to contribute to the development of reliable, rapid, and accurate AI-driven decision support systems that can bolster clinical workflows [21,22] .

In this study, contemporary deep learning methodologies, specifically the state-of-the-art You Only Look Once (YOLO) versions YOLOv9 and YOLOv10, were rigorously applied for the automated detection and classification of brain tumors from Magnetic Resonance Imaging (MRI) scans. Recognizing the rapid evolution and distinct architectural advancements within the YOLO lineage, this research undertook a granular analysis of these two leading models. The diverse architectural configurations of these models, systematically scaled from compact to larger versions (e.g., -S, -M, -L, -X variants), were a central focus. Each configuration was meticulously evaluated based on its parameter count, computational footprint (FLOPs), and a comprehensive suite of performance metrics, including mean Average Precision (mAP) at various Intersection over Union (IoU) thresholds (e.g., mAP50, mAP50-95), precision, recall, and F1-score. This detailed analysis was conducted on a specific, clinically relevant MRI brain tumor dataset to ensure the findings are grounded in a practical application context. The comprehensive investigation aimed not only to benchmark the raw detection and classification accuracy of each model variant but also to thoroughly elucidate their model-specific characteristics. These characteristics include architectural nuances, inference speed, and the critical trade-offs between detection efficacy and computational resource utilization. Ultimately, this multifaceted evaluation seeks to provide critical insights into the operational strengths and limitations of YOLOv9 and YOLOv10 variants, thereby offering evidence-based guidance for their optimal deployment in AI-assisted brain tumor diagnostic workflows.

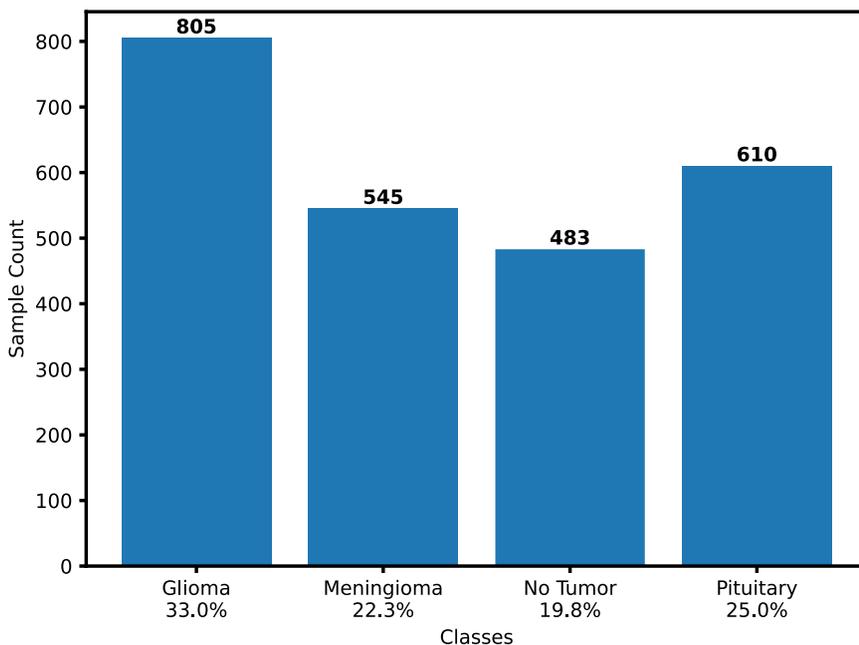
## 2. Materials and Methods

### 2.1 Dataset

Our dataset is taken from the "Labeled MRI Brain Tumor Dataset" intended for brain tumor detection and classification purposes. The source dataset contains 2,443 labeled MRI images, representing 4 different types of brain tumors, and has been split into three structures training, testing, and validation that each play a part in the construction of the model. The training structure is comprised of many samples that allows the model to learn from, while the test and validation structures evaluates the on the whole performance and generalization ability of the model. The distribution of the classes and what each class contains is referenced in detail in Figures 2.1 and 2.2. To evaluate the model's ability to classify brain tumors accurately, it is also pertinent to train the model to evaluate non-tumor instances, as understanding this relationship allows for the model to be robust. This form of organization aids in a more complete and realistic evaluation of the model and improves the model's ability to generalize to unique data points [23].

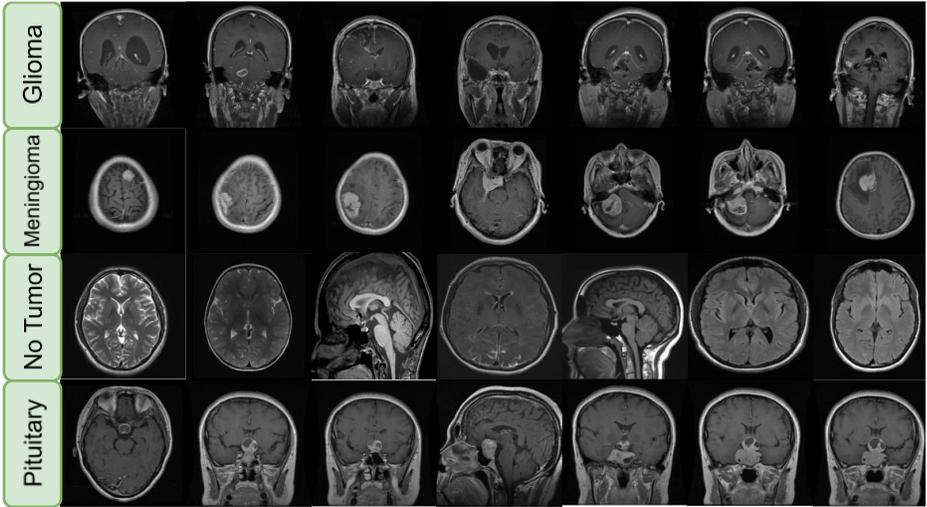


**Figure 2.1** Details of dataset division into 3 parts as train test and validation.



**Figure 2.2** Number of images per class.

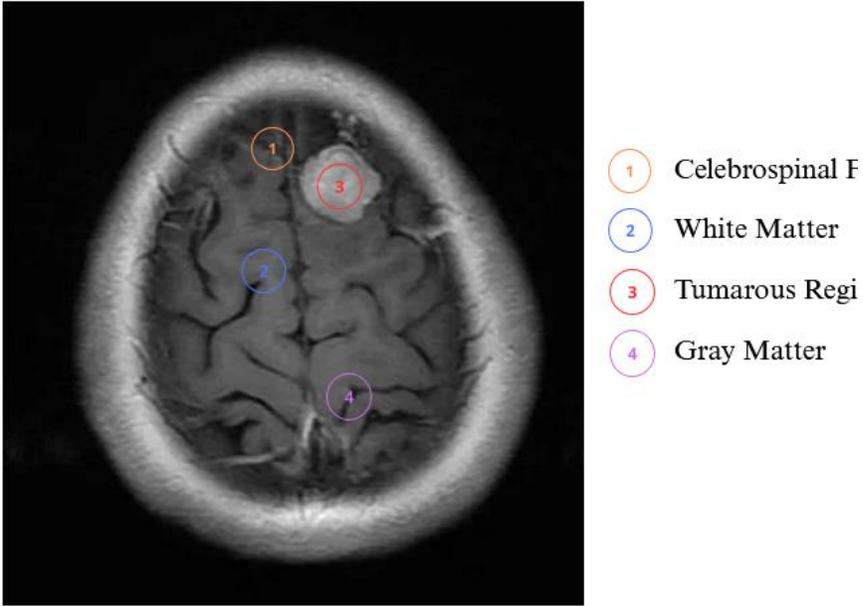
Seven sample images from each class in the dataset is presented in Figure 2.3. This figure displays seven distinct images representing each class, offering a clearer insight into the visual diversity within each category. The selected images highlights the unique features necessary for the model to recognize and differentiate between the various classes. These visual examples have been carefully chosen to emphasize both the variety of appearances within each class and the importance of such diversity in the model's learning process.



**Figure 2.3** Some randomly selected sample images of dataset classes.

## 2.2 Brain Tumors

The uncontrolled proliferation of living and dead cells within the brain is fundamentally responsible for the development of brain tumors [24–26]. The growth rate of these tumors varies from one individual to another and is also influenced by the specific location of the tumor within the brain. Based on this distinction, brain tumors can be categorized into two main types: primary and secondary. A primary brain tumor arises directly within the brain, while a secondary brain tumor originates elsewhere in the body and spreads to the brain [27–29]. The human brain is composed of three essential tissue structures: grey matter, white matter, and cerebrospinal fluid (CSF), as depicted in Figure 2. 4. The process of scanning different MR images relies heavily on the characteristics of these three elements within the brain.



**Figure 2.4** Illustrates the three main tissue structures of the human brain: grey matter, white matter, and cerebrospinal fluid (CSF), which are essential for MRI imaging analysis.

### 2.3 YOLOv9 and YOLOv10 Models

Object detection is extremely important, especially in medical imaging and real-time applications, where time and accuracy are crucial. YOLO is a commonly used deep learning method in these scenarios due to its speed and accuracy of object detection. The YOLO models detect objects by splitting an image into an  $n \times n$  grid of cells, where each cell predicts the class and location of any object within the cell. The entire process is a single prediction of the whole image which makes YOLO faster and more efficient than other detectors that require multiple passes.

YOLOv9 and YOLOv10 models were used in this research for important medical imaging applications, such as brain tumor detection, in various MODELS. Each YOLO model is offered in a configured size optimized for different computational demands, which positions them in a place of usability

across many devices with a level of speed and accuracy. The YOLOv9 and YOLOv10 models can be fine-tuned, if we also consider the different computational power and parameter numbers each model is optimized for. Usually, the bigger the model the more accurate it is; however, the more accurate it is, the cost necessarily goes up.

