Jurnal Masyarakat Informatika, 17(1), 2026
DOI:10.14710/jmasif.17.1.78464
ISSN: 2777-0648

Comparative Analysis of Machine Learning for Stroke Classification Using YOLOv11 Detection and a Radiomics-Based Two-Stage Model

Wahyu Ozorah Manurung[*1)], Ernawati[1)], Widhia KZ Oktoeberza [1)] Desi Andreswari[1)], Endina Putri Purwandari[2)], Rusdi Efendi[3)]

[1)] *Informatics, Engineering Faculty, Universitas Bengkulu, Indonesia*
[2)] *Information Systems, Engineering Faculty, Universitas Bengkulu, Indonesia*
[3)] *Graduate School of Engineering, Tottori University, Japan*
[*] *Corresponding author: wahyuozorahmanurung@gmail.com*

Abstract

Stroke is a leading cause of disability and death worldwide, including in Indonesia. Rapid and accurate diagnosis is crucial, especially during the golden period (3–4.5 hours). CT scans are the primary imaging modality, but manual interpretation is often limited by time, subjectivity, and radiologist availability. This study proposes a two-stage model integrating YOLOv11 for lesion detection and machine learning for classification, using radiomics for feature extraction. In the first stage, YOLOv11 detects lesions and generates bounding boxes, which serve as Regions of Interest (ROIs). In the second stage, radiomics features are extracted and classified using Naïve Bayes, Support Vector Machine (SVM), and Random Forest. Results show YOLOv11 achieved an overall mAP@50 of 0.732, with the highest performance in hemorrhagic stroke (0.741). Radiomics-based classification further improves stability, achieving accuracies of 0.97–0.99 and precision, recall, and F1 scores≥0.94. Among classifiers, SVM performed best, with a test accuracy of 0.97, a false positive rate of 1.23%, total error 0.0218, generalization gap -0.0117, variance 0.0002, standard deviation 0.003635, confidence interval 0.9708 (+/-0.0073), and consistent fold accuracy between 96.5–97.5%, indicating stability without overfitting. These findings confirm that the combination of the YOLOv11 two-stage model, radiomics, and SVM provides a robust approach to support stroke diagnosis.

Keywords: *Digital Image Processing, Stroke, Radiomics, YOLOv11, Machine Learning.*

1 Introduction

Stroke is a serious global health problem, the leading cause of long-term disability and the second leading cause of death worldwide. According to the World Health Organization [1], stroke accounts for approximately 11% of total global deaths [2]. In Indonesia, stroke is the leading cause of death, surpassing heart disease [3] . Clinically, stroke occurs due to the disruption of blood supply to the brain, which can be caused by blockage (ischemic stroke) or bleeding (hemorrhagic stroke). Ischemic stroke, which accounts for approximately 87% of cases, is treated with thrombolytic drugs or endovascular procedures to dissolve blood clots [4],[5]. In contrast, hemorrhagic stroke focuses on stopping the bleeding and controlling the intracranial pressure[6]. An incorrect diagnosis can be fatal, as administering thrombolytic therapy in hemorrhagic cases can worsen the patient's condition.

The urgency of a rapid and accurate diagnosis is crucial, especially during the "golden time" (3–4.5 hours after symptom onset), when medical intervention offers the greatest opportunity to minimize brain damage and improve patient outcomes [7]. Head CT scans remain the gold standard for initial diagnosis due to their ability to quickly differentiate between ischemic (hypodense) and hemorrhagic

Received: 13 October 2025; Revised: 03 February 2026; Accepted: 19 February 2026; Published: 26 February 2026

(hyperdense) strokes [8]. However, in clinical practice, CT scan interpretation is often hampered by queues and radiologists' time constraints, prolonging diagnosis time and reducing the effectiveness of interventions. This situation drives the need for automated artificial intelligence-based systems to support diagnosis. One promising approach is the use of real-time object detection algorithms, such as the latest version of You Only Look Once (YOLOv11), which can rapidly localize lesion regions in medical images. In study [9], YOLOv11 is used specifically for stroke lesion localization rather than final classification, as accurate region detection is essential for guiding subsequent feature extraction and machine learning-based classification. With its ability to detect lesion areas through bounding boxes, YOLOv11 has the potential to accelerate stroke identification and support timely clinical decision-making.

However, bounding box-based detection alone is insufficient to provide in-depth classification information. Clinical diagnosis requires a more detailed analysis of brain tissue characteristics. In this context, radiomics is crucial, as it allows the extraction of quantitative features from medical images, including first-order statistical features, shape, and texture that are difficult to identify by the human eye but have significant clinical value [10], [11]. Recent research has shown that integrating radiomics with machine learning algorithms can improve the accuracy of clinical predictions. For example, study [12] combined radiomics with deep learning on longitudinal CT scans to predict tuberculosis treatment outcomes, resulting in a significant performance improvement when the two approaches were combined. This indicates that integrating radiomics and deep learning can strengthen the capabilities of medical decision support systems.

In CT image-based stroke classification, previous research has been dominated by deep learning approaches using conventional CNNs. For example, several studies have used EfficientNet or ResNet architectures, supported by augmentation and classical feature extraction techniques such as GLCM, and have successfully achieved high accuracy [13–15]. Other studies have also demonstrated the importance of data preprocessing, such as augmentation and noise removal, to improve model performance. Furthermore, YOLO has been extensively tested in other medical domains, such as retinal pathology detection [16] and MRI-based brain tumor detection [17], demonstrating superior performance in detecting small lesion areas with high computational efficiency. However, the specific application of YOLO to stroke detection in CT images remains very limited, especially when integrated with radiomics as a subsequent classification step.

This review reveals that most previous studies have focused on conventional CNNs and classical feature extraction methods, whereas the use of real-time object detection algorithms, such as YOLOv11, and their integration with radiomics has not been widely explored. This gap presents an opportunity to introduce a new, more effective approach to CT scan-based stroke diagnosis. Therefore, based on the identified research gap, this study aims to develop and evaluate a two-stage applied framework for CT-based stroke analysis, in which YOLOv11 is used exclusively for stroke lesion localization and ROI generation, followed by radiomics-based feature extraction and conventional machine learning classification. The main contribution of this study lies in its contextual application and in the comparative evaluation of three machine learning classifiers under two settings, using radiomics features and without radiomics to assess the impact of quantitative feature extraction on stroke classification performance.

2 Literature Review

Research on CT image-based stroke classification has advanced rapidly alongside advances in computer vision and machine learning. Most early approaches focused on conventional CNN-based deep learning. For example, architectures like EfficientNet and ResNet have been shown to extract visual features from CT images and achieve high accuracy in classifying ischemic and hemorrhagic strokes. Studies by Putri et al [13], demonstrated the good performance of CNNs, while research by [15] emphasized the critical role of preprocessing, such as augmentation and denoising, in improving input quality. However, conventional CNN models have limitations in inference speed and transparency of the classification process, making them less than optimal for real-time clinical applications.

On the other hand, research has also begun to explore radiomics feature extraction as a complement to medical image analysis. Radiomics transforms images into quantitative data through statistical features, shape, and texture, which can be used for both detection and classification. Study [12] demonstrated that the combination of radiomics and deep learning improves the accuracy of clinical Detection and classification, while [18] demonstrated that CT-based radiomics can achieve an AUC of >0.90 in predicting brain metastases. Despite its potential, radiomics faces challenges, including variability in image quality and the need for precise ROI segmentation.

Furthermore, real-time object detection algorithms such as YOLO are increasingly being applied in the medical domain to detect lesions and anomalies. Studies by Ardelean [16] on retinal pathology and Monisha & Rahman [17] in MRI-based brain tumor detection, YOLO can identify small regions with high efficiency, even under complex imaging conditions. However, the use of YOLO for stroke cases on CT images remains very limited. Among the YOLO family, YOLOv11 represents a recent architectural advancement that improves detection accuracy while maintaining real-time inference speed. Although YOLOv8 and YOLOv10 are more commonly used as models in studies. Compared to earlier versions, YOLOv11 incorporates more efficient feature extraction and attention mechanisms, enabling better localization of lesion regions in complex medical images [19]. Based on recent studies, YOLOv11 offers a favorable balance between accuracy and computational efficiency, making it suitable for time-critical clinical scenarios such as stroke diagnosis [20]. Therefore, YOLOv11 is adopted in this study specifically for stroke lesion localization rather than end-to-end classification.

