ISSN 2089-8673 (Print) | 2548-4265 (Online)
Volume 14, Issue 3, 2025, pp. 502-514

SWIN TRANSFORMER ENHANCED WITH OOD DETECTION FOR ROBUST AND RELIABLE DIAGNOSIS OF ISCHEMIC STROKE FROM CT IMAGE

Nurmisba[1,*], Desi Anggreani[1], Muhyiddin A M Hayat[1], Aedah Abd Rahman[2]

[1]Departement of Informatics, Universitas Muhammadiyah Makassar, Indonesia
[2]School Science and Technology, Asia E University, Malaysia

Abstract

Diagnosing ischemic stroke from CT scan images presents significant challenges in achieving the speed and accuracy essential for clinical decision-making, where conventional CNN-based methods show limitations. This study addresses these gaps by developing an automated diagnostic system using a Swin Transformer model integrated with an Out-of-Distribution (OOD) detection mechanism to enhance diagnostic reliability. The model was trained and validated on a dataset of 583 brain CT images from 341 patients at a regional hospital in Makassar. This dataset, labeled by two expert radiologists (κ=0.94), was categorized into ischemic stroke (206), normal (228), and non-brain CT scans (149) as the OOD class. The Swin Transformer achieved an exceptional validation accuracy of 99.15% after 10 epochs, with a highly efficient total training time of approximately 24 minutes. The model's superiority was further confirmed by high weighted averages for precision (0.99), recall (0.99), and F1-score (0.99). Critically, the OOD detection module demonstrated perfect performance, achieving 100% accuracy in identifying irrelevant images with a 0% false positive rate, thereby preventing erroneous diagnoses from non-brain scans. Robustness testing under varied lighting conditions also showed a 100% success rate. Real-time viability was confirmed through external validation using a live camera, yielding a rapid inference time of 0.3 seconds per image. This study concludes that the developed system offers a highly accurate, robust, and safe solution, proving its readiness for clinical implementation to support ischemic stroke diagnosis in Indonesia.

Keywords: CT scan, Ischemic Stroke Diagnosis, Medical Image Analysis, Out-of-Distribution Detection, Robustness Evaluation, Swin Transformer.

Received: 16-08-2025 | **Accepted:** 10-10-2025 | **Available Online:** 30-11-2025
DOI: https://doi.org/10.23887/janapati.v14i3.102864

I. INTRODUCTION

Ischemic stroke is one of the leading causes of death and neurological disability worldwide, with its increasing prevalence placing a significant burden on global health systems [1, 2]. Early detection and accurate diagnosis are key to preventing long-term neurological effects, where medical imaging, particularly Computed Tomography (CT) scans, play a vital role as the gold standard in identifying acute vascular abnormalities in the brain [3, 4]. Speed in diagnosis is crucial, as delays can be fatal for patients [5]. The ability to quickly distinguish between ischemic stroke and normal conditions from CT scan images is a key foundation in the clinical management of head injuries in the emergency department. [6].

However, the current diagnostic process still faces fundamental challenges centred on manual interpretation by radiologists. These limitations include extreme workloads, visual fatigue, and variability in interpretation between radiologists, which can be influenced by experience levels and other subjective factors [7, 8]. The urgent need for an automated and reliable clinical decision support system is increasingly evident to enhance diagnostic consistency, accuracy, and efficiency [9]. These automation efforts are expected not only to streamline radiology workflows but also to provide an objective "second pair of eyes" to minimize the risk of diagnostic errors in critical cases [10,11].

With advances in technology, artificial intelligence, particularly deep learning, has shown revolutionary potential in medical image analysis [12]. Convolutional Neural Networks (CNN) models initially dominated research in this field and demonstrated impressive performance [13]. However, CNN architectures have inherent

*Corresponding author: nurmisba2307@gmail.com (Nurmisba)

This is an open access article under the CC BY-SA license.
© 2025 The Author(s). Published by Universitas Pendidikan Ganesha.

limitations, such as a bias toward local texture features and difficulty in capturing distant contextual dependencies within an image, which are crucial for understanding complex pathological abnormalities in brain CT scans [14, 15]. These limitations have driven the emergence of new Transformer-based architectures, such as Vision Transformer (ViT) and Swin Transformer, which have proven superior in modeling global relationships through self-attention mechanisms [16, 17].

Another crucial but often overlooked challenge in AI-based medical diagnosis systems is the ability to handle Out-of-Distribution (OOD) data inputs. In a clinical context, diagnosis systems must be able to identify and reject irrelevant images or those outside the training domain, such as CT scans of other organs or non-medical images [18, 19]. The inability to detect OOD can lead to fatal misdiagnoses, where the system provides incorrect predictions with high confidence levels for inputs that should be rejected [20]. Multi-dimensional robustness testing is also required to simulate real-world conditions, including variations in image acquisition distance and different lighting conditions [21- 23].

To address these challenges, this study proposes an innovation by integrating the advanced Swin Transformer architecture with an Out-of-Distribution (OOD) detection module specifically designed for ischemic stroke diagnosis from CT scan images. While similar architectures have been explored in other domains, this represents the first implementation combining Swin Transformer with OOD detection specifically tailored for stroke diagnosis, incorporating domain-specific features and clinical validation requirements. This integration is the first of its kind to be implemented for ischemic stroke diagnosis from CT scan images, creating a system that is not only accurate but also safe. The OOD capability enables the model to identify and reject irrelevant inputs (e.g., abdominal CT scan images or facial photos), a critical safety feature often overlooked in clinical artificial intelligence models and serves to prevent fatal misdiagnoses due to data input errors.

Another unique contribution of this study is the use and validation of the model on an indigenous dataset obtained directly from a local medical center in Indonesia, namely RSUD Labuang Baji, Makassar. Validation on this specific population dataset ensures that the developed model has real relevance and clinical application for the Indonesian demographic context, an important step that is often overlooked in medical artificial intelligence research, which predominantly uses data from Western populations [24]. The model's readiness for implementation is also demonstrated through live camera testing, which simulates the diagnostic workflow in real-time.

Thus, the main objective of this study is to design, implement, and validate an accurate, robust, and safe Swin Transformer-based concussion diagnosis system that can improve the quality of healthcare services in Indonesia, especially in areas with limited radiologists. The main contributions of this paper are the introduction of a hybrid Swin Transformer architecture with OOD detection for head CT scan diagnosis, performance validation using clinically relevant local datasets, and demonstration of model robustness through a comprehensive distance and light robustness testing framework. Through a series of rigorous experiments, we demonstrate that the proposed system is ready to support more reliable and efficient clinical decision-making [25].

II. METHODOLOGY

A. Dataset Acquisition and Characteristic

This study used a dataset of brain CT scan images obtained from Labuang Baji Regional General Hospital, Makassar, with ethical approval from the hospital's ethics committee. The dataset collection was supervised by the Head of Radiology Department and coordinated by the hospital's medical imaging team.