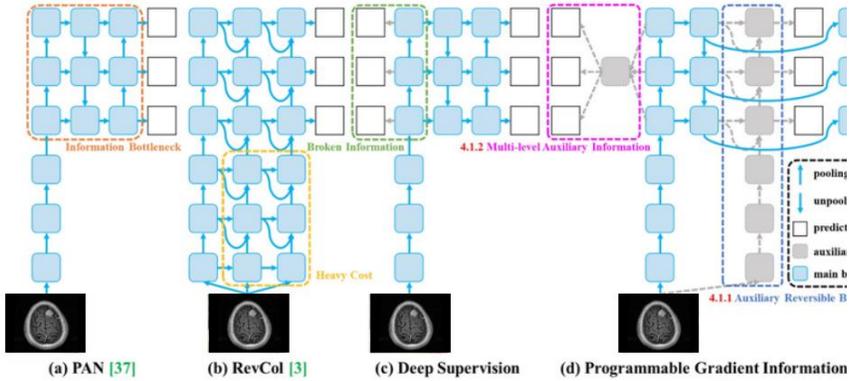
The YOLOv9 series provides several model variants that cater to various use cases. For example, the YOLOv9t fits the requirement for movement and/or computer systems that are limited in computing power, thus offering fewer parameters and enabling more rapid predictions. The YOLOv9e is larger and more capable of achieving higher accuracy; however, it demands more computational power. Other YOLOv9 variants provide various trade-offs between the YOLOv9e and YOLOv9t for systems with different capabilities.

In a similar manner, the YOLOv10 series offers a better trade-off between speed and accuracy. The YOLOv10n is smaller and has less computational requirements than the YOLOv10x which improves accuracy with a higher number of parameters; however, it requires more computing capabilities. The YOLOv10 models provide decent versatility offering different trade-offs to address the performance and efficiency needs of the application.

The computational costs of these models originate primarily from the number of layers, number of parameters, and the computational costs. The larger YOLOv9 and YOLOv10 models typically provides better accuracy, but where we run them would make it difficult to run them on a limited system. Therefore, finding the right model for the application would require deciding on an optimal level of speed and accuracy between the models.

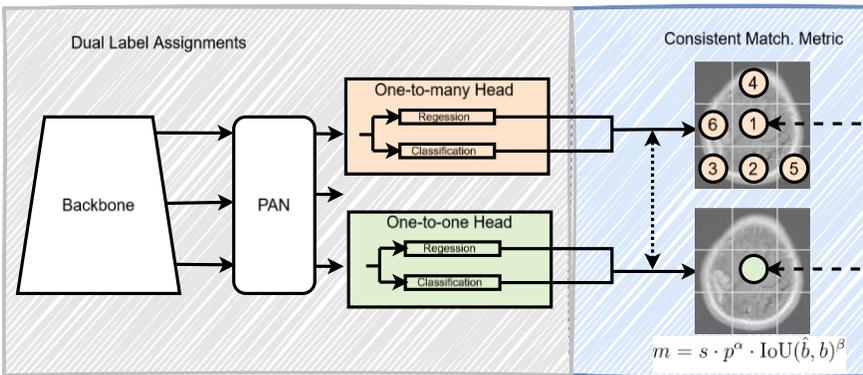
As a aggregate, the YOLOv9 and YOLOv10 family of models provide significant improvements in object detection tasks, where each variant might provide a different advantage depending upon the use case. Computational costs, while intuitive to calculate, may not have a direct relationship with accuracy and speed such that it is critical to consider the accuracy, speeds, definitions of speed,

etc. for each model. The architecture of the models is shown in Figure 2.5 and 2.6 [30,31].



**Figure 2.5** The PGI method and related network architectures are as follows:

(a) Path Aggregation Network (PAN), (b) Reversible Columns (RevCol), (c) traditional deep supervision, and (d) our proposed Programmable Gradient Information (PGI). PGI consists of three main components: (1) the primary branch, which is the architecture used during inference, (2) the auxiliary reversible branch, which generates reliable gradients to support the primary branch during backward propagation, and (3) multi-level auxiliary information, which enables the main branch to adaptively learn from various levels of semantic data [32,33].



**Figure 2.6** Consistent dual assignments for NMS-free training.

## 2.4 Performance measures

To assess the effectiveness of a model or algorithm, a range of evaluation metrics is commonly utilized to determine its reliability and applicability. In the field of medical image analysis, these performance indicators are particularly crucial for evaluating diagnostic accuracy and consistency. The key metrics used in this study are summarized in Table 2.1. These measures not only reflect the model's predictive capabilities but also offer insight into the validity of the broadly speaking framework and the specific components it employs [34].

**Table 2.1** Evaluation metrics.

Parameter	Formula	Definition/functionality
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Represents the proportion of all correct predictions both positive and negative relative to the total number of instances.
Sensitivity/Recall	$\frac{TP}{TP + FN}$	Indicates the model's ability to correctly identify actual positive cases among all true positives.
Precision	$\frac{TP}{TP + FP}$	Reflects the ratio of true positive predictions to the total number of instances that were predicted as positive.
Specificity	$\frac{TN}{TN + FP}$	Measures the proportion of actual negative cases that were correctly predicted by the model.
F1	$2 \times \frac{Precision \times Recall}{Precision + Recall}$	A harmonic mean of precision and recall, providing a balanced metric for assessing overall model effectiveness.
mAP	$\frac{1}{n} \sum_{k=1}^n AP_k$	Used in object detection tasks, it represents the average precision across all classes or object categories.

## 2.5 Data Augmentation

In this study, no custom data augmentation methods were applied manually. However, the YOLOv9 and YOLOv10 models used in the experiments inherently

employ a variety of default augmentation strategies as part of their standard training configuration. These augmentations were automatically applied during training and contributed to the models' performance by increasing the diversity of visual input. The augmentations include both color and geometric transformations. Color augmentations are performed in the HSV color space, where the hue is adjusted by  $\pm 0.015$ , saturation by  $\pm 0.7$ , and value (brightness) by  $\pm 0.4$ . These changes help the model handle variations in lighting and color conditions. Geometric augmentations consist of random translations up to 10% along both axes (translate: 0.1), random scaling (scale: 0.5), and horizontal flipping with a probability of 50% (fliplr: 0.5), enabling the model to learn from different orientations and spatial arrangements [35].

In addition to these, the Mosaic augmentation technique is applied throughout the training process (mosaic: 1.0), merging four different images into one to present objects in varying contexts. This augmentation is turned off during the last 10 epochs (close\_mosaic: 10) to allow more stable learning in the final training phase. The Rand Augment policy is also used as an Auto Augment strategy, randomly selecting and applying augmentation operations to further diversify the training data. Moreover, a Random Erasing technique with a 40% probability (erasing: 0.4) is employed to simulate partial occlusions by masking random rectangular areas within images. Collectively, these default augmentation strategies enhance the model's ability to generalize by exposing it to a wide range of real-world variations, even though no manual augmentation was conducted [36].

## **2.6 Experimental design**

All experiments in this study were carried out on a Linux-based system running Ubuntu 22.04. The machine was equipped with an Intel Core i5-13600K processor, 32 GB of DDR5 RAM, and an NVIDIA RTX 3090 GPU. PyTorch served as the primary deep learning framework, leveraging NVIDIA CUDA for GPU acceleration. To ensure consistency and fairness across all models, the same computational environment and hyperparameters were used throughout both the training and testing phases.

### 3. Results and Discussion

Table 2.2 summarizes the performance of the YOLOv9 and YOLOv10 models in terms of precision, recall, mAP50, and mAP50-95. The given metrics make it easy to compare the aggregate performance of the various models and their ability to detect tumors across classes. The results show that there are large differences in performance depending on the model architecture, layer arrangement and total number of parameters.

**Table 2.2.** The results observed as a result of the experiments in this study

Model	Layer	Params (M)	FLOPs (G)	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)
Yolo 9 t	197	1.971	7.6	94.4	93.3	96.1	61
Yolo 9 s	197	7.168	26.7	94.2	92.1	95.8	62.2
Yolo 9 m	151	20.015	76.5	96.2	91	97	62.8
Yolo 9 c	156	25.322	102.3	95.4	91.3	96.2	62.8
Yolo 9 e	279	57.379	189.1	93.6	90.6	95.3	62
Yolo 10 n	102	2.265	6.5	93.9	91.5	95.3	60.4
Yolo 10 s	106	7.219	21.4	91.3	91.1	95	61
Yolo 10 m	136	15.315	58.9	94.4	87.8	94.4	61.8
Yolo 10 l	174	24.312	120.0	92.8	90.5	94.5	63.3
Yolo 10 x	192	29.400	160.0	95.6	89.6	94.9	63

Within the YOLOv9 series, the smallest model, YOLOv9-t, exhibited a precision of 94.4%, a recall of 93.3%, and a mean Average Precision (mAP) at a 50% Intersection over Union (IoU) threshold (mAP50) of 96.1%. However, it achieved a more modest score of 61% on the mAP50-95 metric. An increase in model size generally correlated with improvements in precision, recall, and mAP scores. For instance, the YOLOv9-m model attained an mAP50-95 score of 62.8%, with 96.2% precision and 91% recall. Conversely, the largest model, YOLOv9-e, with 279 million parameters, demonstrated a precision of 93.6% and an mAP50-95 value of 62%. This indicates a trade-off between model complexity

and performance, where a high computational cost (57.379 G FLOPs) did not proportionally translate to performance gains.