For further classification, various machine learning algorithms, including Naïve Bayes, Random Forests, and SVMs, have also been used. Dharshini et al [15] compared five ML algorithms and found that deep learning had the highest accuracy, but classical models such as Random Forest remained competitive and easier to interpret. This suggests that the choice of classification algorithm should be tailored to the type of features and clinical needs.

Previous studies show that most research still focuses on conventional CNNs or classical feature-extraction methods. This approach demonstrates good accuracy but remains limited in terms of computational efficiency and generalizability when applied to clinical cases requiring rapid diagnosis. Meanwhile, YOLO-based research for stroke analysis using CT scans is still very rare, even though this algorithm has proven superior in other medical domains such as brain tumor detection and retinal analysis. This limitation highlights a significant research gap, particularly in the application of YOLOv11 as a more advanced brain lesion detection model.

Therefore, this study contributes by presenting a novel two-stage model-based approach that combines the speed of real-time object detection with YOLOv11 and the depth of quantitative

radiomics analysis for stroke classification using machine learning algorithms. The main contributions of this study are introducing the underexplored integration of YOLOv11 and radiomics for stroke diagnosis, conducting a systematic comparison of three machine learning algorithms to identify the optimal classification model, and offering support for clinical decision-making during the critical golden time period. This combination is expected to provide a new foundation for the development of intelligent medical image-based diagnostic systems.

3 Research Methods

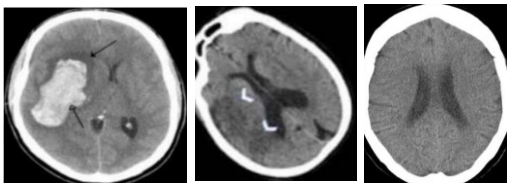
Deep learning-based object detection is increasingly used in medical image analysis due to its fast and accurate performance. One prominent method is You Only Look Once (YOLO), particularly the latest version, YOLOv11, which can detect stroke lesions in head CT scan images in real time using bounding boxes [16]. The detection results are then further refined using radiomics feature extraction, a process that converts medical images into quantitative data representing tissue intensity, texture, and shape. This approach is important because it captures morphological details of the brain that are difficult to identify visually, while also strengthening the basis for subsequent classification [10], [18].

The obtained radiomics features are then processed through three machine learning algorithms: Naïve Bayes, Support Vector Machine (SVM), and Random Forest. Naïve Bayes is efficient and effective, SVM excels at separating classes with optimal margins, and Random Forests provide stable results on complex data. The final stage of this research is a performance evaluation using accuracy, precision, recall, F1-score, and mAP metrics for object detection. With this framework, this research is an experimental study focused on the development and evaluation of an automated system based on a two-stage model. This approach was chosen because it enables the direct application of artificial intelligence algorithms to solve real-world problems in the medical field, with a focus on building the YOLOv11 architecture for detecting stroke lesions and extracting radiomic features to support machine learning classification.

3.1 Research Data

The dataset used in this study is sourced from the Brain Stroke CT Dataset available on the Kaggle platform. This dataset was collected in Istanbul in 2021 using data provided by the Ministry of Health of the Republic of Turkey through the Directorate General of Health Information Systems and is managed by the Turkish Health Institute (TUSEB). It consists of three classes: ischemic stroke, hemorrhagic stroke, and no stroke, as shown in Figure 1. All images in the Brain Stroke CT Dataset are fully anonymized and publicly available for research purposes. Since the dataset does not contain any identifiable patient information, ethical approval and individual patient consent were not required, in accordance with the data usage policies of the Kaggle platform and the data provider.

From the original dataset, 2,431 CT images were selected for this study, comprising 666 no-stroke cases, 973 hemorrhagic stroke cases, and 792 ischemic stroke cases. Lesion regions were manually annotated by the authors using the overlay masks provided in the original dataset as guidance for delineating region of interest (ROI) boundaries. These annotations were subsequently used for YOLOv11-based lesion localization and radiomics feature extraction.



(a) (b) (c)

Figure 1 Sample data set: (a) Hemorrhagic; (b) Ischemic; (c) Normal

3.2 Proposed Stage

Figure 2 shows the proposed research design aims to evaluate the performance of a two-stage framework in which YOLOv11 is employed for stroke lesion detection and region-of-interest (ROI) localization, followed by radiomics-based feature extraction and machine-learning-based stroke classification. In this study, YOLOv11 is utilized specifically for lesion detection and ROI localization, exploiting its ability to learn high-level semantic representations for accurate spatial identification [19]. However, deep features extracted by YOLOv11 are primarily optimized for visual discrimination and are often limited in clinical interpretability. To address this limitation, radiomics feature extraction is applied to the detected ROIs to obtain handcrafted quantitative descriptors of lesion morphology, intensity, and texture, which have been shown to provide clinically meaningful and interpretable information [21]. This process does not constitute redundancy, as YOLOv11 and radiomics operate at different representational levels. This decoupled strategy follows prior radiomics and hybrid learning studies, where traditional machine learning classifiers are preferred for processing handcrafted features due to their stability and robustness in medical imaging tasks [22].

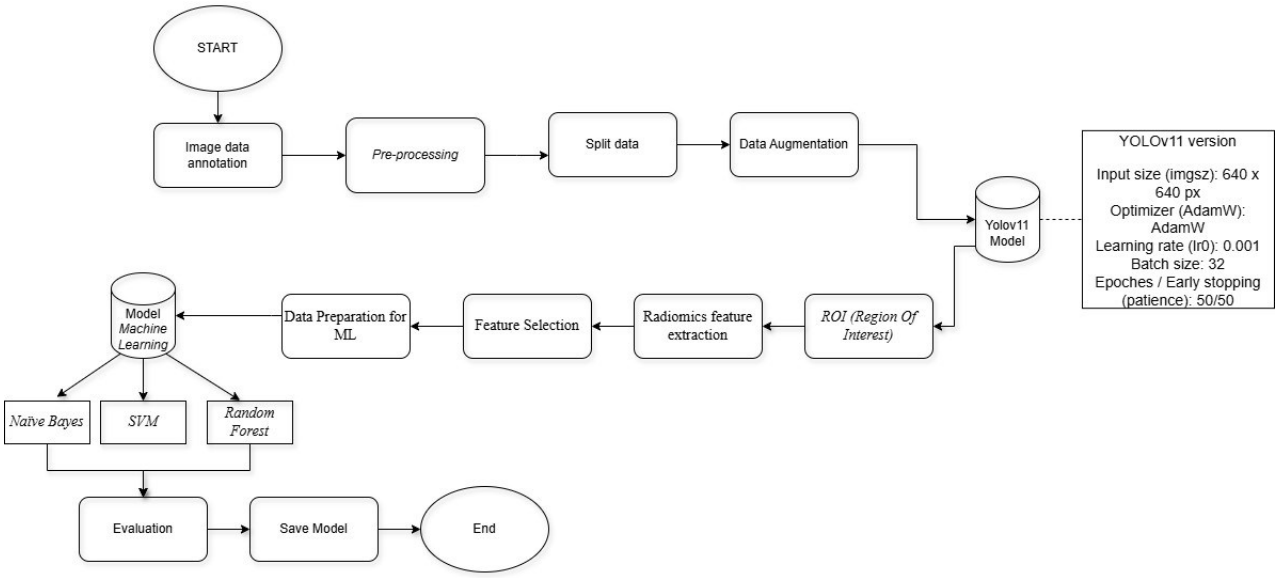


Figure 2 Research Design Flowchart

1. Preprocessing

The annotated images undergo a preprocessing stage to match the YOLOv11 input requirements. Auto orientation is applied to ensure consistent image alignment, adaptive equalization is used to enhance contrast and highlight lesion details, and images are resized to 640 × 640 to ensure uniform input dimensions and stable model training.

2. Split Data and Augmentation

The original dataset consisted of 2,431 CT images. To prepare the data for model training and evaluation, the data was first split into training, validation, and test sets with ratios of 80%, 10%, and 10%, respectively. This resulted in 1,945 images for training, 243 images for validation, and 243 images for testing.

To enrich data variety and improve the model's generalization, data augmentation was applied only to the training set. Augmentation techniques included horizontal flipping, rotation between -20° and +20°, brightness adjustment between -15% and +15%, and exposure adjustment between -10% and +10%. After augmentation, the training set contained 5,835 images, which were generated by applying a 3× augmentation factor to the original 1,945 training images.