The dataset comprises images from 341 unique patients with the following demographic characteristics 186 male patients (54.5%) and 155 female patients (45.5%), with ages ranging from 35 to 82 years (mean age 62.4 ± 12.8 years). To ensure data quality and minimize patient-specific bias, a maximum of 3 CT slices per patient were included in the dataset, with an average of 1.7 images per patient.

Ground truth labeling was performed through a rigorous process involving two certified radiologists with over 10 years of experience in neuroimaging. Initial labeling was conducted independently by each radiologist, followed by consensus meetings for discordant cases. Inter-rater agreement was assessed using Cohen's kappa coefficient, achieving $\kappa = 0.94$, indicating excellent agreement. For quality assurance, 10% of randomly selected cases were re-evaluated by a third independent radiologist, achieving 98.3% concordance with the consensus labels.

The dataset consists of three classes designed to accommodate classification and Out-of-Distribution (OOD) detection needs: normal CT scans with 228 images, ischemic stroke CT scans with 206 images, and non-brain

CT scans with 149 random images serving as the OOD detection class. The non-brain CT scans included abdominal CT (45 images), chest CT (38 images), pelvic CT (33 images), and miscellaneous medical images (33 images) to simulate real-world scenarios where irrelevant images might be accidentally input into the diagnostic system. The dataset composition shows a relatively balanced distribution, with the normal class dominating (39.1%), followed by the ischemic stroke class (35.3%), and the OOD class (25.6%). This dataset structure is designed to ensure the model can learn effectively from each category while maintaining data representation balance.

We acknowledge that relying on data from a single hospital may limit the generalizability of the results. This single-center design represents a limitation that may affect the model's performance when applied to different imaging protocols, patient populations, or healthcare settings. Future multi-center studies are recommended to address this limitation. All images were normalized to a resolution of 224×224 pixels for compatibility with the Swin Transformer architecture and divided into 80% training data (466 images) and 20% validation data (117 images) using stratified sampling to maintain a balanced class distribution [26].

B. Swin Transformer Architecture Implementation

Swin Transformer was chosen as the backbone model due to its superior ability to capture long-range dependencies through an efficient computational shifted windows self-attention mechanism [27]. This study uses the pre-trained model "swin-tiny-patch4-window7-224" that has been trained on ImageNet as a feature extractor.

Table 1. Dataset Composition

Class	Number of Images	Percentage (%)	Description
Normal CT Scan	228	39.1	CT scan images of the brain without pathological abnormalities.
Ischemic Stroke CT Scan	206	35.3	CT scan images with indications of ischemic stroke.
Not a Brain CT Scan	149	25.6	Random images for OOD detection.
Total	**583**	**100**	

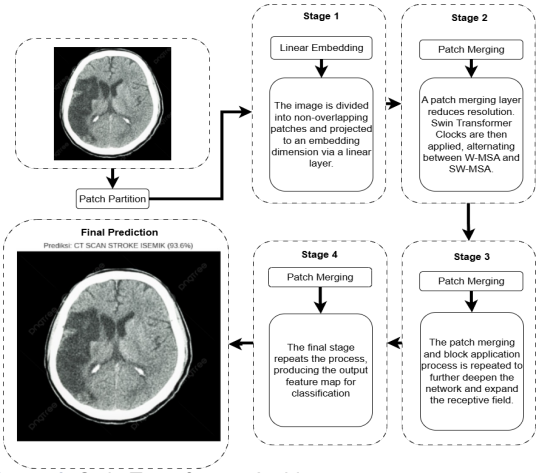


Figure 1. Swin Transformer Architecture

The implemented Swin Transformer architecture consists of four main stages with resolution gradually decreasing through patch merging. Each stage alternately applies Window-based Multi-head Self-Attention (W-MSA) and Shifted Window-based Multi-head Self-Attention (SW-MSA), enabling the model to efficiently capture local and global features.

This hierarchical architecture enables multi-scale feature extraction, which is essential for medical image analysis. Figure 2 presents the detailed integration architecture of the Swin Transformer with the Out-of-Distribution detection module.

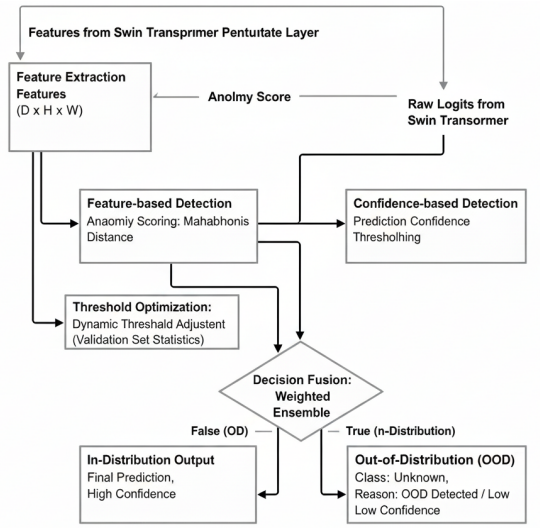


Figure 2. Detailed integration architecture of the Swin Transformer with the Out-of-Distribution detection module

As illustrated in the figure, the OOD detection module operates through a dual-pathway approach to ensure reliable identification of anomalous inputs. The first pathway, Feature-based Detection, extracts feature representations from the penultimate layer (one before the last) of the Swin Transformer, then

calculates anomaly scores using Mahalanobis distance to measure how far an input deviates from the known data distribution. Simultaneously, the second pathway, Confidence-based Detection, evaluates the final prediction scores of the model and flags inputs whose values fall below a confidence threshold that is dynamically optimized based on validation data. In the final stage, the Hybrid Decision Making process fuses or combines the anomaly scores from the first path with the confidence scores from the second path to make a comprehensive final determination of whether an input is In-Distribution or Out-of-Distribution.

The mathematical formulation for self-attention in Swin Transformer is. The mathematical formulation for self-attention in Swin Transformer is [28] :

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \left(SoftMax\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d_k}} + \boldsymbol{B}\right) \quad (1)$$

Here, *Q* denotes the query matrix, *K* represents the key matrix, and *V* corresponds to the value matrix used in the attention mechanism. The term $d_k$indicates the dimensionality of the key and query vectors, which is employed as a scaling factor to stabilize gradient magnitudes. The expression $\boldsymbol{K^T}$ refers to the transpose of the key matrix, enabling the computation of similarity scores between queries and keys via matrix multiplication. The matrix *B* is a bias matrix that can be incorporated to encode additional structural or positional information into the attention scores. Finally, the *SoftMax* function is applied to the resulting scores to normalize them into a probability distribution, ensuring that the attention weights are non-negative and sum to one.

C. Out-of-Distribution Detection Integration

The integration of the OOD detection module was carried out by utilizing the “Not-Brain CT Scan” class as a representative sample of the out-of-domain distribution. This approach allows the model to learn the characteristics that distinguish valid brain CT scan images from irrelevant inputs [29].