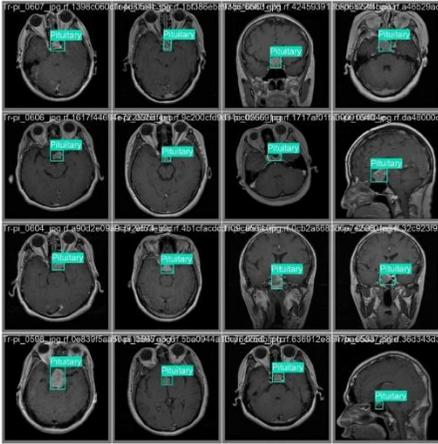
A similar trend was observed in the YOLOv10 models. The smallest variant, YOLOv10-n, delivered 93.9% precision, 91.5% recall, and 95.3% mAP50, while recording the lowest mAP50-95 score at 60.4%. As model size increased, there was a consistent rise in precision and mAP50-95 scores. Notably, the YOLOv10-x model outperformed others with 95.6% precision, 94.9% mAP50, and an mAP50-95 score of 63%, achieving these results with a reasonable number of parameters. These high scores, coupled with an acceptable parameter count, suggest that the YOLOv10-x model may be the most suitable for tumor detection tasks in terms of both accuracy and computational efficiency.

The results underscore the general trend that models with more parameters typically perform better, yet also reveal that performance gains diminish after a certain model size. For example, while the YOLOv10-l model achieved an mAP50-95 score of 63.3%, the YOLOv10-x model reached an mAP50-95 score of 63% but offered better precision and recall values compared to YOLOv10-l. Therefore, YOLOv10-x is considered to provide the best performance at a reasonable model size, making it an evident choice for real-world tumor detection applications.

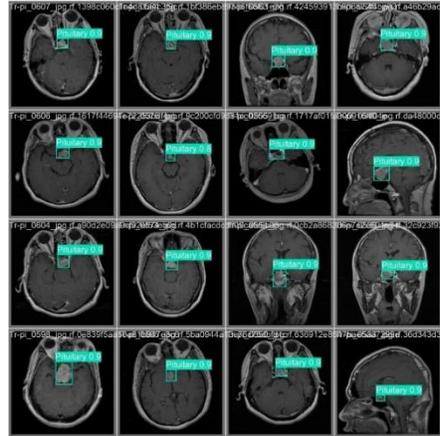
The trade-off between performance and computational cost is also apparent when examining FLOPs and parameter counts. Although the YOLOv9-e model had a significantly higher FLOPs value (57.379 G), its performance increase was limited compared to smaller models like YOLOv9-m and YOLOv10-x, which achieved very high detection accuracy without an excessive computational burden. This suggests that smaller, well-optimized models can often offer comparable or better efficiency in daily operational use than larger counterparts. Furthermore, while larger models typically enhance detection performance, this study has demonstrated that there are diminishing returns from increasing model size. Thus, our findings emphasize the importance of weighing the fundamental decision points of balancing model complexity and parameter size with

computational cost and time. Accurately weighting these trade-offs is crucial in determining the most appropriate model for real-world contexts such as tumor detection in medical imaging, where both accuracy and efficiency are critical. Further research into such trade-offs is necessary to achieve more advancements in medical image analysis.

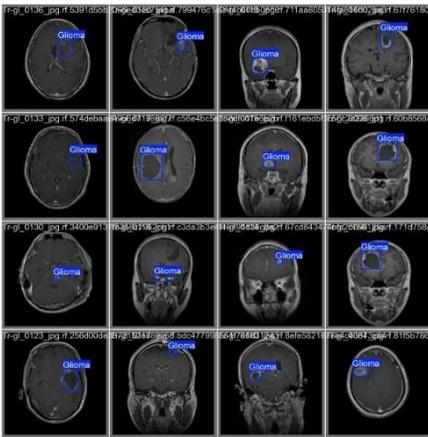
Figure 2.7 presents examples to visualize the performance of the YOLOv9-m model. Subfigures (a) and (c) display the ground truth labels of tumors in the input images, whereas subfigures (b) and (d) illustrate the corresponding predictions made by the YOLOv9-m model for the same images. This comparison allows for a qualitative assessment of the model's detection capabilities and potential error margins.



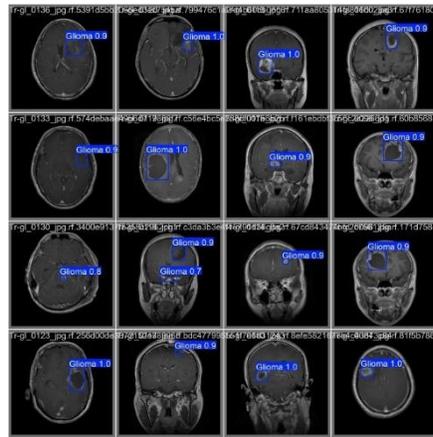
(a)



(b)



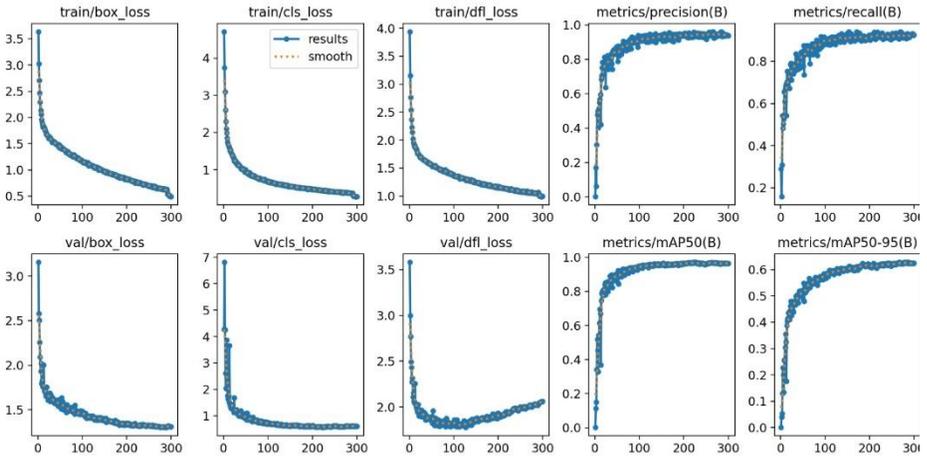
(c)



(d)

**Figure 2.7** In subfigure (a) and (c), the ground truth labels for YOLO 9M are displayed, whereas subfigure (b) and (d) illustrate the corresponding predictions.

Figure 2.8 shows a resultant image obtained using the YOLOv9-m algorithm during the experiment. This image provides a practical example of how the model detects and marks tumors with bounding boxes in an MRI scan. Such visual results are important for understanding the model's applicability and performance in real-world scenarios.



**Figure 2.8** Results from the YOLO9M (Best model) model in the experiment.

The provided Figure 2.8 illustrates the training and validation performance of an object detection model over 300 epochs. The top row of graphs depicts the training progress, specifically showing the box loss, classification (cls) loss, and distribution focal loss (DFL), alongside precision and recall metrics. These training loss curves exhibit a desirable decreasing trend, indicating that the model is effectively learning from the training dataset and minimizing errors in localization, classification, and bounding box refinement. Concurrently, the training precision and recall metrics demonstrate a consistent increase, eventually plateauing, which signifies improved accuracy in identifying positive instances and correctly classifying them. The bottom row of graphs presents the corresponding validation metrics: validation box loss, classification loss, and DFL loss, as well as mean Average Precision at an IoU threshold of 0.50 (mAP50) and mean Average Precision over IoU thresholds from 0.50 to 0.95 (mAP50-95). Similar to the training losses, the validation box and classification losses decrease and then stabilize, suggesting good generalization to unseen data. While the validation DFL loss initially decreases, a slight upward trend is noticeable towards the later epochs, which might hint at the early signs of overfitting for that specific loss component. The mAP50 and mAP50-95 curves show

a significant improvement in the initial epoch followed by a plateau, indicating that the model's as a whole detection and segmentation performance on the validation set has stabilized at a high level.

#### **4. Conclusion**

This comparative performance analysis of YOLOv9 and YOLOv10 architectures for brain tumor detection underscores that optimal model selection extends beyond mere architectural scale, necessitating careful consideration of efficiency alongside detection accuracy. While larger models initially demonstrated enhanced precision and recall, our findings confirm a clear point of diminishing returns with increasing size and parameter count. Notably, YOLOv9m and YOLOv10x variants emerged as exemplars of an effective balance between robust detection capabilities and resource economy, with YOLOv10-x particularly excelling in precision and mAP50-95 metrics. Furthermore, this study highlights a critical trade-off between computational expenditure and performance gain. For instance, the YOLOv9e model, despite its substantially higher FLOP value (57.379 G), did not yield commensurate improvements in detection accuracy when compared to the more streamlined YOLOv9m and YOLOv10x. This evidence substantiates that smaller, well-optimized models can achieve comparable, and at times superior, practical performance relative to their larger, more resource-intensive counterparts. Consequently, these findings advocate holistic evaluation in model selection for real-world applications. This is particularly crucial within demanding fields such as medical imaging for brain tumor diagnostics, where both high accuracy and operational efficiency are paramount. A judicious balance between model complexity, parameter load, and computational cost against achievable performance is essential. Future research focused on the advanced optimization of this performance-efficiency nexus is pivotal in advancing the capabilities and clinical integration of artificial intelligence in medical image analysis, specifically enhancing tools for brain tumor assessment.