3. Model Architecture YOLOv11

In this stage, the YOLOv11 model is configured for CT scan lesion detection with an input size of 640 × 640 px to preserve spatial detail, and the AdamW optimizer is used for stable convergence and improved generalization. A learning rate of 0.001 and a batch size of 32 are selected to balance training stability and computational efficiency, and the model is trained for up to 50 epochs with early stopping to prevent overfitting. These settings enable fast and accurate object detection on the dataset.

4. Region of Interest

YOLO detects and generates a bounding box. The bounding box results are then forwarded to the Region of Interest (ROI) stage, which is an area in the CT scan image that potentially indicates a stroke. This area becomes the main focus for the subsequent feature Radiomics process.

5. Radiomic Feature Extraction and Selection Feature

The ROIs from YOLO detection are then extracted into radiomic features, a set of quantitative characteristics (e.g., texture, shape, pixel intensity, image filter transformation) that represent the characteristics of a stroke in the image. After that, feature selection was performed to identify the most relevant features for training the selected model using RFECV with 5-fold cross-validation analysis. This step is shown in Figure 3.

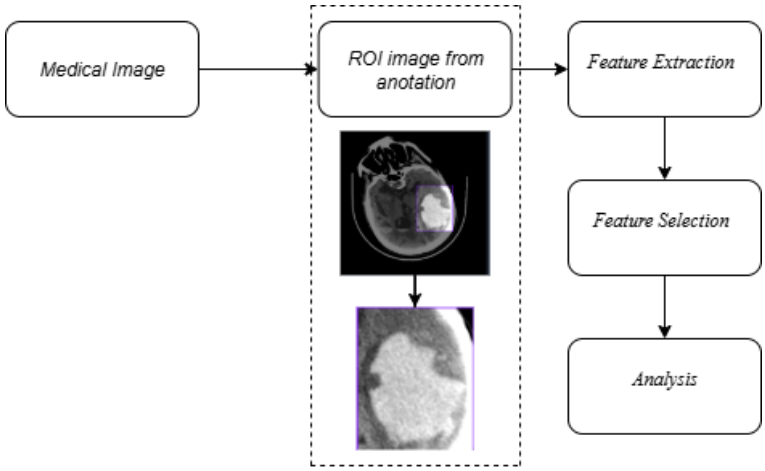


Figure 3 Radiomics Workflow

The application of radiomix in medical diagnostics follows a standardized workflow that transforms visual images into quantitative data. The following are the general process steps:

a) Acquisition & Segmentation: Capture medical images (CT/MRI) and determine specific areas (ROIs) to be studied, such as lesion locations.
b) Feature Extraction: Convert visual images into numerical data based on tissue texture and intensity.
c) Feature Selection: Filter the data to select the most relevant features to avoid overfitting.
d) Predictive Analysis: Process the selected data with Machine Learning for rapid classification, for example, accurately distinguishing stroke types.

6. Tabular Data Preprocessing (Feature and Label Merge)

Before being fed into the machine learning algorithm, additional preprocessing is performed: combining radiomic features with annotation labels, converting categorical labels into numeric form (label encoding), and separating the variables X (radiomic features) and y (stroke class).

7. Machine Learning Modeling

The extracted radiomic features were used as input for three machine learning algorithms: Naïve Bayes, Support Vector Machine, and Random Forest. Random Forest, a tree-based ensemble method, handles non-linear data; SVM, a margin-based method, excels in high-dimensional feature spaces; and Naïve Bayes, a fast probabilistic approach, serves as a baseline. The parameters used for each algorithm are adjusted to obtain optimal performance.

Table 1 Hyperparameter configuration of the evaluated machine learning models

Random Forest	SVM	Gaussian Naive Bayes
n_estimators=200	kernel='rbf'	GaussianNB()
max_depth=8	C=1.0	StandardScaler()
random_state=42	random_state=42	
n_jobs=-1	probability=True	
	StandardScaler()	

Specific hyperparameters were selected to optimize generalization on high-dimensional radiomics data and ensure a fair comparison across models as shown in Table 1. Random Forest used 200 estimators with a limited tree depth (max_depth = 8) to balance model capacity and control overfitting. SVM employed an RBF kernel to capture non-linear patterns in radiomic features, with C = 1.0 providing a balanced trade-off between margin and error. StandardScaler was applied to SVM and Naïve Bayes to normalize feature scales. Overall, hyperparameters were chosen based on common practice and preliminary empirical testing to promote and reduce overfitting.

8. Evaluation

The evaluation phase assessed the system's overall performance at both the detection and classification stages. In the detection phase using YOLOv11, the evaluation focused on mAP@50 and mAP@50-95, the primary evaluation metrics for object detection models. These two metrics assess detection accuracy based on the level of overlap tolerance between the predicted bounding

box and the ground truth, thus demonstrating the model's ability to precisely locate stroke lesions. Furthermore, the confidence score parameter was analyzed to assess the model's confidence in its detection results.

Next, in the classification phase using machine learning namely: Naïve Bayes, Support Vector Machine, and Random Forest, evaluation was conducted using accuracy, precision, recall, and F1-score metrics. Precision and recall were selected to assess the model's accuracy and its ability to identify all positive cases. The F1-score was used to provide a balanced assessment of precision and recall, especially in imbalanced datasets. This evaluation provides an overview of the model's performance in classifying stroke types as ischemic, hemorrhagic, and normal.

9. Save the Model

The concluding step involves saving the trained and evaluated model. The model is stored in both 'joblib' and 'best.pt' formats to facilitate future deployment, additional testing, or further development.

3.3 Model Evaluation

In this study, performance evaluation was conducted using several metrics commonly used in classification and object detection. Accuracy is the most common evaluation metric in machine learning, measuring the proportion of correct predictions to the total test data. This metric provides an overview of how often the model makes correct decisions, covering both true positives and true negatives, making it a simple yet widely used measure in classification evaluation.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision measures the accuracy of the model's positive predictions against the actual class, indicating how reliable the model is in identifying stroke cases. A high precision value means positive predictions are more trustworthy, which is particularly important in medical diagnosis to minimize false positives and their potential clinical impact.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

Recall (Sensitivity) assesses the model's ability to correctly identify all positive cases. It is particularly important in medical contexts, where missing a true case can be critical. In stroke detection, high recall ensures that patients with stroke are not overlooked.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

F1-Score balances precision and recall by combining them into a single metric. It provides a harmonic mean that reflects both accuracy and completeness of predictions. This makes it especially useful for fair evaluation on imbalanced datasets.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

A confusion matrix is a table that compares model predictions with the actual labels, showing correct and incorrect predictions. In this study, the FPR was used to assess how often the model produced false-positive predictions.

$$\text{FPR} = \frac{\text{FP}}{\text{FP+TN}} \tag{5}$$

Based on the formula, where TP, TN, FP, and FN indicate true positives, true negatives, false positives, and false negatives, respectively. Next, an evaluation will be carried out using the Empirical Risk & Generalization Gap. Statistical analysis in machine learning model evaluation often focuses on the Empirical Risk R and the Generalization Gap G. Empirical Risk Minimization (ERM) is an approach that measures model performance based on training data, while the Generalization Gap describes the difference between the error on training data (empirical risk) and the error on unseen data (expected risk). In theory, the smaller the generalization gap, the better the model's ability to generalize to new data, rather than simply memorizing the training data [23]. With this combination of metrics, the study not only assesses the model's classification ability but also the accuracy and consistency of stroke lesion detection in CT scan images as shown in Equations (6) and (7).

$$R = \frac{1}{m}\sum_{i=1}^{m} \text{l}(h(x_i), y_i) \tag{6}$$

$$\text{G} = R_{\text{test}} - R_{\text{train}} \tag{7}$$

where m denotes the total number of samples, x_i and y_i represent the input data and corresponding ground truth labels, $h(\cdot)$ denotes the trained model (hypothesis), and $l(\cdot)$ represents the loss function.

And also an evaluation was carried out to see the bias variance as shown in Equation (8). Bias is an error caused by a model that is too simple so that it fails to capture data patterns (underfitting), while variance is an error caused by a model that is too sensitive to training data so that it is unstable on new data (overfitting) so this is very good for research to be accurate in predicting.

$$E_{S,\epsilon}\left[\left(y - \hat{f}(x)\right)^2\right] = \underbrace{\left(E_S[\hat{f}(x)] - f(x)\right)^2}_{\text{Bias}^2} + \underbrace{E_S\left[\left(\hat{f}(x) - E_S[\hat{f}(x)]\right)^2\right]}_{\text{Variance}} + \underbrace{\sigma^2}_{\text{Irreducible Noise}} \tag{8}$$

where $\hat{f}(x)$ is the model prediction, f(x) is the true underlying function, E_S [·] denotes the expectation over different training sets, and σ² represents the irreducible noise.