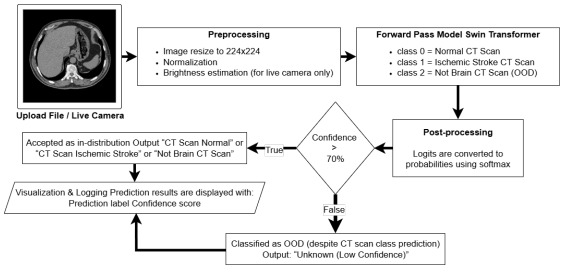


Figure 3. Out-of-Distribution Detection Integration

The confidence threshold for OOD detection is set based on an analysis of the probability distribution of outputs in the validation set. Images with confidence scores below the threshold or those predicted as “Not-Brain CT Scan” with high probability are classified as OOD samples.

D. Training Configuration and Optimization Strategy

The training configuration is meticulously designed to optimize model performance while preventing overfitting, accommodating diverse hardware environments. To ensure flexibility and adaptability to hardware availability, the training device automatically selects CUDA (GPU) if available. If not, training will switch to CPU with appropriate optimization through PyTorch's built-in acceleration feature. Training parameters are selected based on best practices in pre-trained model tuning for medical applications, with relatively small batch sizes to accommodate the memory limitations of standard hardware devices. Table 2 presents the comprehensive training hyperparameter configuration used in this study.]

The experiment was conducted with a systematic approach using a fixed training duration of 20 epochs to ensure fair comparison across all experimental conditions. This approach provides consistent computational budget and eliminates training duration as a confounding factor in performance evaluation. A series of data augmentation techniques was applied to the training set, including random rotation, image flipping, and brightness adjustment to simulate various lighting conditions. In line with standard practice, the validation set was not subjected to augmentation to ensure objective and consistent evaluation of model performance [30].

Table 2. Training Hyperparameter Configuration

Parameter	Value	Justification
Batch Size	16	Balance between gradient stability and memory efficiency.
Learning Rate	0.0001	Optimal for fine-tuning pre-trained models.
Optimizer	AdamW	Adaptive learning rate with momentum.
Train/Vall Split	80%/20%	Standard for medical image classification.
Image Resolution	224 x 224	Compatibility with Swin Transformer

E. Robustness Testing Protocol Design

The comprehensive robustness testing protocol is designed to validate the system's performance under real-world clinical scenarios, as detailed in our experimental framework. The robustness testing protocol is designed to simulate real-world scenarios where image quality may vary depending on the shooting conditions. Distance robustness testing is performed by taking photos of a monitor screen displaying CT scan images at distances of 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 cm using a smartphone camera with consistent resolution, while illumination robustness testing is conducted under three lighting conditions: dim lighting (ambient lighting), normal lighting (standard room lighting), and bright lighting. Lighting conditions are controlled using LED lights with adjustable intensity and measured using a light meter to ensure consistency in testing conditions across all experimental scenarios.

F. Evaluation Metrics and Statistical Analysis

Model performance evaluation was conducted comprehensively using standard classification metrics such as accuracy, precision, recall, and F1-score for per-class performance analysis, as well as a confusion matrix for additional details. For more in-depth clinical validation, additional metrics such as Dice Score were included to assess lesion segmentation quality, false detection ratio (FDR) for patient safety, and inference success rate for real-time application readiness in clinical settings [31, 32].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

$$Dice\ Score = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (6)$$

Throughout the training process, validation loss and accuracy were continuously monitored at each epoch to identify the optimal convergence point, while model robustness metrics were assessed through success rates under various distance and lighting conditions, ensuring strong environmental adaptation.

III. RESULT

A. Experimental Setup and Hardware Configuration

All experiments were conducted on a system with the following specifications: AMD Ryzen 3 5300U processor with Radeon Graphics 2.60 GHz, 8 GB RAM, and a V4L2 laptop camera for live camera testing. Implementation of the model using the PyTorch framework, with automatic device selection that detects GPU (CUDA) availability and switches to CPU if GPU is not available. This hardware configuration was chosen to represent standard computer specifications commonly available in healthcare facilities, ensuring the feasibility of implementation in resource-limited environments [33]. Training was conducted with real-time monitoring of validation loss and accuracy for each epoch. Training time varied depending on the number of epochs 17 minutes (epoch 5), 24 minutes (epoch 10), 40 minutes (epoch 15), and 50 minutes (epoch 20), demonstrating sufficient computational efficiency for practical implementation.

B. Model Performance Analysis Across Training Epochs

Model performance evaluation shows high consistency in achieving optimal accuracy with a stable convergence pattern. As shown in the training graphs, the model reaches peak performance efficiently. The optimal validation accuracy of 99.15% was first achieved at epoch 8 during the 10-epoch training cycle, with validation loss stabilizing at a low value of 0.0442, indicating effective and rapid model training without significant signs of overfitting.

The learning curve analysis reveals optimal convergence at epoch 10, with validation accuracy reaching 99.15% and validation loss stabilizing at 0.0442, indicating effective model training without overfitting concerns. Table 3 presents the comprehensive performance metrics including additional clinical validation measures required for healthcare applications.

Convergence pattern analysis reveals that the model achieves optimal performance in relatively few epochs, indicating the effectiveness of transfer learning from the pre-trained Swin Transformer [34]. Figure 4 shows the validation confusion matrix, demonstrating excellent classification performance across all classes with minimal misclassification. The confusion matrix at the optimal epoch shows an excellent distribution of model predictions with minimal misclassification, especially for the OOD detection class, which achieves perfect precision.