## References

- [1] R. Ahsan, I. Shahzadi, F. Najeeb, H. Omer, Brain tumor detection and segmentation using deep learning, *Magnetic Resonance Materials in Physics, Biology and Medicine* 38 (2024) 13–22. <https://doi.org/10.1007/S10334-024-01203-5/FIGURES/4>.
- [2] R.L. Siegel Mph, A.N. Giaquinto, | Ahmedin, J. Dvm, R.L. Siegel, Cancer statistics, 2024, (2024). <https://doi.org/10.3322/caac.21820>.
- [3] I. Pacal, A novel Swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images, *International Journal of Machine Learning and Cybernetics* 15 (2024) 3579–3597. <https://doi.org/10.1007/s13042-024-02110-w>.
- [4] Z.N.K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, J. Lu, Content-Based Brain Tumor Retrieval for MR Images Using Transfer Learning, *IEEE Access* 7 (2019) 17809–17822. <https://doi.org/10.1109/ACCESS.2019.2892455>.
- [5] A. Selvapandian, K. Manivannan, Fusion based Glioma brain tumor detection and segmentation using ANFIS classification, *Comput Methods Programs Biomed* 166 (2018) 33–38. <https://doi.org/10.1016/J.CMPB.2018.09.006>.
- [6] S. Ince, I. Kunduracioglu, B. Bayram, I. Pacal, U-Net-Based Models for Precise Brain Stroke Segmentation, *Chaos Theory and Applications* 7 (2025) 50–60. <https://doi.org/10.51537/CHAOS.1605529>.
- [7] I. Pacal, O. Celik, B. Bayram, A. Cunha, Enhancing EfficientNetv2 with global and efficient channel attention mechanisms for accurate MRI-Based brain tumor classification, *Cluster Comput* 27 (2024) 11187–11212. <https://doi.org/10.1007/S10586-024-04532-1/TABLES/4>.
- [8] S.K. Pandey, A.K. Bhandari, YOLOv7 for brain tumour detection using morphological transfer learning model, *Neural Comput Appl* 36 (2024) 20321–20340. <https://doi.org/10.1007/S00521-024-10246-7/TABLES/9>.

- [9] I. Pacal, O. Attallah, InceptionNeXt-Transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis, *Biomed Signal Process Control* 110 (2025) 108116. <https://doi.org/10.1016/J.BSPC.2025.108116>.
- [10] I. Pacal, O. Attallah, Hybrid deep learning model for automated colorectal cancer detection using local and global feature extraction, *Knowl Based Syst* 319 (2025) 113625. <https://doi.org/10.1016/J.KNOSYS.2025.113625>.
- [11] B. Ozdemir, I. Pacal, A robust deep learning framework for multiclass skin cancer classification, *Sci Rep* 15 (2025) 1–19. <https://doi.org/10.1038/S41598-025-89230-7>;SUBJMETA=166,2321,631,639,67,985;KWRD=BIOMEDICAL+ENGINEERING,CANCER+IMAGING.
- [12] B. Ozdemir, E. Aslan, I. Pacal, Attention Enhanced InceptionNeXt Based Hybrid Deep Learning Model for Lung Cancer Detection, *IEEE Access* (2025). <https://doi.org/10.1109/ACCESS.2025.3539122>.
- [13] I. Aruk, I. Pacal, A.N. Toprak, A novel hybrid ConvNeXt-based approach for enhanced skin lesion classification, *Expert Syst Appl* 283 (2025) 127721. <https://doi.org/10.1016/J.ESWA.2025.127721>.
- [14] S. Ince, I. Kunduracioglu, A. Algarni, B. Bayram, I. Pacal, Deep learning for cerebral vascular occlusion segmentation: A novel ConvNeXtV2 and GRN-integrated U-Net framework for diffusion-weighted imaging, *Neuroscience* 574 (2025) 42–53. <https://doi.org/10.1016/J.NEUROSCIENCE.2025.04.010>.
- [15] M. Kang, C.M. Ting, F.F. Ting, R.C.W. Phan, BGF-YOLO: Enhanced YOLOv8 with Multiscale Attentional Feature Fusion for Brain Tumor Detection, *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 15008 LNCS (2024) 35–45. [https://doi.org/10.1007/978-3-031-72111-3\\_4](https://doi.org/10.1007/978-3-031-72111-3_4).

- [16] A. Elhanashi, S. Saponara, P. Dini, Q. Zheng, F.Z. Ali, Y. Singh, S. Kuanar, Y. Huo, K. Younis, Advanced Deep Learning in Medical Imaging: Brain Tumor Detection and Localization with YOLOv9, *Lecture Notes in Electrical Engineering* 1369 LNEE (2025) 113–120. [https://doi.org/10.1007/978-3-031-84100-2\\_14](https://doi.org/10.1007/978-3-031-84100-2_14).
- [17] S.K. Pandey, A.K. Bhandari, Morphological transfer learning based brain tumor detection using YOLOv5, *Multimed Tools Appl* 83 (2024) 49343–49366. <https://doi.org/10.1007/S11042-023-17367-6/FIGURES/5>.
- [18] Y. Du, W. Liu, Y. Wang, R. Li, L. Xie, YOLO-CPC: a breast tumor detection and identification algorithm based on improved YOLOv7, *Signal Image Video Process* 19 (2025) 1–11. <https://doi.org/10.1007/S11760-024-03811-Z/TABLES/4>.
- [19] Y. Cakmak, S. Safak, M.A. Bayram, I. Pacal, Comprehensive Evaluation of Machine Learning and ANN Models for Breast Cancer Detection, *Computer and Decision Making: An International Journal* 1 (2024) 84–102. <https://doi.org/10.59543/COMDEM.V1I.10349>.
- [20] Y. Cakmak, I. Pacal, Enhancing Breast Cancer Diagnosis: A Comparative Evaluation of Machine Learning Algorithms Using the Wisconsin Dataset, *Journal of Operations Intelligence* 3 (2025) 175–196. <https://doi.org/10.31181/JOPI31202539>.
- [21] K.R. Bhatele, S.S. Bhadauria, Brain structural disorders detection and classification approaches: a review, *Artif Intell Rev* 53 (2020) 3349–3401. <https://doi.org/10.1007/S10462-019-09766-9/TABLES/13>.
- [22] M.I. Sharif, J.P. Li, J. Naz, I. Rashid, A comprehensive review on multi-organs tumor detection based on machine learning, *Pattern Recognit Lett* 131 (2020) 30–37. <https://doi.org/10.1016/J.PATREC.2019.12.006>.
- [23] Labeled MRI Brain Tumor Dataset Object Detection Dataset (v1, 2024-12-16 3:21pm) by denver, (n.d.). <https://universe.roboflow.com/denver-csxhq/labeled-mri-brain-tumor-dataset-jwgzs-zjflf/dataset/1> (accessed May 11, 2025).

- [24] S.A. Swapnil, V.S. Girish, Image Mining Methodology for Detection of Brain Tumor: A Review, Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020 (2020) 232–237. <https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00044>.
- [25] N. Kumari, S. Saxena, Review of Brain Tumor Segmentation and Classification, Proceedings of the 2018 International Conference on Current Trends towards Converging Technologies, ICCTCT 2018 (2018). <https://doi.org/10.1109/ICCTCT.2018.8551004>.
- [26] L. Sahoo, L. Sarangi, B.R. Dash, H.K. Palo, Detection and Classification of Brain Tumor Using Magnetic Resonance Images, Lecture Notes in Electrical Engineering 665 (2020) 429–441. [https://doi.org/10.1007/978-981-15-5262-5\\_31](https://doi.org/10.1007/978-981-15-5262-5_31).
- [27] M. Sharif, J. Amin, M. Raza, M.A. Anjum, H. Afzal, S.A. Shad, Brain tumor detection based on extreme learning, Neural Comput Appl 32 (2020) 15975–15987. <https://doi.org/10.1007/S00521-019-04679-8/TABLES/6>.
- [28] M. Mudda, R. Manjunath, N. Krishnamurthy, Brain Tumor Classification Using Enhanced Statistical Texture Features, IETE J Res 68 (2022) 3695–3706. <https://doi.org/10.1080/03772063.2020.1775501>.
- [29] S. Saurav, A. Sharma, R. Saini, S. Singh, An attention-guided convolutional neural network for automated classification of brain tumor from MRI, Neural Comput Appl 35 (2023) 2541–2560. <https://doi.org/10.1007/S00521-022-07742-Z/TABLES/11>.
- [30] C.-Y. Wang, I.-H. Yeh, H.-Y.M. Liao, YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information, (n.d.). <https://github.com/WongKinYiu/yolov9>. (accessed May 10, 2025).
- [31] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, G. Ding, YOLOv10: Real-Time End-to-End Object Detection, (2024). <https://github.com/THU-MIG/yolov10>. (accessed May 10, 2025).

- [32] S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, Path Aggregation Network for Instance Segmentation, n.d.
- [33] Y. Cai, Y. Zhou, Q. Han, J. Sun, X. Kong, J. Li, X. Zhang, Reversible Column Networks, (2022). <http://arxiv.org/abs/2212.11696>.
- [34] T. Shelatkar, D. Urvashi, M. Shorfuzzaman, A. Alsufyani, K. Lakshmana, Diagnosis of Brain Tumor Using Light Weight Deep Learning Model with Fine-Tuning Approach, *Comput Math Methods Med* 2022 (2022) 2858845. <https://doi.org/10.1155/2022/2858845>.
- [35] Z. Wang, P. Wang, K. Liu, P. Wang, Y. Fu, C.-T. Lu, C.C. Aggarwal, J. Pei, Y. Zhou, A Comprehensive Survey on Data Augmentation, (2024). <https://arxiv.org/abs/2405.09591v3> (accessed May 16, 2025).
- [36] A. Mumuni, F. Mumuni, N.K. Gerrar, A Survey of Synthetic Data Augmentation Methods in Machine Vision, *Machine Intelligence Research* 2024 21:5 21 (2024) 831–869. <https://doi.org/10.1007/S11633-022-1411-7>.