In addition to classification metrics, detection performance was evaluated using Mean Average Precision (mAP), the standard metric for object detection models such as YOLOv11. mAP is computed as the average of Average Precision (AP) across classes, reflecting the model's consistency in detecting different stroke types.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{9}$$

where N denotes the number of classes and AP_i represents the average precision of the i-th class. With this combination of metrics, the study not only assesses the model's classification ability but also the accuracy and consistency of stroke lesion detection in CT scan images.

4 Results and Discussion

In this study, a two-stage model was developed to detect and classify three stroke types in head CT images: ischemic stroke, hemorrhagic stroke, and normal (no stroke). This approach was chosen because it combines the strengths of deep learning–based object detection with quantitative radiomics analysis, resulting in a system that is more robust, faster, and more accurate than either method alone. In the first stage, brain lesion areas were detected using YOLOv11, which is known for its high inference speed and detection accuracy. The detected lesion areas were then processed in the ROI stage to derive radiomics features (texture, intensity, and shape), which were converted to numerical representations. These features were used as input for three different algorithms: Naïve Bayes (probabilistic), SVM (margin-based), and Random Forest (ensemble). The evaluation was comprehensive: YOLOv11 was assessed using mAP@50, while the classifiers were evaluated using accuracy, precision, recall, F1-score, and confusion matrix, along with additional statistical analyses such as Empirical Risk, Generalization Gap, and the Bias-Variance trade-off. This methodology not only ensures technical accuracy but also highlights the strengths and limitations of each algorithm.

4.1 Stroke area annotation

In the initial stage, ischemic and hemorrhagic stroke regions were annotated as Regions of Interest (ROIs) using bounding boxes in YOLOv11 format to enable lesion localization. Normal images were included without annotations to serve as negative samples, helping the model distinguish between normal and abnormal cases and reduce false-positive detections. Examples of the annotation results are shown in the following Figure 4.

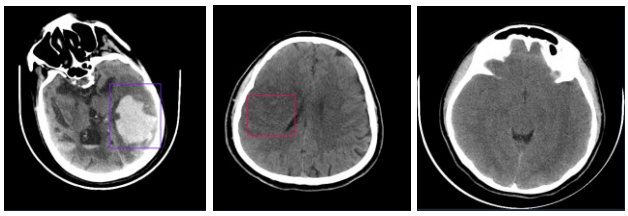


Figure 4 Area annotation results on CT scan images for stroke region detection

4.2 Model Yolov11

After the annotation was completed, preprocessing, augmentation, and the dataset were split into training, validation, and test sets for objective evaluation. YOLOv11 was then trained with the training configuration described in the previous section. YOLOv11 converged within 50 epochs, achieving stable detection performance as shown in Table 2.

Table 2 Results of the YOLO model train

Class	Images	Instances	mAP50
all	243	169	0.732
Hemoragik	97	97	0.741
Iskemik	72	72	0.722

The YOLOv11 testing results demonstrate effective stroke lesion detection on CT scan images, achieving an overall mAP@50 of 0.732 across 243 images with 169 annotated instances. The hemorrhagic stroke class achieved the highest detection performance, with a mAP@50 of 0.741 across

97 images, reflecting the model's strong ability to identify high-contrast bleeding regions. In comparison, ischemic stroke obtained a slightly lower mAP@50 of 0.722 across 72 images.

Compared to previous studies, these results are quite good. The study by Liu [24] which also used CT scan images, reported mAP@50 values of 0.594 on various multicenter validation datasets, indicating that this study's results are comparable to those of extensively validated detection models. Meanwhile, a benchmarking study by Tapia [25] on YOLOv5 to YOLOv10 showed more variable mAP@50 performance: in internal testing, the best value was only around 0.55, and in external testing, it dropped significantly to the 0.2–0.4 range. Therefore, the mAP@50 of 0.732 achieved in this study confirms that YOLOv11 is capable of providing competitive detection performance for CT scan images, and efficiency of newer attention mechanisms, especially when compared to previous studies.

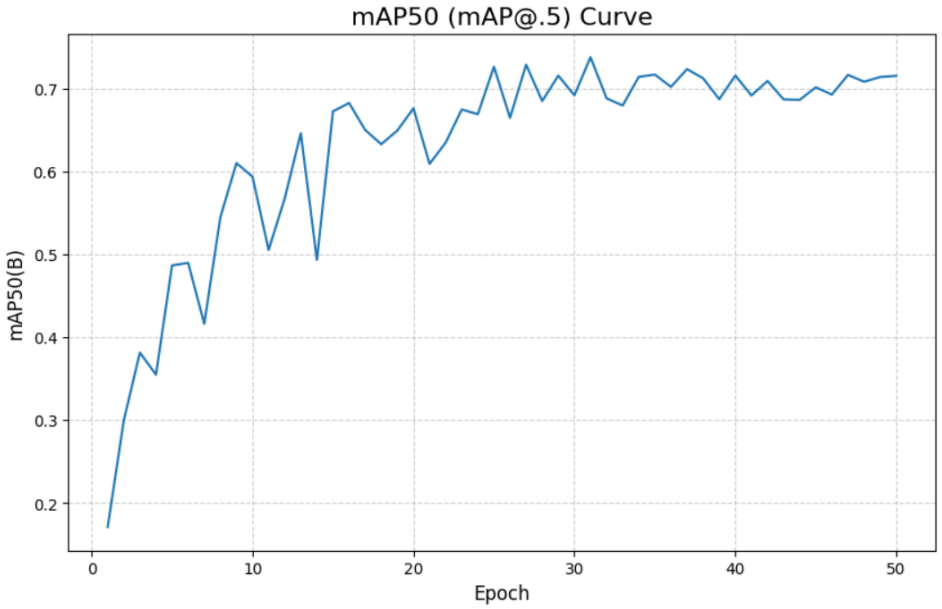
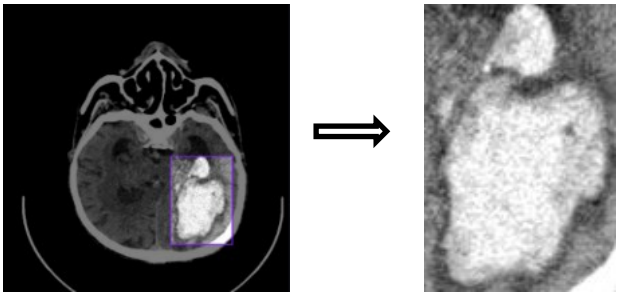


Figure 5 mAP50 (mAP@0.5) curve over training epochs for the YOLOv11 detection model

Figure 5 illustrates the mAP@50 performance of YOLOv11 over 50 training epochs. Detection performance improves rapidly in the early training phase, rising from approximately 0.17 at epoch 1 to approximately 0.60 at epoch 10. Between epochs 25 and 40, mAP@50 values range from 0.69 to 0.74, with peak performance around epochs 27–31. After epoch 30, performance stabilizes around 0.70, indicating that the model has reached an optimal learning point and that further training yields diminishing returns or no improvement. This behavior indicates mild overfitting, motivating the incorporation of radiomics feature extraction to improve classification stability and discriminatory performance beyond detection alone.

4.3 Region Of Interest

Detection uses YOLOv11 to determine the Region of Interest (ROI). ROI represents the area on the CT scan image containing stroke lesions, both ischemic and hemorrhagic. By using the bounding boxes of the detected results, the ROI is isolated from other irrelevant image regions as shown in Figure 6. This process is important because it allows the analysis to focus solely on the brain areas affected by stroke, thereby reducing noise and increasing accuracy during the radiomics feature stage. Thus, ROI serves as a bridge between object detection and subsequent classification.



(Detection) (ROI)

Figure 6 Region of interest (ROI) extraction from the YOLOv11 detection output

4.4 Radiomics feature

In this study, radiomic features were extracted from head CT scan images to represent the quantitative characteristics of stroke lesions. Generally, radiomics consists of four categories of features: shape-based, first-order intensity statistics, texture features, and image filter transformation results. However, this study focuses only on first-order intensity statistics and texture features, with a total of 93 extracted features. Feature selection was performed using RFECV analysis.