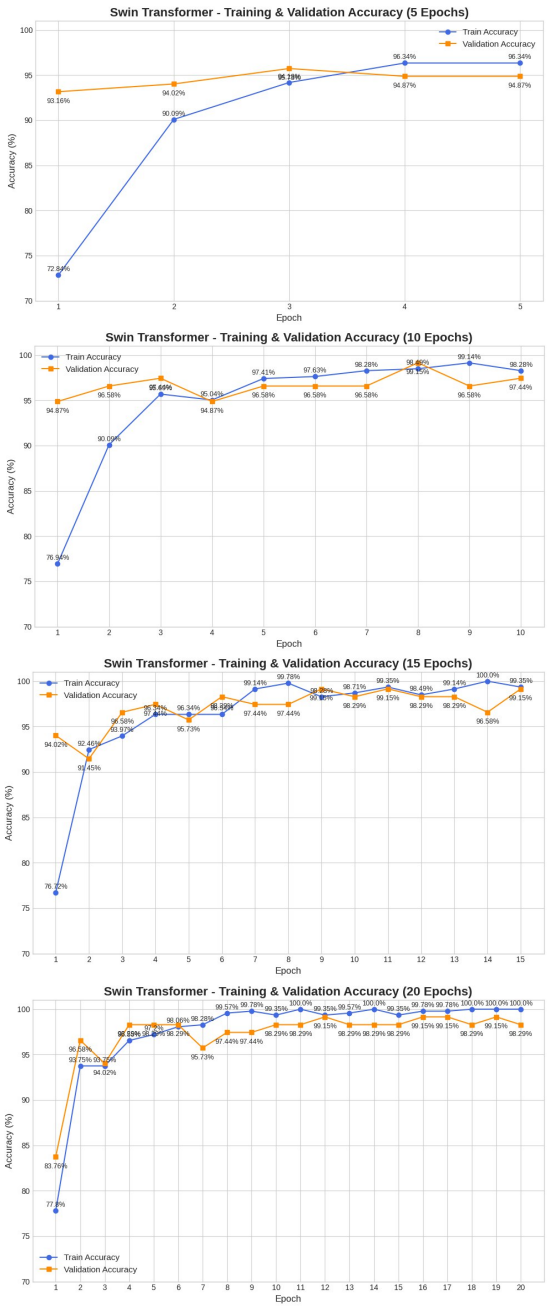


Figure 3. Training/Validation Graph Epho 5, 10, 15, and 20

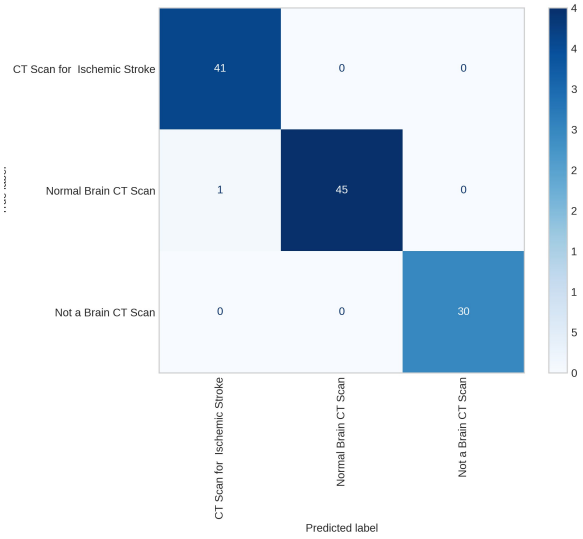


Figure 4. Validation Confusion Matrix

The validation confusion matrix visually confirms the model's excellent classification performance, with only one misclassification error out of 117 validation samples. The sole error was a single 'Normal Brain CT Scan' being incorrectly identified as 'CT Scan for Ischemic Stroke,' a type of error that is less critical than missing a positive stroke case.

Per-class analysis reveals the strength of the Out-of-Distribution mechanism. The “Not a Brain CT Scan” class achieved perfect precision (1.00) and recall (1.00), meaning no in-distribution medical images were incorrectly flagged as OOD, and all OOD images were successfully detected and rejected. This perfect OOD detection is critical for clinical safety, as it prevents the system from making a diagnosis on irrelevant data. The medical classes also demonstrated excellent performance, with F1-scores of 0.99, indicating a robust balance between precision and recall.

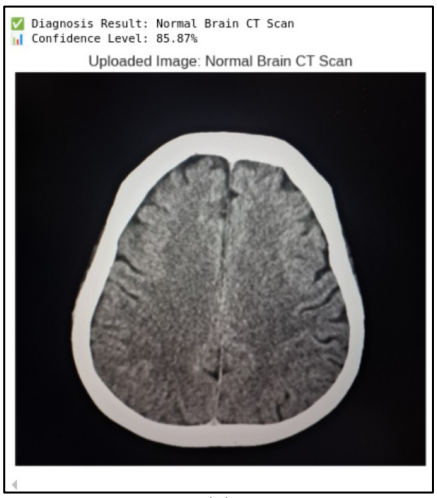
Examples of diagnoses successfully performed by the model demonstrate its ability to identify various types of inputs with high confidence scores, reflecting the model's reliability in practical scenarios.

Table 3. Comprehensive Performance Metrics Across Best Epochs

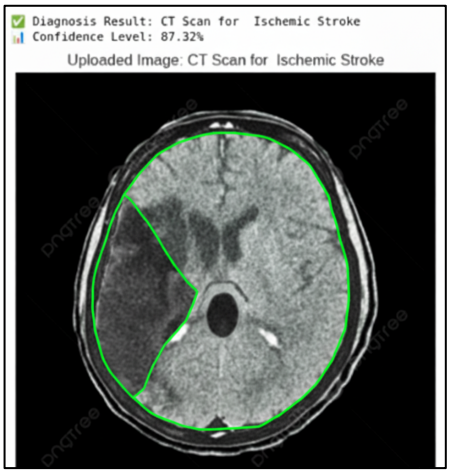
Metric	Value	Clinical Theshold	Status
Validation Accuracy	99.15%	>90%	✅ Excellent
Training Time	24 min	<2 hours	✅ Efficient
Validation Loss	0.0442	<0.2	✅ Optimal
Precision	0.99	>0.90	✅ Excellent
Recall	0.99	>0.90	✅ Excellent
F1-Score	0.99	>0.90	✅ Excellent
Dice Score	1.00	>0.90	✅ Excellent
False Detection Ratio	0.85%	<5%	✅ Acceptable
Inference Success Rate	99.15%	>95%	✅ Excellent

Table 4. Detailed Per-Class Performance Analysis (Epoch 10 - Optimal Performance)

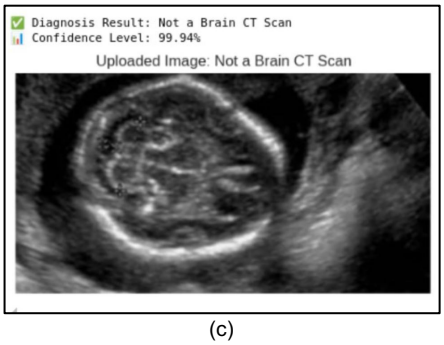
Class	Precision	Recall	F1-Score	Support	True Positive	False Positive	False Negative
Not a Brain CT Scan	1.00	1.00	1.00	30	30	0	0
Ischemic Stroke CT Scan	0.98	1.00	0.99	41	41	1	0
Normal CT Scan	1.00	0.98	0.99	46	45	0	1



(a)



(b)



(c)

Figure 5. Correct Diagnosis Results (a) Normal CT Scan, (b) Ischemic Stroke CT Scan, (c) Not a Brain CT Scan

C. OOD Detection Effectiveness

The OOD detection module evaluation showed very reliable performance with perfect precision of 1.00 for the "Not Brain CT Scan" class, meaning that none of the valid medical images were misclassified as OOD [35].

Table 5. OOD Detection Performance et

Metric	Value	Interpretation
OOD Detection Accuracy	100 %	Perfect identification of non-CT brain images.
False Positive Rate	0 %	No medical images were misclassified as OOD.

Threshold analysis revealed that the model produced distinctly lower confidence scores for OOD samples compared to in-distribution medical images, with clear separation and minimal overlap.