# Chapter 3

---

## Evaluating YOLOv10 and YOLOv11 Variants for Efficient Polyp Detection

### Abstract

Colorectal cancer (CRC) is a significant global source of morbidity and mortality; however, it is highly preventable through the early detection and removal of pre-invasive lesions, namely polyps. While colonoscopy is the gold standard for CRC screening and polyp detection, human factor-related miss rates have spurred interest in artificial intelligence (AI)-based computer-aided detection (CADe) and diagnosis (CADx) systems, with deep learning models showing promise. This study conducted a comprehensive performance evaluation of various versions of the latest YOLO (You Only Look Once) architectures, YOLOv10 and YOLOv11, for polyp detection. The models were assessed not only on standard detection metrics precision, recall, F1 score, and mean Average Precision (mAP50-95)—but also on structural and efficiency characteristics, including model architecture, layer depth, parameter count, and computational cost per inference per image. Analyses revealed that larger YOLO models generally achieved higher detection accuracy, though diminishing returns in performance gains were observed beyond a certain model size. Notably, the YOLOv10m, YOLOv11n, and YOLOv11s versions demonstrated an optimal balance between detection accuracy and model efficiency. Specifically, the YOLOv10m model, with 84.5% precision and 60.0% mAP50-95, emerged as a strong candidate for practical clinical applications. Similarly, the YOLOv11n model delivered competitive detection performance despite the efficiency advantages of its compact structure. These findings indicate that YOLOv10m, YOLOv11n, and -s represent different optimization points on the accuracy-computational cost spectrum, highlighting the potential for effectively deploying smaller, hyper-optimized models.

## 1. Introduction

When it comes to total random mortality, cancer is still one of the most significant global public health challenges. Although this socially important disease has many diagnostic tools and treatment options, sometimes it is still challenging to detect cancer early enough to add value to the patient's outcome [1]. Each year, millions of new cancer cases are diagnosed, many of which are preventable or could be treated promptly if identified early, according to the World Health Organization (WHO). Colorectal cancer (CRC) is one of those cases with very high incidence and death rates across all societies [2].

CRC is a sizable health burden to all transitional and developing countries. According to cancer statistics, in 2024, there were approximately 152,810 CRC diagnoses in the United States and around 53,010 deaths related to this disease [3]. Clearly for CRC, it is the third most common cancer in the world and the second leading cause of cancer deaths. The worldwide figures continued to escalate to nearly 1.9 million new CRC cases and around 900,000 deaths due to CRC by 2022 thus advising that society continue to face challenges with the ever-increasing rates of incidence and mortality in CRC and new and effective methods of early detection/early treatment [4].

CRC pathogenesis generally follows a stepwise process starting with benign polyps in the mucosal layer of the colon or rectum that can, eventually, become tumors via malignant transformation. This process, the "adenoma-carcinoma sequence", takes 5 to 15 years, which is why there is a window to intervene, identify, and intervene. Identifying and removing adenomatous polyps early greatly reduces the risk of developing CRC. But such an opportunity depends on the existence of real-time, highly sensitive detection. Colonoscopy is considered gold standard for diagnosis and prevention of CRC because it can identify polyps and remove them at the same time. However, studies have reported that up to 25% of polyps may go undetected at colonoscopic investigation, especially when they are flat or small [5–7].

Through the digital transformation of medicine, AI-based computer-aided diagnosis (CADx) systems are emerging as significant platforms to improve early cancer diagnosis [8–12]. Deep learning, particularly convolutional neural network (CNN)-based architectures, is a game-changer for the analysis of medical images, providing unparalleled accuracy at tremendous speed [13,14]. The "You Only Look Once" (YOLO) algorithm, in particular, has recently gained interest mainly in the field of object detection in real-time which is useful for identifying polyps during an endoscopic procedure [15–21]. YOLO models perform an analysis by bounding the whole image as a singular pass producing high-speed high-precision analysis well suited for a clinical environment [22,23].

The current literature suggests deep-learning methods have led to substantial improvements in medical research applications and the deeper that endoscopists engage with relevant literature and artificial intelligence (AI) based technologies, the more likely they are to win over their peers. Hospitals are also investing in collecting their own proprietary datasets. For example, Catlow and colleagues found that endoscopists with lower polyp detection rates were associated with higher CRC incidence and mortality. A study of the UK's National Endoscopy Database (NED) explored the ability of performance feedback based on their adjusted mean number of polyps (aMNP) for complexity of case to improve endoscopist's performance [24].

Raseena et al. emphasized the global burden of CRC and the significance of early diagnosis. They set forth the idea of missed polyps at a 25% rate during colonoscopy, and proposed the DeepCPD model, which combined a transformer-based architecture and a Linear Multi-head Self-Attention (LMSA) mechanism with data augmentation. On the four separate datasets that DeepCPD was tested on, DeepCPD attained accuracy, precision and recall of above 98%, and a training time decreasing by about 1.2 times. The high recall performance reduced false negatives, which can ultimately increase clinical diagnostic capabilities [25].

Wu et al. also developed PATM-YOLO as a modified version of the YOLOv5 model for the detection of small polyps. Their design included a detection head,

a new PATM attention module, and an optimized ASFF function for effective feature fusion. Now with the Swin Transformer, their model was even better! Inferences on both a proprietary and public dataset showed that PATM-YOLO exceeded YOLOv5, with a detection accuracy of 91.3% which was a 8.5% improvement from baseline [26].

In a different noteworthy study, Sherif et al. compared the performance of EndoCuff-assisted colonoscopy (EAC) with standard colonoscopy (SC) for polyp detection. During 2018-2020 at Cairo University Hospitals with 214 patients, EAC performed best for the detection of small polyps ( $\leq 9$  mm) with a detection rate of 32.24% and SC with a lower detection rate of 26.64% ( $p < 0.05$ ). The adenoma detection rates were also higher with EAC as they reported an comprehensive 17.2% versus SC 14.9% ( $p < 0.05$ ). Minor mucosal erosions were noted in % 2.8 of patients [27].

Finally, Sushama and Menon highlighted the increasing incidence of CRC in India emphasizing the necessity for early detection in the proposed patients. They introduced a CNN-based model to minimize glare influences in colonoscopic images while minimizing selection bias. Trained on the PolypGen dataset, the model was well-trained and able to generalize due to being trained on polyp and normal images. The assessment of the model on multiple datasets, including the Gastrointestinal Atlas-Colon Polyp, Gastrolab-Polyp, and ETIS-LARIB datasets, we observed the model scored precision, recall, F1 and F2 all above 75% [28].

In recent years, the increasing use of machine learning and deep learning algorithms in the medical field has drawn significant attention [29,30]. This research examines the relative performance of the various YOLOv10 and YOLOv11 model variants nano, small, medium, large, and xlarge in polyp detection [31,32]. We utilize some popular datasets, namely CVC-ClinicDB, CVC-ColonDB, Kvasir-SEG, and ETIS-LARIB. We held out a subset known as CVC-ColonDB as the validation set and merged them to create training data from the CVC-ClinicDB, Kvasir-SEG, ETIS-LARIB datasets. This merging of datasets to create the training data was intended to increase the model's ability to

learn on a pooled dataset allowing greater transferability of the model to other diverse datasets [33,34].

## **2. Materials and Methods**

### **2.1 Datasets**

Timely identification and therapy of gastrointestinal polyps is essential in the prevention of colorectal cancer. For this purpose, researchers at the intersection of medical imaging and computer vision have invested considerable time and effort developing complex algorithms for the automatic detection and segmentation of polyps. The success of these efforts largely rests on the quality and correct annotation of the datasets used. In our study, we combined four well-known datasets CVC-ClinicDB, CVC-ColonDB, ETIS-LARIB Polyp DB, and Kvasir-SEG to create a combined framework for polyp detection. These collections include multiple endoscopic medical images and the corresponding annotated segmentation masks upon which to train deep learning models.

The combined dataset includes 2,188 images in total, with 1,808 images entered the training split and 380 into the validation split. These datasets include polyps that are quite different in size, shape, and texture. The ambiguity in both features allows researchers to develop models that are more robust and more generalizable. For instance, images from CVC-ClinicDB and CVC-ColonDB are from colonoscopies, whereas ETIS-LARIB includes images of polyps that have more varying characteristics and tissue diversity. Kvasir-SEG provides validated segmentation masks for 1,000 polyps, with gastroenterology experts validating each segmentation mask to increase the quality of automated assessment. Putting together these datasets provides models that can be more realistically aligned with clinical reality and likely improves performance in a variety of scenarios. All of these datasets together provide a strong foothold for future research related to polyp detection. The distribution of each of the datasets presented in the training and validation splits related to both CVC-ClinicDB and Kvasir-SEG can be found

in Table 3.1. Example images from the combined datasets are shown in Figure 3.1 [35–38].

**Table 3.1** Distribution of images by dataset and assigned tasks.

Dataset	Task	Images
CVC-ClinicDB	Train	612
CVC-ColonDB	Test	380
Kvasir-SEG	Train	1000
ETIS-LARIB	Train	196
<b>Total</b>		2188



**Figure 3.1** Sample polyp images from different datasets.

## 2.2 Data preprocessing steps

In this study, various YOLO algorithm variants were employed to detect polyps, making full use of the built-in data augmentation techniques these models offer by default. Data augmentation plays a pivotal role in enhancing a model’s ability to generalize, as it helps the system learn to recognize objects under diverse conditions such as varying lighting, scales, positions, and orientations. The augmentation strategies automatically implemented by YOLO encompass a range of transformations, including adjustments in color space (notably HSV

modifications), geometric manipulations, mosaic augmentation, Auto Augment, and random erasing [39].

Specifically, the HSV adjustments involved subtle shifts in hue, saturation, and brightness by factors of 0.015, 0.7, and 0.4 respectively, fine-tuning the color properties of the images to reflect natural variations. Geometrically, images were translated by up to 10%, scaled by as much as 50%, and randomly flipped horizontally with a 50% chance, thus diversifying the spatial presentation of polyps. Mosaic augmentation applied with full intensity combined four distinct images into one, thereby exposing the model to a broader spectrum of backgrounds and spatial arrangements. Complementing this, the Auto Augment technique (specifically Rand Augment) randomly applies various transformations to increase data variety and effectively curb overfitting. Meanwhile, Random Erasing, activated with a 40% probability, randomly obscures parts of the images, simulating scenarios where visual data might be partially missing or corrupted, which in turn bolsters the model's resilience.