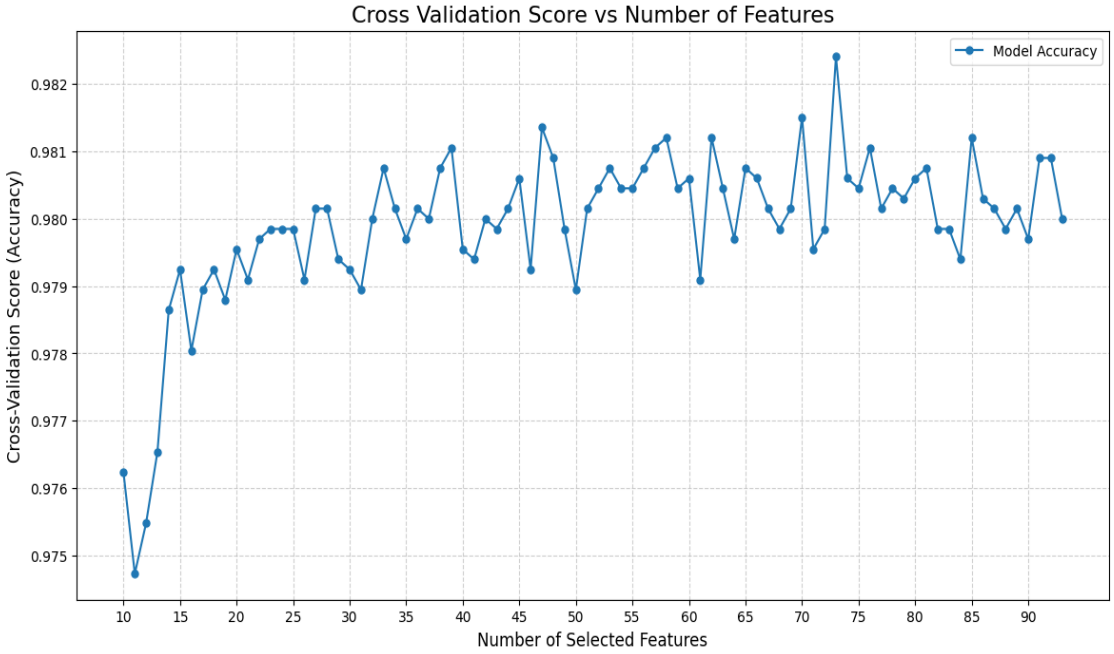


Figure 7 Cross-validation accuracy as a function of the number of selected features using RFE

The feature selection process was conducted using Recursive Feature Elimination (RFE) with cross-validation-based accuracy evaluation. Figure 7 showed that model accuracy increased substantially up to approximately 47 features and then stabilized, with no meaningful improvement from adding more features. Therefore, 47 features were selected as optimal, balancing high accuracy and model efficiency while reducing the risk of overfitting. The selected features are listed in Table 3.

4.5 Model Performance

The objective of this study was to develop a two stage model based on YOLOv11 and radiomics to detect and classify strokes in head CT scan images. Unlike previous studies that often used

conventional CNNs, this analysis approach combines object detection and radiomics. Model performance was tested comparatively using Naive Bayes, SVM, and Random Forest, while also being compared with a model without radiomics features to assess its contribution to improving classification accuracy.

Table 3 List of 47 radiomic features categorized into first-order intensity and texture features

First-order intensity statistics	Texture features
original_firstorder_90Percentile, *original_firstorder_Energy,* *original_firstorder_Entropy,* *original_firstorder_InterquartileRange,* *original_firstorder_Kurtosis,* *original_firstorder_MeanAbsoluteDeviation,* *original_firstorder_Range,* *original_firstorder_Robust,MeanAbsoluteDeviation* *original_firstorder_Skewness,* *original_firstorder_TotalEnergy,* *original_firstorder_Uniformity,* *original_firstorder_Variance,*	GLCM (11 feature)*: original_glcm_ClusterProminence, original_glcm_ClusterShade, original_glcm_ClusterTendency, original_glcm_Correlation, original_glcm_Imc1, original_glcm_Imc2, original_glcm_JointEnergy, original_glcm_JointEntropy, original_glcm_MCC, original_glcm_SumEntropy, original_glcm_SumSquares* GLRLM (7 feature)*: original_glrlm_GrayLevelNonUniformity, original_glrlm_GrayLevelNonUniformityNormalized, original_glrlm_GrayLevelVariance, original_glrlm_LongRunEmphasis, original_glrlm_LongRunLowGrayLevelEmphasis, original_glrlm_RunLengthNonUniformity, dan original_glrlm_RunVariance* GLSZM (10 feature)*: original_glszm_GrayLevelNonUniformity, original_glszm_LargeAreaEmphasis, original_glszm_LargeAreaLowGrayLevelEmphasis, original_glszm_SizeZoneNonUniformity, original_glszm_SizeZoneNonUniformityNormalized, original_glszm_SmallAreaEmphasis, original_glszm_SmallAreaHighGrayLevelEmphasis, original_glszm_ZoneEntropy, original_glszm_ZonePercentage, dan original_glszm_ZoneVariance.* GLDM (4 feature)*: original_gldm_DependenceEntropy, original_gldm_DependenceNonUniformity, original_gldm_DependenceNonUniformityNormalized, dan original_gldm_GrayLevelNonUniformity.* *NGTDM(3 feature)* *original_ngtdm_Coarseness, original_ngtdm_Complexity, dan original_ngtdm_Strength.*

Table 4 Classification performance of the Random Forest model based on the Radiomics feature

Data Split	Class	Precision	Recall	F1 Score	Accuracy
Test Set	Normal	1.00	1.00	1.00	0.99
	Hemorrhagic	0.99	0.97	0.98	
	Ischemic	0.96	0.99	0.97	
Train Set	Normal	1.00	1.00	1.00	0.99
	Hemorrhagic	0.99	0.99	0.99	
	Ischemic	0.98	0.98	0.98	

The Random Fores model achieved 0.99 accuracy on test data as shown in Table 4, with precision and recall ranging from 0.96 to 1.00 across classes. The normal class achieved a perfect score (Precision, Recall, F1 = 1.00), while the hemorrhagic and ischemic classes performed slightly lower, with F1 scores of 0.98 and 0.97, respectively. On training data, accuracy further improved to 0.99, with all metrics remaining consistently high above 0.98.

Table 5. Classification performance of the SVM model based on the Radiomics feature

Data Split	Class	Precision	Recall	F1 Score	Accuracy
Test Set	Normal	1.00	1.00	1.00	0.99
	Hemorrhagic	0.99	0.97	0.98	
	Ischemic	0.96	0.99	0.97	
Train Set	Normal	1.00	1.00	1.00	0.97
	Hemorrhagic	0.96	0.97	0.97	
	Ischemic	0.96	0.95	0.95	

Table 5 shows that the SVM model achieved 0.99 accuracy on the test set, with precision, recall, and F1 score values above 0.96 for all classes. On the train set, accuracy increased to 0.97, with consistently high metrics ≥ 0.95.

Table 6. Classification performance of the Naïve Bayes model based on the Radiomics feature

Data Split	Class	Precision	Recall	F1 Score	Accuracy
Test Set	Normal	0.99	1.00	1.00	0.97
	Hemorrhagic	0.96	0.97	0.97	
	Ischemic	0.96	0.94	0.95	
Train Set	Normal	1.00	1.00	1.00	0.95
	Hemorrhagic	0.94	0.94	0.94	
	Ischemic	0.91	0.90	0.91	

Table 6 shows that the Naïve Bayes model achieved an accuracy of 0.97 on the test set, with precision, recall, and F1 scores above 0.94. On the train set, the accuracy was 0.95, with consistently high precision, recall, and F1 scores exceeding 0.90.