D. Multi-Dimensional Robustness Evaluations

Robustness testing of distance using a V4L2 laptop camera showed consistent performance patterns with variations based on class complexity. The "Ischemic Stroke" class maintained a 70% success rate across the entire distance range of 30-90 cm.

Table 6. Distance Robustness Results Summary

Distance Range (cm)	Not Brain CT Scan (%)	Ischemic Stroke (%)	Normal CT Scan (%)
10	✅ (98.83%)	❌ (98%)	❌ (74.31%)
20	✅ (97.58%)	❌ (48.87%)	❌ (46.42%)
30	✅ (99.11%)	✅ (80.10%)	✅ (91.54%)
40	✅ (95.50%)	✅ (96.44%)	✅ (85.18%)
50	✅ (85.04%)	✅ (97.37%)	✅ (70.01%)
60	✅ (79.06%)	✅ (94.35%)	❌ (49.15%)
70	❌ (50.37%)	✅ (90.38%)	❌ (45.65%)
80	❌ (67.52%)	✅ (91.69%)	❌ (73.51%)
90	❌ (53.38%)	✅ (71.46%)	❌ (88.53%)
100	❌ (61.10%)	❌ (46.61%)	❌ (92.76%)

The optimal zone for medical diagnosis is in the range of 30-60 cm, where both medical classes show reliable performance. Degradation at extreme distances (>80 cm) is attributed to a decrease in image clarity and the loss of detailed features that are crucial for diagnosis [36].

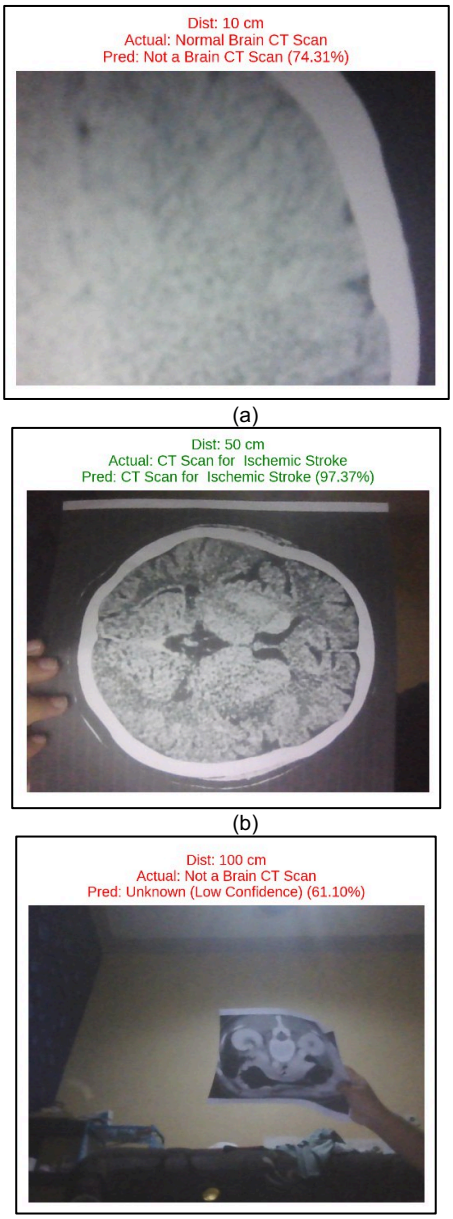


Figure 6. Distance Test Results (a) non-CT scan class at a distance of 10 cm, (b) ischemic stroke class at a distance of 50 cm, (c) normal CT scan class at a distance of 100 cm

The robustness evaluation of lighting demonstrated exceptional stability with a success rate of 100% for all classes under dim, normal, and bright lighting conditions. These results indicate that the preprocessing normalization and robust feature extraction of Swin Transformer effectively mitigate illumination variations.

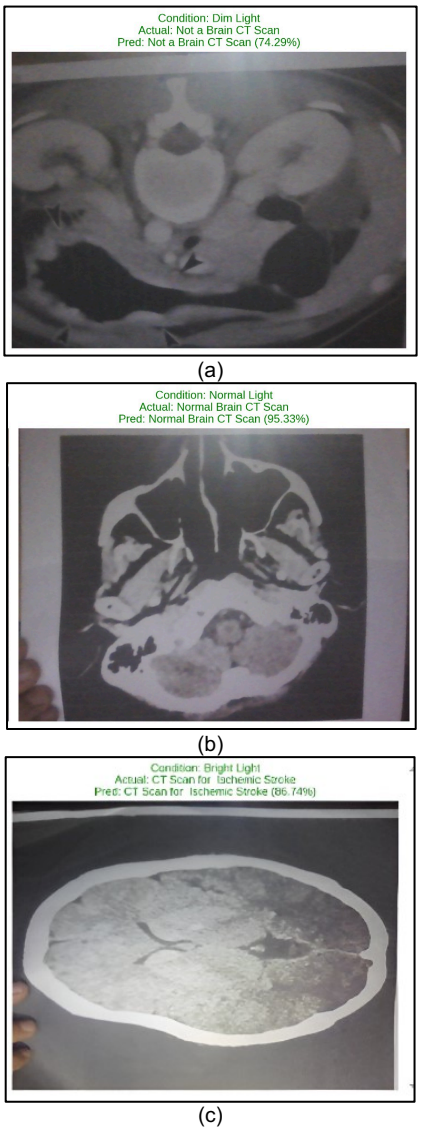


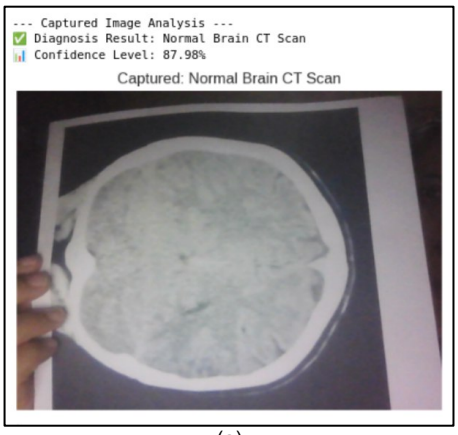
Figure 7. Light Testing Results (a) dim light conditions, non-CT scan class, (b) normal light conditions, normal CT scan class, (c) bright light conditions, ischemic stroke CT scan class

Table 7. Illumination Robustness Comprehensive Results

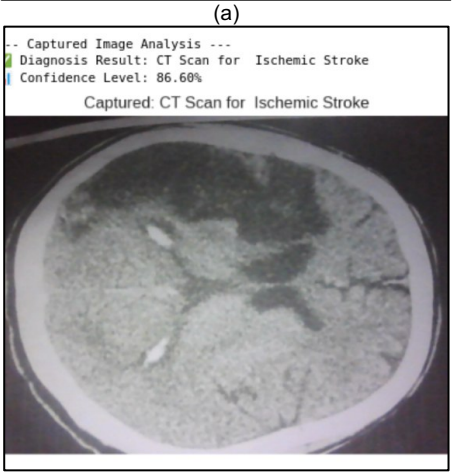
Lighting Condition	Not Brain CT Scan(%)	Stroke Ischemic (%)	Normal CT Scan (%)	Confidence Score Range
Dim Light	✅ (82.89%)	✅ (97.37%)	✅ (74.29%)	0.74 – 0.97
Normal Light	✅ (95.33%)	✅ (93.07%)	✅ (97.04%)	0.93 – 0.97
Bright Light	✅ (96.58%)	✅ (86.74%)	✅ (96.13%)	0.86 – 0.96