By integrating these augmentation methods, the training process encourages the development of models that are more robust and adaptable to real-world challenges such as fluctuations in illumination, changes in scale and orientation, partial occlusions, and diverse environmental conditions. This holistic approach not only improves training efficiency but also yields models with stronger generalization capabilities, ensuring their performance remains reliable when applied beyond controlled datasets [40].

### **2.3 Yolo Algorithms**

Detecting polyps, especially in medical imaging techniques like colonoscopy, are critical for early detection and actionable solutions. In this case polyps, YOLO algorithms represent one of the best types of models available as they can process input in real time, which is imperative in clinical scenarios. This is enabled by methodologically balancing speed and accuracy. YOLO models use a grid-based approach, splitting the image into an n-by-n grid, predicting the locations and

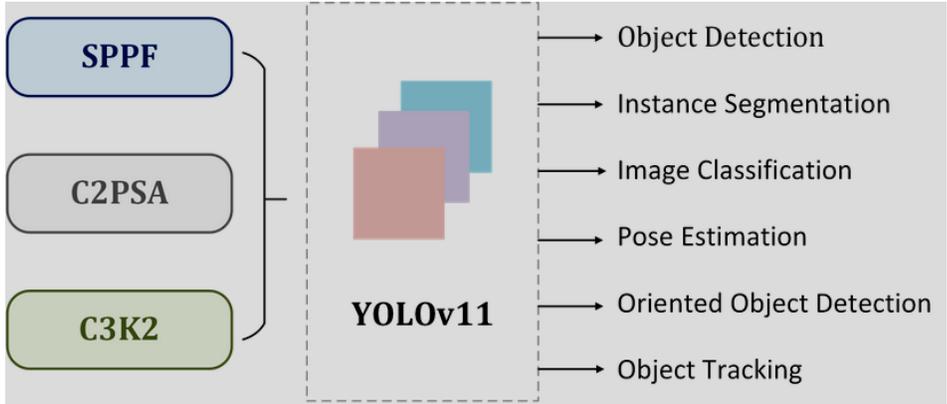
classes of objects in each cell. YOLO does this with a single pass over the image, unlike some detection methods, more efficiently with faster results.

In this paper, we used the YOLOv10 and YOLOv11 families of models specifically for detecting polyps in colonoscopy images. Both families of models are available in multiple configurations to leverage all levels of computational capacity of the hardware platform. For example, the YOLOv10 series offers a number of models of varying sizes, ranging from very small networks like YOLOv10n with 2.69 million parameters and 8.2 GFLOPs of required compute, to larger, more accurate networks like YOLOv10x with 31.58 million parameters and 169.8 GFLOPs. There are also intermediate sizes YOLOv10-s, m, l, and b, providing relatively the same detection performance for various computational costs, which allow researchers or developers to select based on the application they are deploying and resources available.

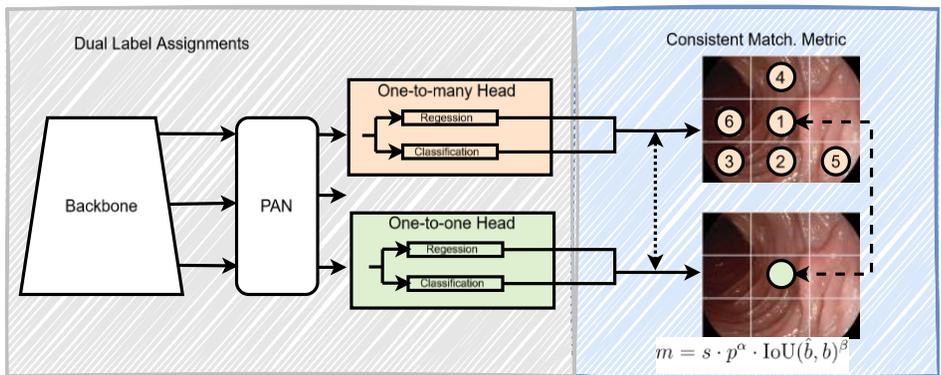
Likewise, the YOLOv11 series was examined with respect to the possible use for polyp detection and followed with an examination of accuracy metrics. The YOLOv11n model, for the most efficient resource consumption, is only 2.58 million parameters, and 6.3 GFLOPs. Clearly, if hardware is limited then this would be a reasonable option. YOLOv11x, on the other hand, uses 56.82 million parameters, and 194.4 GFLOPs and would be used if it were essential to have as accurate a classified detection are possible. YOLOv11s, m, and l represent different compromises achieved between processing time and accuracy which would allow disclosure of YOLOv11 into almost any context, clinical or otherwise.

The computing requirements of these models are influenced, mostly by other parameters of the neural network, including how many layers and total parameters, and the floating-point operations per second (FLOPs). As mentioned in this discussion, while larger and deeper models have more capacity and generally have better accuracy, they almost universally have higher computing demands. Therefore, the choice of a model in clinical-application polyp detection requires finding the best model from a hardware-capacity standpoint that also

provides as much diagnostic accuracy level possible. Detailed architectures of these models can be found in Figures 3.2 and 3.3 [33,34].



**Figure 3.2** Key architectural modules in YOLOv11.



**Figure 3.3** Consistent dual assignments for NMS-free training.

## 2.4 Evaluation Metrics

In this study, the performance of YOLOv9 and YOLOv10 model variants for polyp detection was examined through a set of well-established evaluation metrics commonly employed in object detection tasks. Among these, Precision stands out as a measure of the proportion of correctly identified objects among all predictions made by the model, effectively highlighting how well the model avoids false positives. On the other hand, Recall quantifies the fraction of actual objects that the model successfully detects, with higher values indicating a

reduced likelihood of missing true positives. Together, these two metrics provide a balanced perspective on the model's ability to accurately and comprehensively identify relevant objects.

In addition to Precision and Recall, the Mean Average Precision (mAP) metric was also critically examined. mAP is a more inclusive measure of detection quality. mAP @50 demonstrates how well the predicted bounding boxes overlap with ground truth annotations when the Intersection over Union (IoU) is set to 50%. Here we see our mAP scores are mostly above 0.97, which shows excellent performance within a relatively wide window for overlap. mAP @50-95 gives us an in general picture of performance across the different IoU thresholds (50%-95%). mAP @50-95 shows us how accurately predicted bounding boxes delineate the ground truth annotations across different degrees of tolerated overlap presented by the various thresholds. The IoU metric allows for an objective and quantitative measure of accuracy, where higher values reflect greater precision regarding where the predicted bounding boxes and true bounding boxes overlap.

Together, these measurements allowed us to achieve a comprehensive understanding of not only the accuracy of the models, but the robustness and generalizability of the models for various labels. The results presented in a quantitative manner provide immensely valuable management-based evidence to assess the most suitable model for the careful selection and implementation of accurate detection for clinical applications such as polyp detection, where both precision and recall are necessary to reliably support clinical judgements.

## **2.5 Experimental design**

All experiments conducted as part of this study were performed on a Linux-based system running Ubuntu 22.04. The hardware configuration included an Intel Core i5-13600K CPU, 32 GB of DDR5 RAM, and an NVIDIA RTX 3090 GPU, which collectively provided a robust environment for deep learning tasks. PyTorch was used as the primary framework, with GPU acceleration enabled

through NVIDIA's CUDA toolkit. To maintain consistency and ensure a fair comparison between different models, the same computing setup and identical hyperparameter settings were applied throughout the entire training and evaluation processes.

### 3. Results and Discussion

Table 3.2 provides a complete aggregate of the results for the different YOLOv10 and YOLOv11 model types from the trials. The analysis utilized mainly the primary metrics of precision, recall, mAP50, and mAP50-95 in combination to provide a well-rounded understanding of how these models perform in polyp detection. These metrics provide a clear framework for assessing the models' aggregate accuracy, sensitivity, and generalizability across classes.

**Table 3.2** Summary of key architectural details and performance metrics for YOLOv10 and YOLOv11 models, highlighting parameter size, computational cost, and detection accuracy.