Table 7. Performance results of the three models without the Radiomics feature on the Test Set

Models	Class	Precision	Recall	F1 Score	Accuracy
Random Forest	Normal	1.00	1.00	1.00	1.00
	Hemorrhagic	1.00	0.99	0.99	
	Ischemic	0.99	1.00	0.99	
SVM	Normal	1.00	1.00	1.00	1.00
	Hemorrhagic	1.00	0.99	0.99	
	Ischemic	0.99	1.00	0.99	
Naïve Bayes	Normal	1.00	0.98	0.99	0.97
	Hemorrhagic	0.98	0.94	0.96	
	Ischemic	0.93	1.00	0.96	

On the test dataset, the Random Forest classifier achieved near-perfect performance, with an accuracy of 1.00 for the Normal and Hemorrhagic classes and 0.99 for the Ischemic class, while precision, recall, and F1-score were close to 1.00 across all classes, with a slight reduction for the Ischemic class, as shown in Table 7. The SVM model demonstrated consistently high performance, achieving accuracy, precision, recall, and F1-score values of 1.00 for the Normal and Hemorrhagic classes and 0.99 for the Ischemic class. In comparison, the Naïve Bayes model showed lower but still

competitive performance, with an overall accuracy of approximately 0.96, achieving strong precision and recall in the Normal class, moderate performance in the Hemorrhagic class, and reduced precision but perfect recall in the Ischemic class, indicating higher sensitivity but lower specificity for ischemic stroke detection.

Table 8. Performance results of the three models without the Radiomics feature on the Train Set

Models	Class	Precision	Recall	F1 Score	Accuracy
Random Forest	Normal	1.00	1.00	1.00	1.00
	Hemorrhagic	1.00	0.99	0.99	
	Ischemic	0.99	1.00	0.99	
SVM	Normal	1.00	1.00	1.00	1.00
	Hemorrhagic	1.00	1.00	1.00	
	Ischemic	1.00	1.00	1.00	
Naïve Bayes	Normal	0.97	0.96	0.97	0.90
	Hemorrhagic	0.89	0.87	0.88	
	Ischemic	0.84	0.87	0.86	

Based on Table 8, the Random Forest model demonstrated strong performance, achieving 1.00 in accuracy for the Normal and Hemorrhagic classes and 0.99 for the Ischemic class, with precision, recall, and F1-score values consistently above 0.99. The SVM model exhibited the most stable performance, achieving perfect scores of 1.00 for accuracy, precision, recall, and F1-score across all classes. In contrast, the Naïve Bayes classifier showed lower performance, with an accuracy of 0.90. It achieved higher precision in the Normal class at 0.97, but reduced precision in the Hemorrhagic class at 0.87 and in the Ischemic class at 0.88. Recall ranged from 0.84–0.97, and the F1-score ranged from 0.86–0.88, indicating higher variability compared to the other models.

4.6 Discussion

4.6.1 Comparison using radiomics extraction with no radiomics

Based on the findings presented in Tables 4 to 8, models that incorporate radiomics features consistently demonstrate more balanced and stable performance than those trained without them. As illustrated in Tables 4 to 6, the use of radiomics features enabled Random Forest and Support Vector Machine classifiers to achieve high test accuracies (0.97–0.99), with closely aligned training accuracies and F1 scores (0.94–0.99). This indicates improved generalization across Normal, Hemorrhagic, and Ischemic classes. The enhanced stability is attributable to radiomics encoding clinically meaningful tissue characteristics rather than raw pixel patterns, thereby reducing sensitivity to dataset-specific noise and mitigating overfitting risks. Texture features effectively capture lesion heterogeneity, which is vital for representing structural variations within stroke regions, while intensity-based features reflect CT density differences that distinguish ischemic from hemorrhagic lesions. Conversely, models lacking radiomics, as depicted in Tables 7 and 8, exhibited near-perfect accuracy, suggestive of overfitting driven by dataset-specific pixel distributions instead of robust pathological cues. This phenomenon is most apparent in the Naïve Bayes model, where the absence of radiomics resulted in diminished performance and imbalance. Overall, extracting radiomics features enhances model robustness and stability by providing structured, interpretable descriptors of lesion properties. This approach helps to mitigate overfitting, fosters better generalization, particularly for probabilistic models, and reinforces the reliability of ensemble and margin-based classifiers.

4.6.2 Comparison of model performance curves with feature extraction

The visualization of model performance with feature extraction in Figure 8 shows that Random Forest and SVM maintain a balance of precision, recall, and F1 scores on both training and test data, indicating good generalization. Meanwhile, Naïve Bayes still shows lower performance than the other two algorithms, especially in the ischemic class, but it still performs better when not using radiomics. This confirms that radiomics extraction not only improves accuracy but also provides consistent performance across classes across all algorithms.

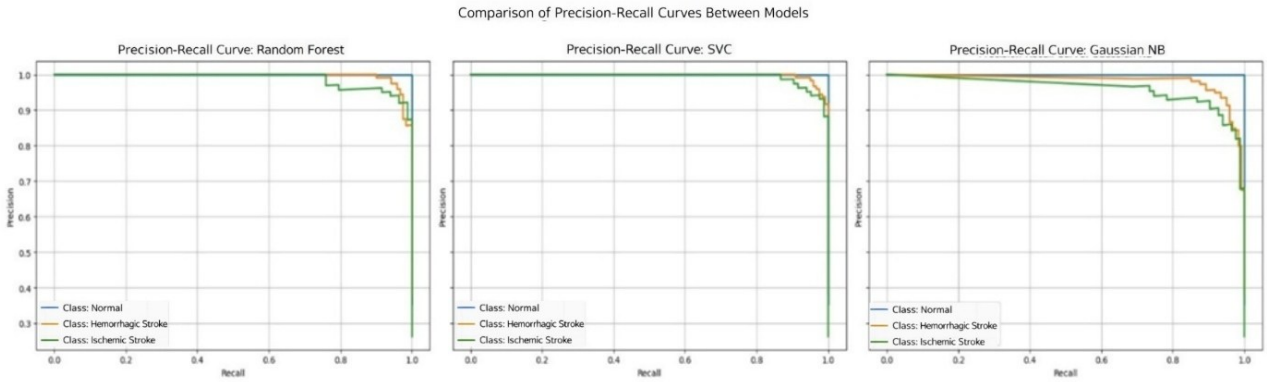


Figure 8 Comparison of model performance curves with the Radiomics feature

Based on the confusion matrix analysis, Random Forest achieved the lowest average false positive rate (FPR) at 1.23%, SVM at 1.23%, and Gaussian Naïve Bayes at 2.20%. Even though SVM and Random Forest have the same FPR, when considering other evaluation metrics such as accuracy, precision, recall, and F1-score, SVM consistently outperformed the others. Moreover, the SVM's FPR remained below the 2% threshold, which is often recommended in medical contexts. This finding is consistent with the literature, which states that an extremely low FPR under 2% is only critical in general population screening, such as for ovarian cancer [26]. In the context of this study, the primary focus is on maximizing detection accuracy; therefore, SVM remains the best-performing model.

4.6.3 Cross-Validation Stability, Overfitting Analysis, and Learning Curve

Table 9 presents the results of the model stability evaluation using 5-fold cross-validation. The results indicate that Gaussian Naive Bayes consistently performed worst across all folds, followed by the other two methods. In contrast, Random Forest and SVM demonstrated comparable and stable performance, with only negligible differences between folds. These findings suggest that Random Forest and SVM are more reliable than Gaussian Naive Bayes for the dataset used.

Table 9 Cross-Validation Stability of the Proposed Models

Fold	Random Forest	SVM	Gaussian Naive Bayes
Fold 1	0.978195	0.975188	0.967669
Fold 2	0.978947	0.970677	0.954135
Fold 3	0.980451	0.973684	0.945113
Fold 4	0.975940	0.964662	0.929323
Fold 5	0.975169	0.969902	0.955606

Table 10 Robustness Analysis Results of the Proposed Models

Model	Mean Accuracy	Std Deviation	Confidence Interval
Random Forest	0.977741	0.001942	0.9777 (+/- 0.0039)
SVM	0.970823	0.003635	0.9708 (+/- 0.0073)
Gaussian Naive Bayes	0.950369	0.012739	0.9504 (+/- 0.0255)

Based on the evaluation results, both Random Forest and SVM demonstrate strong, comparable performance, as indicated by their high mean accuracies. Random Forest shows slightly better stability with a lower standard deviation, while SVM remains competitive despite exhibiting slightly higher variability. In contrast, Gaussian Naive Bayes performs worst, with the lowest mean accuracy and the largest standard deviation and confidence interval, indicating lower stability and weaker classification performance than the other two models.