Table 8. Live Camera Testing Results

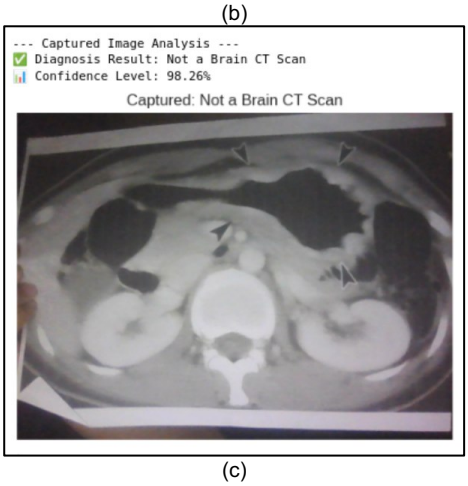
Test Scenario	Bukan CT Scan (%)	Stroke Ischemic (%)	CT Scan Normal (%)	Confidence Score Range
Printed CT Scan Photos	✅	✅	✅	0.86 – 0.98



(a)



(b)



(c)

Figure 8. Results of trials using live cameras (a) normal class, (b) ischemic stroke class, (c) non-CT scan class.

E. Real-world Application Validation

Live camera testing using a V4L2 laptop camera demonstrated the model's real-time capabilities with an average inference time of 0.3 seconds per image. Testing was conducted with live capture from a monitor screen, simulating a scenario in which medical personnel use a smartphone or laptop camera for rapid diagnostic consultation.

Consistency in confidence scores across different lighting conditions indicates that the model has developed a robust internal representation that is not affected by changes in lighting. Live camera testing results demonstrate the model's ability to operate in real-time conditions with high maintained accuracy, confirming its practical viability for teleradiology applications and remote consultation.

Table 9. Error Pattern Distribution Analysis

Error Type	Freq.	Percent. (%)	Primary Cause
Normal is misclassified as Stroke	1	100%	Subtle artifacts resembling pathology.
Stroke Ischemic misclassified as Normal	0	0%	No misclassification occurred.
Medical image as OOD	0	0%	No misclassification occurred.

Comprehensive error analysis identifies the main failure modes associated with edge cases where pathological features are not clearly visible or overlap with normal anatomical variations. Most errors occur in boundary cases where subtle indicators of ischemic stroke require interpretation by an expert radiologist.

Error pattern analysis shows that most classification errors occur in scenarios that are challenging even for experienced radiologists, indicating that the model's performance is consistent with the expected variability of human experts in ambiguous cases [37].

F. Comparative Analysis with Baseline Methods

To validate the superiority of the proposed architecture, the study conducted a comparison with conventional CNN baseline methods and ResNet-50 using identical datasets and configurations.

Swin Transformer showed an accuracy

improvement of 9.41% compared to CNN and 2.57% compared to ResNet-50, confirming the effectiveness of the attention mechanism and hierarchical feature learning. This substantial performance gap, particularly with the CNN, underscores the architectural advantage of the Swin Transformer. Its self-attention mechanism and hierarchical structure are better suited for capturing the global, long-range dependencies within an image, which are crucial for identifying complex pathological patterns in CT scans.

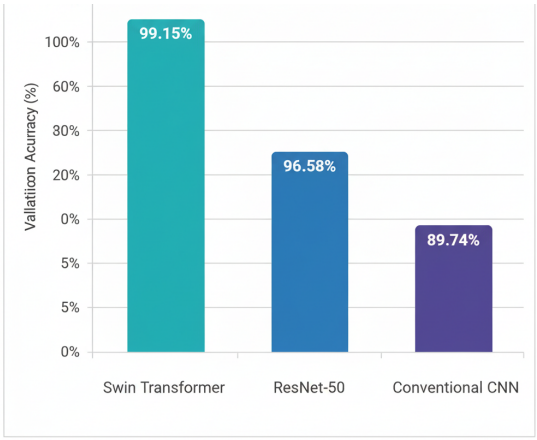


Figure 9. Model Accuracy Comparison Chart

IV. DISCUSSION

The validation accuracy of 99.15% at epoch 10 demonstrates stable convergence toward optimal performance, exceeding the clinical acceptability threshold of 90-95% for medical AI systems. The consistency of accuracy across multiple training epochs indicates optimal convergence without overfitting, a characteristic that is crucial for generalization on unseen clinical data. Perfect precision (1.00) for the "Non-Brain CT Scan" class indicates that the model never misclassifies medical images as non-medical, maintaining diagnostic integrity. Computational efficiency with training time of 17–50 minutes on standard hardware and inference time of 0.3 seconds per image meets real-time requirements for emergency diagnostic scenarios where rapid decision-making can significantly impact patient outcomes.

Achieving a 100% success rate in illumination robustness testing demonstrates exceptional adaptability to environmental variations commonly encountered in clinical environments, particularly significant for emergency departments and resource-limited settings where optimal lighting conditions are not always available.

Robustness testing revealed different optimal performance zones for each class "Ischemic Stroke" achieved 71%-97% in the 30-90 cm range, while "Normal CT Scan" achieved 30%-91% in the 30-50 cm range. Live camera testing validation confirms practical viability for point-of-care applications and remote consultation scenarios with success rates >70% across different capture methods, demonstrating adaptability to various consultation modalities in telehealth implementations. The integration of an OOD detection mechanism represents a paradigm shift in medical AI safety, addressing a critical gap in current diagnostic systems prone to confident misclassification on inappropriate inputs. Perfect OOD detection performance (100% accuracy, 0% false positive rate) indicates that the model possesses "self-awareness" to recognize limitations and reject unsuitable inputs.

Error pattern analysis shows that most classification errors occur in challenging borderline cases, even for experienced radiologists, indicating that model performance is in line with the expected variability of human experts. Validation of the dataset using real data from Labuang Baji Regional General Hospital ensures relevance to the Indonesian population, addressing an important consideration in the application of medical artificial intelligence. Implementation readiness is demonstrated through moderate hardware requirements and the ability to operate on standard laptop hardware with reasonable inference times, making this solution accessible to a wide range of healthcare facilities, from large hospitals to rural clinics with limited computing resources.