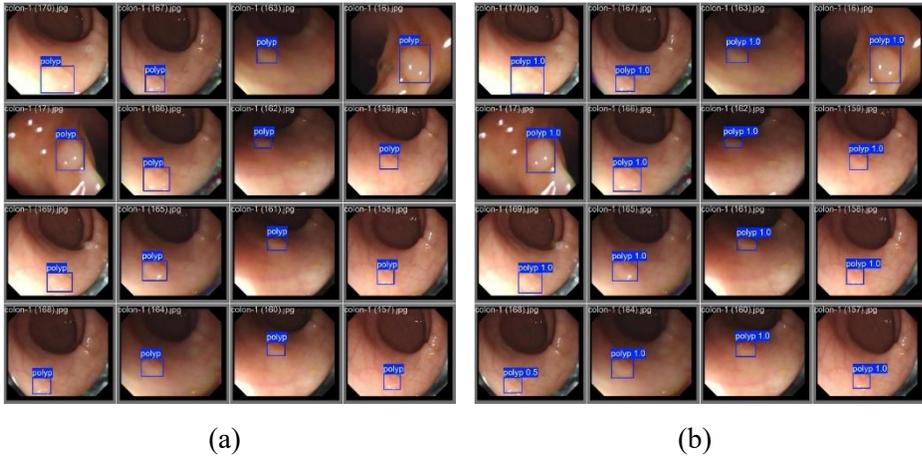
<b>Models</b>	<b>Layer</b>	<b>Params</b>	<b>FLOPs (G)</b>	<b>Precision</b>	<b>Recall</b>	<b>mAP50</b>	<b>mAP50-95</b>
		<b>(M)</b>		<b>(%)</b>	<b>(%)</b>	<b>(%)</b>	<b>(%)</b>
Yolo 10 n	125	2.69	8.2	89.2	68.1	80.3	56.9
Yolo 10 s	129	8.03	24.4	90.9	71.4	82.8	57.8
Yolo 10 m	159	16.45	63.4	84.5	68.8	81.5	60.0
Yolo 10 l	197	25.71	126.3	89.9	69.2	82.5	58.9
Yolo 10 b	165	20.41	97.9	85.5	70.6	81.5	57.4
Yolo 10 x	215	31.58	169.8	86.7	69.1	81.5	59.4
Yolo 11 n	100	2.58	6.3	91.1	73.6	84.9	59.3
Yolo 11 s	100	9.41	21.3	83.5	74.4	83.1	59.1
Yolo 11 m	125	20.03	67.6	83.1	74.2	82.8	57.6
Yolo 11 l	190	25.28	86.6	88.1	70.4	82.3	55.9
Yolo 11 x	190	56.82	194.4	86.3	72.1	83.7	58.4

The data presented in Table 3.2 reveals a clear correlation between model complexity, defined by parameter count and computational cost (FLOPs), and object detection performance. However, this trend is not linear and demonstrates diminishing returns beyond a certain threshold. Critically, the largest models do not uniformly yield the best results. For instance, the YOLOv10x, despite being the most computationally demanding model in its series with 31.58M parameters and 169.8 GFLOPs, achieves a lower mAP50-95 (59.4%) than the more streamlined YOLOv10m model (60.0%). A similar pattern is observed in the YOLOv11 series, where the largest variant, YOLOv11x, offers only marginal performance gains over its more compact counterparts.

This analysis underscores a crucial trade-off between detection accuracy and computational efficiency. For practical applications such as real-time polyp detection, models like YOLOv10m, YOLOv11n, and YOLOv11s emerge as superior candidates. The YOLOv11n model, for example, achieves a robust 59.3% mAP50-95 with a minimal footprint of 2.58M parameters and 6.3 GFLOPs, showcasing exceptional efficiency. Likewise, the YOLOv10m provides the highest mAP50-95 in its series while maintaining a moderate computational load. To visualize its practical efficacy, Figure 3.4 presents experimental results from the YOLOv10m model, displaying its predictions against the ground truth labels on images from the CVC-ColonDB dataset.

Ultimately, the empirical evidence suggests that merely increasing model size is a suboptimal strategy for enhancing detection accuracy. The primary challenge in domains such as medical imaging is to identify architectures that strike an optimal balance between diagnostic precision and computational feasibility. Future research should therefore focus on developing lightweight, optimized models that build upon these insights. By refining the accuracy-efficiency equilibrium, subsequent work can advance the development of more effective and reliable automated diagnostic tools. Figure 3.4 presents experimental results for the YOLOv10m model on images from the CVC-ColonDB dataset. Subfigure (a)

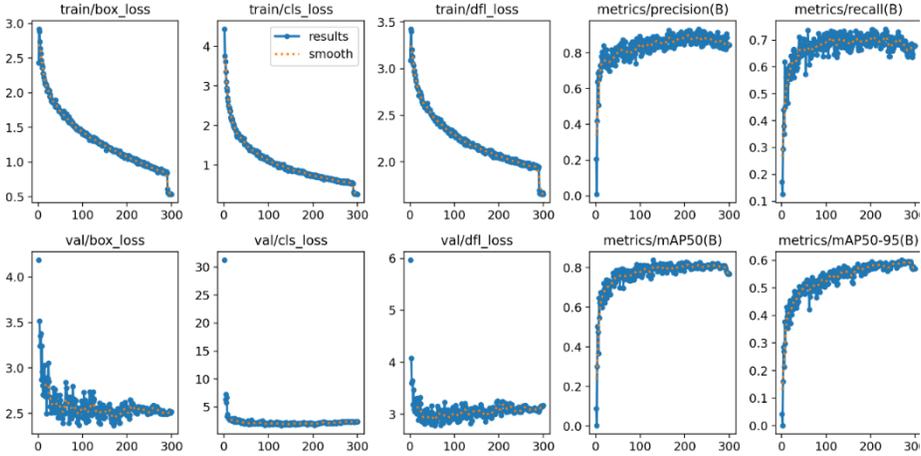
displays the ground truth labels for polyps, while subfigure (b) illustrates the corresponding polyp predictions made by the model.



**Figure 3.4** In subfigure (a), the ground truth labels for YOLOv10m are displayed, whereas subfigure (b) illustrates the corresponding predictions.

Figure 3.4 offers a qualitative validation of the quantitative results by juxtaposing the manual annotations in subfigure (a) with the YOLOv10m detections in subfigure (b). Across the sampled frames, the predicted bounding boxes are almost perfectly congruent with the reference outlines, and the associated confidence scores frequently reach one, reflecting a decisive classifier response. Minor spatial offsets appear only rarely and never lead to missed lesions or extraneous markings, underscoring the detector’s capacity to preserve lesion localization under substantial intra-procedural variability. Notably, the model maintains accurate delineation despite luminance fluctuations, specular highlights, and variations in mucosal texture, which attests to the effectiveness of its multi scale feature aggregation in capturing illumination-invariant cues. This visual concordance reinforces the numerical performance metrics reported earlier and confirms that the proposed system delivers both precision and robustness—qualities that are essential for reliable integration into computer-aided colonoscopy workflows. Figure 3.5 presents the visualization of various metrics,

including loss functions and performance measures, obtained during the training and validation phases of the YOLOv10m model.



**Figure 3.5** The results of YOLOv10m (best model) algorithm in the experiments.

The training and validation curves depicted in Figure 3.5 illustrate dynamics and model performance throughout 300 epochs. The training losses, including box regression, classification, and distribution focal loss, exhibit a consistent and steady decline, indicating effective optimization and improved model fitting over time. Correspondingly, the precision and recall metrics on the training set show progressive increases, plateauing near 0.85 and 0.7 respectively, which suggest balanced improvements in both sensitivity and specificity. However, the validation losses, especially for classification and distribution focal loss, demonstrate a notable plateau and slight fluctuations after initial rapid decreases, signaling potential overfitting or dataset complexity challenges. Despite this, validation precision, recall, and mean average precision metrics continue to improve gradually, reaching respectable levels above 0.75 for mAP50 and above 0.55 for mAP50-95. This pattern underscores the model's strong generalization capability while highlighting areas for further tuning, such as potential regularization or data augmentation strategies, to enhance stability and mitigate

validation loss variability. Aggregate, these results confirm that the model achieves robust and reliable polyp detection performance across both training and validation datasets.

#### **4. Conclusion**

The comprehensive evaluation of YOLOv10 and YOLOv11 architectures for polyp detection performance presented in this study highlights a nuanced balance between model complexity and efficiency. Our key findings indicate that while larger architecture generally offers improvements in raw detection metrics, performance gains are subject to diminishing returns when considering computational cost and parameter count. Specifically, more compact models such as YOLOv10m (Precision: 84.5%, mAP50-95: 60.0%) and YOLOv11n were found to exhibit a noteworthy balance between detection accuracy and operational efficiency. These optimized smaller models, while requiring significantly fewer computational resources, can in some cases offer competitive or even superior performance compared to their larger counterparts. These results strongly suggest that in resource-constrained and rapid-inference demanding fields, such as medical imaging, priority should be given to architectural efficiency and performance optimization rather than merely increasing model size. Consequently, this study affirms the importance of seeking an optimal compromise between model complexity, parameter burden, and detection efficacy in practical applications. Future research should focus on the further refinement and hyper-optimization of such efficient models for reliable deployment in clinical settings where precision is critical.

## References

- [1] I. Pacal, D. Karaboga, A robust real-time deep learning based automatic polyp detection system, *Comput Biol Med* 134 (2021). <https://doi.org/10.1016/J.COMPBIOMED.2021.104519>.
- [2] J. Regula, M. Rupinski, E. Kraszewska, M. Polkowski, J. Pachlewski, J. Orłowska, M.P. Nowacki, E. Butruk, Colonoscopy in Colorectal-Cancer Screening for Detection of Advanced Neoplasia, *New England Journal of Medicine* 355 (2006) 1863–1872. <https://doi.org/10.1056/nejmoa054967>.
- [3] I. Pacal, A. Karaman, D. Karaboga, B. Akay, A. Basturk, U. Nalbantoglu, S. Coskun, An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets, *Comput Biol Med* 141 (2022). <https://doi.org/10.1016/J.COMPBIOMED.2021.105031>.
- [4] R.L. Siegel Mph, A.N. Giaquinto, | Ahmedin, J. Dvm, R.L. Siegel, Cancer statistics, 2024, (2024). <https://doi.org/10.3322/caac.21820>.
- [5] J. Bernal, N. Tajkbaksh, F.J. Sanchez, B.J. Matuszewski, H. Chen, L. Yu, Q. Angermann, O. Romain, B. Rustad, I. Balasingham, K. Pogorelov, S. Choi, Q. Debard, L. Maier-Hein, S. Speidel, D. Stoyanov, P. Brandao, H. Cordova, C. Sanchez-Montes, S.R. Gurudu, G. Fernandez-Esparrach, X. Dray, J. Liang, A. Histace, Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results from the MICCAI 2015 Endoscopic Vision Challenge, *IEEE Trans Med Imaging* 36 (2017) 1231–1249. <https://doi.org/10.1109/TMI.2017.2664042>.
- [6] D.A. Corley, C.D. Jensen, A.R. Marks, W.K. Zhao, J.K. Lee, C.A. Doubeni, A.G. Zauber, J. De Boer, B.H. Fireman, J.E. Schottinger, V.P. Quinn, N.R. Ghai, T.R. Levin, C.P. Quesenberry, Adenoma Detection Rate and Risk of Colorectal Cancer and Death A BS TR AC T, *N Engl J Med* 14 (2014) 1298–306. <https://doi.org/10.1056/NEJMoa1309086>.
- [7] H.A. Qadir, Y. Shin, J. Solhusvik, J. Bergsland, L. Aabakken, I. Balasingham, Toward real-time polyp detection using fully CNNs for 2D