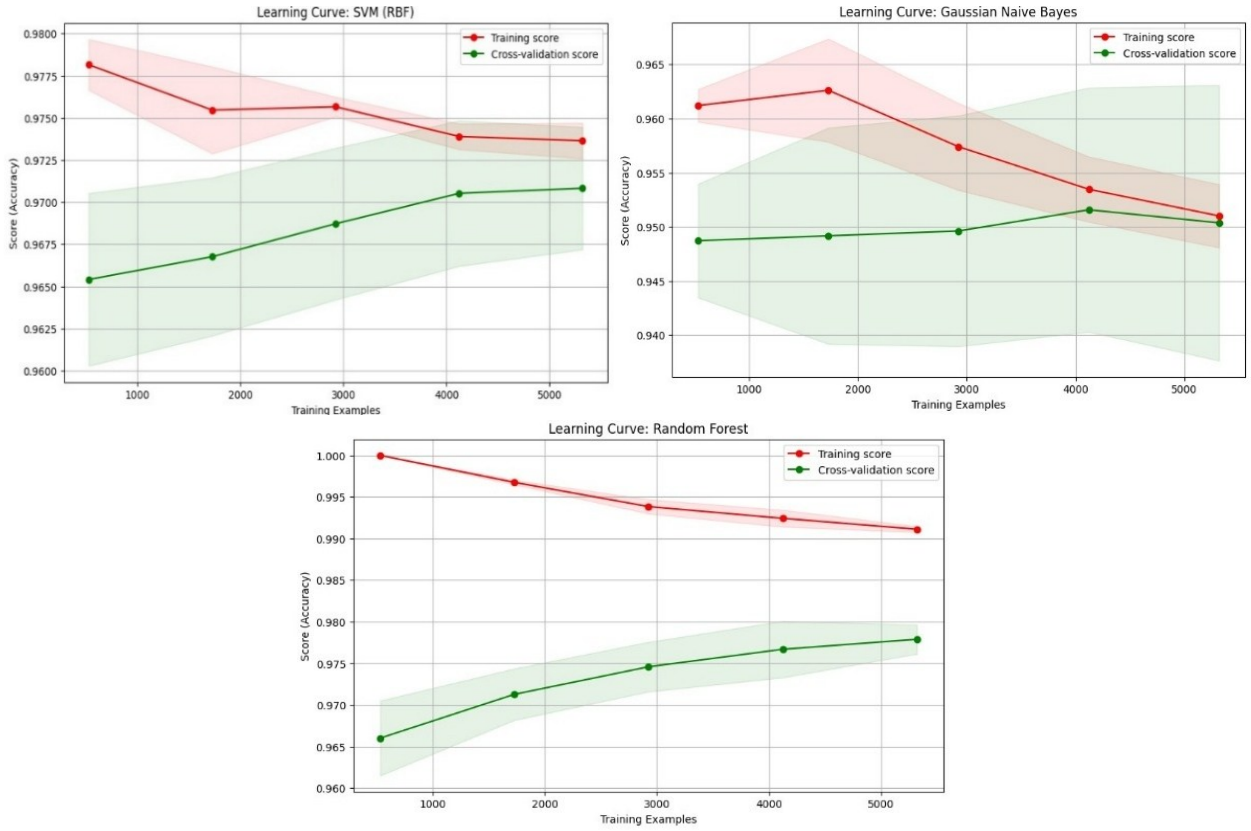


Figure 9 Learning curves of SVM, Naïve Bayes, and Random Forest using Radiomic features

Technically and quantitatively, the three graphs in Figure 9 show distinct performance patterns. Random Forest demonstrates the highest capacity, maintaining accuracy above 0.975; despite slight overfitting where training score drops from 1.0 to 0.991, the rising cross-validation trend suggests strong generalization potential with more data. SVM shows the most stable generalization, with the narrowest gap between training and validation scores at 5000+ samples, indicating consistent prediction despite slightly higher variance. In contrast, Gaussian Naive Bayes exhibits high bias or underfitting, with performance around 0.950 and a sharp decline in training score after 3000 samples, indicating that its simpler linear/probabilistic nature limits its capacity compared to non-parametric or kernel-based models.

4.6.4 Bias-Variance Analysis

Based on the bias-variance analysis in Table 11, the SVM model exhibits the most optimal trade-off, achieving the lowest total error of 0.0218. This result arises from a combination of low bias squared with $Bias^2$ of 0.0216 and very minimal variance of 0.0002, indicating strong generalization capability. The Random Forest model yields a total error of 0.0281, which is slightly higher due to larger contributions from both bias and variance than in the SVM. In contrast, the Naïve Bayes model demonstrates the weakest performance, with the highest total error of 0.0589, dominated by a high Bias^2 value of 0.0575. This suggests underfitting relative to the other two models. Overall, SVM achieved the most balanced bias-variance relationship, confirming its robustness and superior generalization among the three models.

Table 11 Bias-Variance Analysis with Feature Extraction

Models	Estimated $Bias^2$	Estimated Variance	Total Error ($Bias^2$ + Variance)
Random Forest	0.0277	0.0004	0.0281
SVM	0.0216	0.0002	0.0218
Naïve Bayes	0.0575	0.0014	0.0589

4.6.5 Model Generalization Analysis

Based on Table 12, the Random Forest model demonstrates good stability, with a very small positive generalization gap of 0.0050, indicating its ability to maintain performance from training to test data. Meanwhile, the SVM and Gaussian Naïve Bayes models exhibit a unique phenomenon, with negative generalization gaps of -0.0117 and -0.0214, respectively. This suggests that neither model suffers from overfitting; in fact, for Naïve Bayes, test performance is slightly better than training performance, likely because the test data distribution closely matches the model's Gaussian assumptions. Overall, SVM remains the best model, balancing low error rates across both datasets while maintaining a stable generalization gap.

Table 12 Model Generalization Analysis with Feature Extraction

Models	Train Risk (Error Rate)	Test Risk (Error Rate)	Generalization Gap
Random Forest	0.0090	0.0140	0.0050
SVM	0.0257	0.0140	-0.0117
Naïve Bayes	0.0495	0.0281	-0.0214

5 Conclusion

This study demonstrates that a two-stage framework integrating YOLOv11 for lesion detection and radiomics-based feature extraction significantly improves CT-based stroke classification. The incorporation of radiomic features enhances model stability by encoding clinically meaningful intensity and texture characteristics, leading to more reliable generalization. Models with radiomics achieved consistently high test accuracies of 0.97–0.99, with balanced precision, recall, and F1 Scores, whereas models trained without radiomics exhibited near-perfect scores of 1.00, indicating potential overfitting and limited real-world robustness.

Among the evaluated classifiers, the SVM proved to be the most reliable. SVM achieved a mean cross-validation accuracy of 0.9708 with low variability, with a standard deviation of 0.0036, the lowest total error in bias-variance analysis at 0.0218, and a small negative generalization gap of -0.0117, confirming strong generalization without overfitting. Its extremely low variance of 0.0002 further indicates consistent and stable performance. Random Forest also showed competitive accuracy of 0.9777 but exhibited a slightly higher total error of 0.0281, while Gaussian Naïve Bayes demonstrated the weakest performance, with higher bias and lower stability across folds. Overall, the integration of YOLOv11, radiomics, and SVM provides the most robust and clinically reliable solution for CT-based stroke detection and classification, prioritizing generalization and error minimization over superficially perfect accuracy. Therefore, combining YOLOv11 with radiomics and SVM provides the optimal, most reliable solution for stroke detection and classification.

The models in this study were tested using publicly available CT scan images sourced from multiple online datasets, ensuring reproducibility. However, limitations remain: real-world performance may vary when applied to hospital-acquired data, the dataset size for certain stroke subtypes is relatively limited, and additional clinical variables, such as patient history or comorbidities, were not incorporated. Addressing these limitations in future studies would further enhance the clinical applicability and robustness of the models. As a contribution to Future research, this study could improve efficiency by selectively incorporating other extraction methods, such as Local Binary Pattern (LBP). Then, future studies could use hospital-acquired datasets and include additional stroke subtypes.

Declaration of the Use of Generative AI

The use of generative AI (ChatGPT) was limited to language translation and grammatical improvement. All aspects of conceptualization and design, data collection, analysis, and interpretation were conducted by the authors. All AI-assisted processes were thoroughly reviewed by the authors and remain fully responsible for the accuracy, integrity, and originality of the manuscript in accordance with publication ethics standards.