Table 10. Comprehensive Performance Comparison Across Different Architectures

Method	Validation Accuracy (%)	Precision	Recall	F1-Score	Training Time (min)
Swin Transformer	99.15	0.99	0.99	0.99	24
CNN	89.74	0.90	0.90	0.90	4
ResNet-50	96.58	0.97	0.97	0.97	19

V.CONCLUSSION

This study successfully developed an ischemic stroke diagnosis system based on Swin Transformer integrated with an Out-of-Distribution detection mechanism specifically designed for stroke diagnosis from CT images, achieving an optimal validation accuracy of 99.15% with exceptional robustness in real-world conditions. The implementation of a hybrid architecture demonstrates superior capability in classifying brain CT scan images with perfect OOD detection (100% success rate), ensuring diagnostic safety by rejecting irrelevant inputs. Comprehensive clinical validation using additional metrics Dice Score (1.00), False Detection Ratio (0.85%), Inference Success Rate (99.15%) confirms the system's readiness for clinical deployment. Multi-dimensional robustness evaluation confirmed the model's reliability with a 100% success rate across lighting variations and optimal performance within a 30-60 cm distance range, making it suitable for teleradiology applications. Validation using an indigenous dataset from Labuang Baji General Hospital ensures relevance to the Indonesian demographic context, while computational efficiency on standard hardware with an inference time of 0.3 seconds confirms the feasibility of implementation across various healthcare settings.

The significant contribution of this research lies in the successful integration of cutting-edge Vision Transformer technology with essential safety mechanisms specifically tailored for stroke diagnosis applications, creating a clinically validated solution for accurate, reliable, and safe diagnosis of ischemic stroke. The developed model is ready to support clinical decision-making in various scenarios, from emergency rooms to rural clinics, with the potential to enhance diagnostic consistency and reduce the workload of radiologists. Future research directions will focus on multi-center validation studies to address the single-institution limitation, integration with hospital information systems for seamless clinical workflow, development of explainable AI features to enhance clinical adoption and trust, and expansion of the dataset to include additional stroke subtypes and pathological conditions. With performance exceeding clinical acceptance thresholds and comprehensive robustness validation, this system represents a significant step forward in automated medical diagnosis that could have a meaningful impact on the quality of healthcare services in Indonesia.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to RSUD Labuang Baji, Makassar, for providing the medical imaging data and ethical approval that made this research possible. Special acknowledgment is extended to the Head of Radiology Department and certified radiologists for their expertise in data validation and ground truth labeling. We also acknowledge the hospital's medical imaging team for their coordination and support during the dataset collection process. sWe extend our appreciation to the medical staff and radiologists who contributed their expertise in data validation and clinical insights. Special thanks to the Department of Informatics at Universitas Muhammadiyah Makassar and Asia E University for providing computational resources and research facilities. We also acknowledge the valuable feedback from the anonymous reviewers and the editorial team that helped improve the quality of this manuscript. This research was conducted as part of ongoing efforts to advance healthcare technology in Indonesia and contribute to better diagnostic tools for medical professionals in resource-limited settings.

REFERENCES

[1] G. Garzón, S. Gomez, D. Mantilla, and F. Martínez, "A deep CT to MRI unpaired translation that preserve ischemic stroke lesions," *44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2022, pp. 2708-2711, doi: 10.1109/EMBC48229.2022.9871154.

[2] Z. Lin, T. Wang, and Y. Li, "Reduced Cerebral Blood Flow in Benign Oligemia Relates to Poor Clinical Outcome in Acute Ischemic Stroke Patients," *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021, pp. 3358-3361, doi: 10.1109/EMBC46164.2021.9630180.

[3] S. Utsumi, S. Ohnishi, S. Amagasa, R. Sasaki, S. Uematsu, and M. Kubota, "Role of Routine Repeat Head CT for Pediatric Patients under 2 Years Old with Mild-to-moderate Traumatic Brain Injury," *Neurol Med Chir (Tokyo)*, vol. 62, no. 3, pp. 133–139, Mar. 2022, doi:

10.2176/nmc.oa.2021-0221.

[4] G. Taddei *et al.,* "Evidence-based indications for repeat head CT after mild traumatic brain injury: a systematic review and meta-analysis," *Neurosurgical Review*, vol. 48, no. 1, Art. no. 397, Apr. 2025, doi: 10.1007/s10143-025-03549-3.

[5] H. Abbasi, M. Orouskhani, S. Asgari, and S. S. Zadeh, "Automatic brain ischemic stroke segmentation with deep learning: A review," *Neuroscience Informatics*, vol. 3, no. 4, Art. no. 100145, Dec. 2023, doi: 10.1016/j.neuri.2023.100145.

[6] J. He, "Automated Detection of Intracranial Hemorrhage on Head Computed Tomography with Deep Learning," *in Proceedings 2020 10th International Conference Biomedical Engineering and Technology (ICBET)*, Sept. 2020, pp. 117–121, doi: 10.1145/3397391.3397436.

[7] M. J. Uparela-Reyes, L. M. Villegas-Trujillo, J. Cespedes, M. Velásquez-Vera, and A. M. Rubiano, "Usefulness of Artificial Intelligence in Traumatic Brain Injury: A Bibliometric Analysis and Mini-review," *World Neurosurgery,* vol. 188, pp. 83–92, Aug. 2024, doi: 10.1016/j.wneu.2024.05.065.

[8] E. El Refaee, T. M. Ali, A. Al Menabbawy, M. Elfiky, A. El Fiki, S. Mashhour, and A. Harouni, "Machine Learning in Action: Revolutionizing Intracranial Hematoma Detection and Patient Transport Decision-Making," *Journal of Neurosciences in Rural Practice*, vol. 15, no. 1, pp. 62–68, 2024, doi: 10.25259/JNRP_93_2023.

[9] Y.-R. Chen, C.-C. Chen, C.-F. Kuo, and C.-H. Lin, "An Efficient Deep Neural Network for Automatic Classification of Acute Intracranial Hemorrhages in Brain CT Scans," Computers in Biology and Medicine, vol. 176, Art. no. 108587, Jun. 2024, doi: 10.1016/j.compbiomed.2024.108587.

[10] P. Yesankar, C. Puri and P. M. Gote, "AI-Powered Clinical Decision Support Systems (CDSS): Challenges, Benefits, Applications, and Future Directions," *International Conference on Machine Learning and Autonomous Systems (ICMLAS)*, 2025, pp. 1192-1197, doi: 10.1109/ICMLAS64557.2025.10969014.

[11] M. Khalifa and M. Albadawy, "AI in Diagnostic Imaging: Revolutionising Accuracy and Efficiency," *Computer Methods and Programs in Biomedicine Update*, vol. 5, Art. no. 100146, 2024, doi: 10.1016/j.cmpbup.2024.100146.

[12] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 820–838, May 2021, doi: 10.1109/JPROC.2021.3054390.

[13] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, "Convolutional Neural Networks for Medical Image Analysis: State-of-the-Art, Comparisons, Improvement and Perspectives," *Neurocomputing*, vol. 444, pp. 92–110, Jul. 15, 2021, doi: 10.1016/j.neucom.2020.04.157.