- Gaussian shapes prediction, *Med Image Anal* 68 (2021).  
<https://doi.org/10.1016/j.media.2020.101897>.
- [8] I. Pacal, O. Attallah, InceptionNeXt-Transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis, *Biomed Signal Process Control* 110 (2025) 108116.  
<https://doi.org/10.1016/J.BSPC.2025.108116>.
- [9] I. Pacal, A novel Swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images, *International Journal of Machine Learning and Cybernetics* 15 (2024) 3579–3597.  
<https://doi.org/10.1007/s13042-024-02110-w>.
- [10] B. Ozdemir, I. Pacal, A robust deep learning framework for multiclass skin cancer classification, *Sci Rep* 15 (2025) 1–19.  
<https://doi.org/10.1038/S41598-025-89230-7>;SUBJMETA=166,2321,631,639,67,985;KWRD=BIOMEDICAL+ENGINEERING,CANCER+IMAGING.
- [11] B. Ozdemir, E. Aslan, I. Pacal, Attention Enhanced InceptionNeXt Based Hybrid Deep Learning Model for Lung Cancer Detection, *IEEE Access* (2025). <https://doi.org/10.1109/ACCESS.2025.3539122>.
- [12] L.F. Sánchez-Peralta, J.B. Pagador, F.M. Sánchez-Margallo, Artificial Intelligence for Colorectal Polyps in Colonoscopy, *Artif Intell Med* (2021) 1–15. [https://doi.org/10.1007/978-3-030-58080-3\\_308-1](https://doi.org/10.1007/978-3-030-58080-3_308-1).
- [13] I. Pacal, O. Attallah, Hybrid deep learning model for automated colorectal cancer detection using local and global feature extraction, *Knowl Based Syst* 319 (2025) 113625.  
<https://doi.org/10.1016/J.KNOSYS.2025.113625>.
- [14] I. Aruk, I. Pacal, A.N. Toprak, A novel hybrid ConvNeXt-based approach for enhanced skin lesion classification, *Expert Syst Appl* 283 (2025) 127721. <https://doi.org/10.1016/J.ESWA.2025.127721>.

- [15] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, G. Ding, YOLOv10: Real-Time End-to-End Object Detection, (2024). <https://github.com/THU-MIG/yolov10>. (accessed February 12, 2025).
- [16] A. Vijayakumar, S. Vairavasundaram, YOLO-based Object Detection Models: A Review and its Applications, *Multimed Tools Appl* 83 (2024) 83535–83574. <https://doi.org/10.1007/S11042-024-18872-Y/TABLES/4>.
- [17] J. Redmon, A. Farhadi, YOLOv3: An Incremental Improvement, (2018). <https://doi.org/10.48550/arxiv.1804.02767>.
- [18] C.-Y. Wang, I.-H. Yeh, H.-Y.M. Liao, YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information, (2024). [https://doi.org/10.1007/978-3-031-72751-1\\_1](https://doi.org/10.1007/978-3-031-72751-1_1).
- [19] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, G. Ding, YOLOv10: Real-Time End-to-End Object Detection, *ArXiv.Org* (2024). <https://doi.org/10.48550/ARXIV.2405.14458>.
- [20] C.Y. Wang, A. Bochkovskiy, H.Y.M. Liao, Scaled-YOLOv4: Scaling cross stage partial network, *ArXiv* (2020).
- [21] C.Y. Wang, A. Bochkovskiy, H.Y.M. Liao, Scaled-YOLOv4: Scaling Cross Stage Partial Network, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2020) 13024–13033. <https://doi.org/10.48550/arxiv.2011.08036>.
- [22] X. Liu, X. Guo, Y. Liu, Y. Yuan, Consolidated domain adaptive detection and localization framework for cross-device colonoscopic images, *Med Image Anal* 71 (2021). <https://doi.org/10.1016/j.media.2021.102052>.
- [23] M. Misawa, S. ei Kudo, Y. Mori, K. Hotta, K. Ohtsuka, T. Matsuda, S. Saito, T. Kudo, T. Baba, F. Ishida, H. Itoh, M. Oda, K. Mori, Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video), *Gastrointest Endosc* 93 (2021) 960-967.e3. <https://doi.org/10.1016/j.gie.2020.07.060>.
- [24] J. Catlow, L. Sharp, J. Wagnild, L. Lu, R. Bhardwaj-Gosling, E. Ogundimu, A. Kasim, M. Brookes, T. Lee, S. McCarthy, J. Gray, F.

- Sniehotta, R. Valori, C. Westwood, R. McNally, J. Ruwende, S. Sinclair, J. Deane, M. Rutter, Nationally Automated Colonoscopy Performance Feedback Increases Polyp Detection: The NED APRIQOT Randomized Controlled Trial, *Clinical Gastroenterology and Hepatology* 22 (2024) 1926–1936. <https://doi.org/10.1016/j.cgh.2024.03.048>.
- [25] R. T.P, J. Kumar, S.R. Balasundaram, DeepCPD: deep learning with vision transformer for colorectal polyp detection, *Multimed Tools Appl* 83 (2024) 78183–78206. <https://doi.org/10.1007/s11042-024-18607-z>.
- [26] L. Wu, J. Liu, H. Yang, B. Huang, H. Liu, S. Cheng, Small gastric polyp detection based on the improved YOLOv5, *Multimed Tools Appl* 83 (2024) 71773–71788. <https://doi.org/10.1007/s11042-024-18497-1>.
- [27] M. Sherif Naguib, A. Khairy, H. Shehab, H. Abosheishaa, A. Meguid Kassem, The impact of EndoCuff-assisted colonoscopy on the polyp detection rate: A cross-over randomized back-to-back study, *Arab Journal of Gastroenterology* 25 (2024) 102–108. <https://doi.org/10.1016/j.ajg.2023.11.007>.
- [28] G. Sushama, G.C. Menon, Flexible Colon Polyp Detection: A Dual Mode Approach for Detection and Segmentation of Colon Polyps with Optional Inpainting for Specular Highlight Mitigation, *SN Comput Sci* 5 (2024). <https://doi.org/10.1007/s42979-024-02932-z>.
- [29] Y. Cakmak, S. Safak, M.A. Bayram, I. Pacal, Comprehensive Evaluation of Machine Learning and ANN Models for Breast Cancer Detection, *Computer and Decision Making: An International Journal* 1 (2024) 84–102. <https://doi.org/10.59543/COMDEM.V1I.10349>.
- [30] Y. Cakmak, I. Pacal, Enhancing Breast Cancer Diagnosis: A Comparative Evaluation of Machine Learning Algorithms Using the Wisconsin Dataset, *Journal of Operations Intelligence* 3 (2025) 175–196. <https://doi.org/10.31181/JOPI31202539>.
- [31] A. Karaman, I. Pacal, A. Basturk, B. Akay, U. Nalbantoglu, S. Coskun, O. Sahin, D. Karaboga, Robust real-time polyp detection system design

- based on YOLO algorithms by optimizing activation functions and hyper-parameters with artificial bee colony (ABC), *Expert Syst Appl* 221 (2023) 119741. <https://doi.org/10.1016/J.ESWA.2023.119741>.
- [32] I. Pacal, A. Karaman, D. Karaboga, B. Akay, A. Basturk, U. Nalbantoglu, S. Coskun, An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets, *Comput Biol Med* 141 (2022) 105031. <https://doi.org/10.1016/J.COMPBIOMED.2021.105031>.
- [33] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, G. Ding, YOLOv10: Real-Time End-to-End Object Detection, (2024). <https://github.com/THU-MIG/yolov10>. (accessed May 10, 2025).
- [34] R. Khanam, M. Hussain, YOLOv11: An Overview of the Key Architectural Enhancements, (2024). <http://arxiv.org/abs/2410.17725> (accessed May 16, 2025).
- [35] ETIS-LaribPolypDB, (n.d.). <https://www.kaggle.com/datasets/nguyenvoquocduong/etis-laribpolypdb> (accessed May 16, 2025).
- [36] Veritabanı - CVC-Colon, (n.d.). <https://pages.cvc.uab.es/CVC-Colon/index.php/databases/> (accessed February 15, 2025).
- [37] D. Jha, P.H. Smedsrud, M.A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, H.D. Johansen, Kvasir-SEG: A Segmented Polyp Dataset, (2019). <http://arxiv.org/abs/1911.07069>.
- [38] Cvc-Clinicdb - Polyp - Grand Challenge, (n.d.). <https://polyp.grand-challenge.org/CVCClinicDB/> (accessed February 15, 2025).
- [39] A. Mumuni, F. Mumuni, N.K. Gerrar, A Survey of Synthetic Data Augmentation Methods in Machine Vision, *Machine Intelligence Research* 2024 21:5 21 (2024) 831–869. <https://doi.org/10.1007/S11633-022-1411-7>.

- [40] Z. Wang, P. Wang, K. Liu, P. Wang, Y. Fu, C.-T. Lu, C.C. Aggarwal, J. Pei, Y. Zhou, A Comprehensive Survey on Data Augmentation, (2024). <https://arxiv.org/abs/2405.09591v3> (accessed May 16, 2025).