Bibliography

[1] V. L. Feigin and M. Owolabi, “Pragmatic solutions to reduce the global burden of stroke: a World Stroke Organization–Lancet Neurology Commission,” *Lancet Neurol.*, vol. 22, no. 12, pp. 1160–1206, 2023, doi: 10.1016/S1474-4422(23)00277-6.

[2] Kemenkes, “Cegah Stroke Dengan Aktivitas Fisik [Prevent Stroke With Physical Activity],” *Kementerian Kesehatan*, p. 1, 2024, [Online]. Available: https://kemkes.go.id/id/cegah-stroke-dengan-aktivitas-fisik.

[3] D. Dwilaksono, T. E. Fau, S. E. Siahaan, C. S. P. B. Siahaan, K. S. P. B. Karo, and T. Nababan, “Faktor-Faktor yang Berhubungan dengan Terjadinya Stroke Iskemik pada Penderita Rawat Inap [Factors Associated with the Occurrence of Ischemic Stroke in Hospitalized Patients],” *J. Penelit. Perawat Prof.*, vol. 5, no. 2, pp. 449–458, 2023, doi: 10.37287/jppp.v5i2.1433.

[4] A. K. Boehme, C. Esenwa, and M. S. V. Elkind, “Stroke Risk Factors, Genetics, and Prevention,” *Circulation Research*, vol. 120, no. 3, pp. 472–495, 2017, doi: 10.1161/CIRCRESAHA.116.308398.

[5] I. Tarigan, "Study on the Effectiveness of Thrombolytic Drugs in Emergency Treatment of Ischemic Stroke: Time and Outcome Analysis," *Jurnal Farmasimed*, vol. 6, no. 2, pp. 200–204, 2024, [Online]. Available: https://ejournal.medistra.ac.id/index.php/JFM/article/view/2506.

[6] A. Familah, A. F. Arifin, A. H. Muchsin, M. E. Rachman, and Dahliah, "Characteristics of Ischemic Stroke and Hemorrhagic Stroke Patients," *Fakumi Med. J. J. Med. Students*, vol. 4, no. 6, pp. 456–463, 2024, doi: 10.33096/fmj.v4i6.468.

[7] S. Setianingsih, L. E. Darwati, and H. A. Prasetya, "Study Deskriptif Penanganan Pre-Hospital Stroke Life Support Pada Keluarga [Descriptive Study of Pre-Hospital Stroke Life Support Management in Families]," *J. Perawat Indones.*, vol. 3, no. 1, p. 55, 2019, doi: 10.32584/jpi.v3i1.225.

[8] L. Dewi and E. Fitraneti, "Stroke Iskemik [Iscemic Stroke]", *Scientific Journal*, vol. 3, no. 6, pp. 379–388, 2024, doi: 10.56260/sciena.v3i6.173.

[9] M. G. Ragab *et al.*, "A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023)," *IEEE Access*, vol. 12, no. December, pp. 57815–57836, 2024, doi: 10.1109/ACCESS.2024.3386826.

[10] A. Vial *et al.*, "The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: A review," *Transl. Cancer Res.*, vol. 7, no. 3, pp. 803–816, 2018, doi: 10.21037/tcr.2018.05.02.

[11] S. Yousaf, S. M. Anwar, H. RaviPrakash, and U. Bagci, "Brain Tumor Survival Prediction Using Radiomics Features," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12449 LNCS, pp. 284–293, 2020, doi: 10.1007/978-3-030-66843-3_28.

[12] M. Nijiati *et al.*, "Deep learning and radiomics of longitudinal CT scans for early prediction of tuberculosis treatment outcomes," *Eur. J. Radiol.*, vol. 169, no. July, p. 111180, 2023, doi: 10.1016/j.ejrad.2023.111180.

[13] N. H. Putri, J. Jasril, M. Irsyad, S. Agustian, and F. Yanto, "Klasifikasi Citra Stroke Menggunakan Augmentasi dan Convolutional Neural Network EfficientNet-B0 [Stroke Image Classification Using Augmentation and Convolutional Neural Network EfficientNet-B0]," *J. Media Inform. Budidarma*, vol. 7, no. 2, p. 650, 2023, doi: 10.30865/mib.v7i2.5981.

[14] H. Kamozawa, M. Tanaka, "Atrial Fibrillation Detection from Holter ECG Using Hybrid CNN−LSTM Model and P/f-wave Identification", *Advanced Biomedical Engineering*, vol. 14, pp. 46–53, 2025, doi: 10.14326/abe.14.46.

[15] V. Dharshini, S. Deepika, S. B. Devamane, R. Divya, and Gaganashree, "Performance Analysis Machine Learning Algorithms for Stress Detection," *Proc. - 2023 Int. Conf. Comput. Intell. Information, Secur. Commun. Appl. CIISCA 2023*, vol. 7, no. 4, pp. 395–400, 2023, doi: 10.1109/CIISCA59740.2023.00081.

[16] A. I. Ardelean, E. R. Ardelean, and A. Marginean, "Can YOLO Detect Retinal Pathologies? A Step Towards Automated OCT Analysis," *Diagnostics*, vol. 15, no. 14, pp. 1–22, 2025, doi: 10.3390/diagnostics15141823.

[17] S. M. A. Monisha and R. Rahman, "Brain Tumor Detection in MRI Based on Federated Learning with YOLOv11," *arXiv*, 2025, doi: 10.48550/arXiv.2503.04087.

[18] J. Gong *et al.*, "Enhancing brain metastasis prediction in non-small cell lung cancer: a deep learning-based segmentation and CT radiomics-based ensemble learning model," *Cancer Imaging*, vol. 24, no. 1, pp. 1–12, 2024, doi: 10.1186/s40644-023-00623-1.

[19] J. Permatasari, E. U. Armin, E. Sunardi, M. B. Laili, and S. M. Putri, "Evaluasi kinerja YOLOv11 pada deteksi penyakit tanaman cabai: Studi komparatif dengan YOLOv8, YOLOv5, dan SSD [Performance evaluation of YOLOv11 for chili plant disease detection: A comparative study with YOLOv8, YOLOv5, and SSD]," *Jurnal Teknologi*, vol. 25, no. 3, 2025, [Online]. Available: https://e-jurnal.pnl.ac.id/teknologi/article/view/8400.

[20] A. Ardiansyah, A. S. Widagdo, K. N. Qodri, D. Hidayani, and M. Romadhani, "Implementasi deteksi tumor otak menggunakan YOLOv11 dan Flask [Implementation of brain tumor detection using YOLOv11 and Flask]," *Jurnal FASILKOM (Teknologi Informasi dan Ilmu Komputer)*, vol. 15, no. 2, 2025, doi: 10.37859/jf.v15i2.9703.

[21] W. Zhang, Y. Guo, and Q. Jin, "Radiomics and Its Feature Selection: A Review," *Symmetry (Basel).*, vol. 15, no. 10, 2023, doi: 10.3390/sym15101834.

[22] M. R. Salmanpour, M. Shamsaei, G. Hajianfar, H. Soltanian-Zadeh, and A. Rahmim, "Longitudinal clustering analysis and prediction of Parkinson's disease progression using radiomics and hybrid machine learning," *Quant. Imaging Med. Surg.*, vol. 12, no. 2, pp. 906–919, 2022, doi: 10.21037/qims-21-425.

[23] S. B. Lin, Y. Wang, and D. X. Zhou, "Generalization Performance of Empirical Risk Minimization on Over-Parameterized Deep ReLU Nets," *IEEE Trans. Inf. Theory*, vol. 71, no. 3, pp. 1978–1993, 2025, doi: 10.1109/TIT.2025.3531048.

[24] J. Liu *et al.*, "Deep learning-based identification and localization of intracranial hemorrhage in patients using a large annotated head computed tomography dataset: A retrospective multicenter study," *Intell. Med.*, vol. 5, no. 1, pp. 14–22, 2025, doi: 10.1016/j.imed.2024.11.002.

[25] G. Tapia, H. Allende-Cid, S. Chabert, D. Mery, and R. Salas, "Benchmarking YOLO Models for Intracranial Hemorrhage Detection Using Varied CT Data Sources," *IEEE Access*, vol. 12, no. October, pp. 188084–188101, 2024, doi: 10.1109/ACCESS.2024.3510517.

[26] M. S. Pepe, Z. Feng, H. Janes, P. M. Bossuyt, and J. D. Potter, "Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: Standards for study design," *J. Natl. Cancer Inst.*, vol. 100, no. 20, pp. 1432–1438, 2008, doi: 10.1093/jnci/djn326.