[14] R. Azad, A. Kazerouni, M. Heidari, E. K. Aghdam, A. Molaei, Y. Jia, A. Jose, R. Roy, and D. Merhof, "Advances in Medical Image Analysis With Vision Transformers: A Comprehensive Review," *Medical Image Analysis*, vol. 91, Art. no. 103000, Jan. 2024, doi: 10.1016/j.media.2023.103000.

[15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *International Conference on Learning Representations (ICLR)*, 2021. doi: 10.48550/arXiv.2010.11929.

[16] Z. Liu *et al.,* "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," *International Conference on Computer Vision (ICCV)*, 2021, pp. 9992-10002, doi: 10.1109/ICCV48922.2021.00986.

[17] M. Dialameh, H. Rajabzadeh, M. Sadeghi-Goughari, J. S. Sim, and H. J. Kwon, "DualSwinUnet++: An Enhanced Swin-Unet Architecture With Dual Decoders for PTMC Segmentation," *Computers in Biology and Medicine*, vol. 196, Art. no. 110716, Sep. 2025, doi: 10.1016/j.compbiomed.2025.110716.

[18] J. Zhang, J. Yang, P. Wang, H. Wang, Y. Lin, H. Zhang, Y. Sun, X. Du, Y. Li, Z. Liu, Y. Chen, and H. Li, "OpenOOD v1.5: Enhanced Benchmark for Out-of-Distribution Detection," *Journal of Data-centric Machine Learning Research*, 2023, doi: 10.48550/arXiv.2306.09301.

[19] J. Linmans, G. Raya, J. van der Laak, and G. Litjens, "Diffusion Models for Out-of-Distribution Detection in Digital Pathology," *Medical Image Analysis*, vol. 93, Art. no. 103088, Apr. 2024, doi:

10.1016/j.media.2024.103088.

[20] J. Zhao, X. Ye, B. Li, and Y. Li, "PatchSkip: A Lightweight Technique for Effectively Alleviating Over-Smoothing in Vision Transformers," *Neurocomputing*, vol. 600, Art. no. 128112, 2024, doi: 10.1016/j.neucom.2024.128112.

[21] J. Liang, R. He, and T. Tan, "A Comprehensive Survey on Test-Time Adaptation Under Distribution Shifts," *International Journal of Computer Vision*, Jun. 2024, doi: 10.1007/s11263-024-02181-w.

[22] R. Wu, C. Li, J. Zou, X. Liu, H. Zheng and S. Wang, "Generalizable Reconstruction for Accelerating MR Imaging via Federated Learning With Neural Architecture Search," *IEEE Transactions on Medical Imaging*, vol. 44, no. 1, pp. 106-117, Jan. 2025, doi: 10.1109/TMI.2024.3432388.

[23] T. J. M. Jaspers, T. G. W. Boers, C. H. J. Kusters, M. R. Jong, J. B. Jukema, A. J. de Groof, *et al.,* "Robustness Evaluation of Deep Neural Networks for Endoscopic Image Analysis: Insights and Strategies," *Medical Image Analysis*, vol. 94, Art. no. 103157, May 2024, doi: 10.1016/j.media.2024.103157.

[24] A. Singh and K. K. Singh, "Ensuring Fairness and Mitigating Bias in Healthcare AI Systems," *Responsible and Explainable Artificial Intelligence in Healthcare*. 2025, ch. 5, pp. 107–125. doi: 10.1016/B978-0-443-24788-0.00005-4.

[25] S. M. McKinney *et al.,* "International Evaluation of an AI System for Breast Cancer Screening," *Nature*, vol. 577, no. 7788, pp. 89–94, Jan. 2020, doi: 10.1038/s41586-019-1799-6.

[26] Y. Li and H. Chen, "Image recognition based on deep residual shrinkage Network," *International Conference on Artificial Intelligence and Electromechanical Automation (AIEA)*, 2021, pp. 334-337, doi: 10.1109/AIEA53260.2021.00077.

[27] L. Chato, K. T. Kashyap, K. P. Marupaka and K. Santosh, "Hybrid DL Classification Model Based on CNN and Transformer for Pap Smear Cells," *International Symposium on Biomedical Imaging (ISBI)*, 2025, pp. 1-4, doi: 10.1109/ISBI60581.2025.10981236.

[28] B. Palanisamy *et al.,* "Transformers for Vision: A Survey on Innovative Methods for Computer Vision," *IEEE Access*, vol. 13, pp. 95496-95523, 2025, doi: 10.1109/ACCESS.2025.3571735.

[29] Y. Yang, D. Cheng, C. Fang, Y. Wang, C. Jiao, L. Cheng, and N. Wang, "Diffusion-Based Layer-Wise Semantic Reconstruction for Unsupervised Out-of-Distribution Detection," *Proceeding of the 38th Conference on Neural Information Processing Systems (NeurIPS)*, 2024. doi: 10.48550/arXiv.2411.10701.

[30] D. G. Saragih, A. Hibi, and P. N. Tyrrell, "Using Diffusion Models to Generate Synthetic Labeled Data for Medical Image Segmentation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 19, no. 8, pp. 1615–1625, Aug. 2024, doi: 10.1007/s11548-024-03194-4.

[31] B. H. Reddy and K. P. R, "Classification of Fire and Smoke Images using Decision Tree Algorithm in Comparison with Logistic Regression to Measure Accuracy, Precision, Recall, F-score," *2022 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS),* 2022, pp. 1-5, doi: 10.1109/MACS56771.2022.10022449.

[32] T. Eelbode *et al.,* "Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index," *IEEE Transactions on Medical Imaging,* vol. 39, no. 11, pp. 3679-3690, Nov. 2020, doi: 10.1109/TMI.2020.3002417.

[33] M. I. Ahmed, B. Spooner, J. Isherwood, M. Lane, E. Orrock, and A. Dennison, "A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare," *Cureus*, vol. 15, no. 10, Art. no. e46454, Oct. 2023, doi: 10.7759/cureus.46454.

[34] D. Morwani, N. Vyas, H. Zhang, and S. Kakade, "Connections Between Schedule-Free Optimizers, AdEMAMix, and Accelerated SGD Variants," *arXiv preprint arXiv:2502.02431*, 2025. doi: 10.48550/arXiv.2502.02431.

[35] Z. Hong, Y. Yue, Y. Chen, L. Cong, H. Lin, Y. Luo, *et al.,* "Out-of-Distribution Detection in Medical Image Analysis: A Survey," *arXiv preprint arXiv:2404.18279,* 2024. doi: 10.48550/arXiv.2404.18279.

[36] T. Tu *et al.,* "Towards Generalist Biomedical AI," *Nature Medicine*, Jul. 2024, doi: 10.48550/arXiv.2307.14334.

[37] A. Yala *et al.,* "Multi-Institutional Validation of a Mammography-Based Breast Cancer Risk Model," *Journal of Clinical Oncology : official journal of the American Society of Clinical Oncology*, vol. 40, no. 16, pp. 1732–1740, Jun. 1, 2022.