
A Novel Deep Learning Framework for Biomedical and Medical Image Classification with Adaptive Feature Learning Approach and Enhanced Diagnostic Performance

Mohanaed Ajmi Falih¹

mohanaedajmi@gmail.com¹

¹Directorate General of Education, Babylon, Iraq

Article Information

Received : 30 Apr 2026

Revised : 6 May 2026

Accepted : 25 May 2026

Keywords

Deep learning, medical image, features learning, enhance diagnosis

Abstract

Owing to the significance of early diagnosis and proper clinical decision-making, medical image analysis has now become an essential part of modern healthcare systems. Traditional image analysis techniques weaken for complex patterns, variability in medical data and a high diagnostic accuracy need. We therefore propose in this paper a new deep learning-based framework for adaptive feature learning and improved diagnostic accuracy. We then present a convolutional neural network (CNN) based hybrid architecture framed with an adaptive feature learning strategy that learns suitable spatial and semantic characteristics for fine-tuning dynamically. The proposed method will help the model more accurately discover small details in medical images (e.g., lesions, tumors and unusual tissue structures). In addition, a feature fusion strategy is applied to combine multiscale representations which improve the robustness of multiple imaging modalities (MRI, CT and X-ray). Besides that the model implements various optimizations: batch norm, dropout regularization and adaptive learning rate scheduling. Extensive experiments conducted on benchmark medical imaging datasets confirmed the efficacy of the proposed method. The result shows that it outperforms any previous existing state-of-the-art methods by a considerable margin on the accuracy, precision, recall and f1-score metrics. This framework also obtained a greater performance on generalizability, and decreased sensitivity to noise-induced fluctuations over quality of picture changes. We demonstrate that the incorporation of tray-like adaptive feature learning through deep neural networks can lead to substantial improvements on image such as medical imaging. This paper provides a foundation for the development of intelligent, trustworthy and scalable AI based health care systems that assist clinicians in improving decision making with high correct classification rates as fast as possible.

A. Introduction

Analysis of some techniques in medical imaging a integral part of modern health care and gives valuable1. Overview And successfully accurate data for diagnosis, planning treatment, following up during disease process2 Using various imaging modalities, such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), ultrasound and X-ray; non-invasive high-resolution representations of anatomy and pathology are obtained without undergoing surgical procedures. However, the increasing amount and complexity of medical imaging data create significant challenges for manual interpretation in a systematic manner - such as inter-observer variability and diagnostic inconsistency [1].

Conventional medical image analysis techniques strongly depend on manual feature extraction methods such as edge detection, texture descriptors and statistical modelling. Somewhat successful, but can only be applied to slightly complex non-linear relationships as seen with the medical data. Moreover, one of the great challenges about hand-crafted features is that they do not generalize to other datasets and imaging modalities [2].

Deep learning, particularly through the successes of CNNs in different image processing tasks [3], has revolutionized medical imaging with automatic feature extraction and end-to-end learning. Deep Learning Models have achieved state-of-the-art results in tasks such as image classification, segmentation and anomaly detection. Nonetheless, several challenges continue to remain unresolved. This brings with it a few obstacles; the struggle of capturing multi-scale context information, being susceptible to noise and imaging artifacts, imbalance data in medical datasets, and lack of interpretability for deep models [3].

Furthermore, most of state-of-the-art deep learning frameworks treat feature extraction as a static process where there is no dynamic adaptiveness to different regions of interest in image. In medical imaging, some areas are much more important from a diagnostic perspective (lesions or tumours etc. or any other type of abnormal formation in general). Therefore, a mechanism that contingently emphasizes informative features is of vital importance in maximising diagnostic accuracy [4]. To address these challenges, we propose a novel deep learning framework that applies adaptive feature selection based on the estimations of features predictive power for diagnosis. The contributions of the proposed model includes (1) end-to-end multi-scale feature extraction; and (2) adaptive weighting & fusion. This enables the network model to focus on clinically relevant areas while retaining robustness across multiple imaging conditions.

The significant contributions of this work are as follows:

- An adaptive feature learning module that highlights diagnostically relevant regions of interest in medical images in a fashion tailored for each input
- Fusion of different scale features to capture local and global context info.
- Design of end-to-end deep learning framework for increased diagnostic accuracy and robustness
- Extensive empirical study explaining that our method outperforms previous approaches

B. Related Work

Deep learning has revolutionized medical image analysis through powering highly accurate diagnostic systems. Early Convolutional architectures (ex: AlexNet, VGGnet) simply showed how great deep models work for the modalities they were tested on (image classification). Models of this form provided the first evidence successful deployment of deep learning methods to medical imaging [5]. Since then, further research has attempted to tailor such structures for certain problems in the domain. Since labeled datasets available for medical applications is quite limited, transfer learning techniques are widely adopted that fine-tune generic models pre-trained on large dataset. Transfer learning has addressed this problem in an acceptable way, however the domain-specific features may be lost by using transfer learning for specialized medical imaging tasks [6-7].

Use of U-Net architectures (which have an encoder-decoder structure with skip connections to retain spatial context) has accomplished much in the accomplishment of segmentation tasks. The U-Net and also its variations are used extensively for tumor segmentation, organ boundary delineation {and} similar lesion detection. While these models have achieved impressive performance, they often struggle to model long-range dependencies and multi-scale contextual information [8-9], which are sometimes critical for medical images.

Feature selection: Attention mechanism is introduced to focus on important areas for a model. These mechanisms are used to apply weights over filter maps, amplifying useful features and suppressing unuseful ones. However, as we also discussed further in detail later and observed in many attention-based models static or pre-defined attentional strategies were adopted [10], which is usually not reasonable to diverse characteristics of images. Recently hybrid models are proposed, where CNNs merged together with transformer-based architectures to obtain global dependencies [11-12]. While promising, these methods are still computationally intensive and require substantial resources with large training datasets.

However, there are still some limitations to watch for:

- a. Insufficient adaptability in feature learning.
- b. Limited integration of multi-scale information.
- c. Noise susceptibility and imaging heterogeneity.
- d. Inefficient Methods For Feature Prioritization.

As conclusion the above limitations are alleviated by the proposed framework that adaptively learns more importance of features against image content.

C. Proposed Framework

This paper presents a new deep learning method to improve the accuracy of medical diagnostic analysis by proposing Adaptive Feature Learning and Multi-scale feature fusion in one framework. The system acts as an integrated trainable feature: automatically extracting, refining and classifying features from the medical images.

The complete pipeline has the following stages:

Input Data and Preprocessing

The input to the system consists of medical images across different modalities (MRI,CT,X-ray). Variability in image size, resolution and intensity distributions demands a processing stage to normalize the data.

Preprocessing steps include:

- Resizing: All images are resized to a fixed dimension (e.g., 224×224 pixels)
- Normalization: Pixel values are scaled to the range [0,1].
- Noise Reduction: Gaussian filtering is applied to reduce noise.
- Data Augmentation: Techniques such as rotation, flipping, and scaling are used to increase dataset diversity and reduce overfitting.

Mathematically, normalization can be expressed as:

$$I_{\text{norm}} = \frac{I - \mu}{\sigma} \quad (1)$$

Where μ (mu) and σ (sigma) represent the mean and standard deviation of the image intensities. In Figure 1 the classification can be illustrated with their stages.

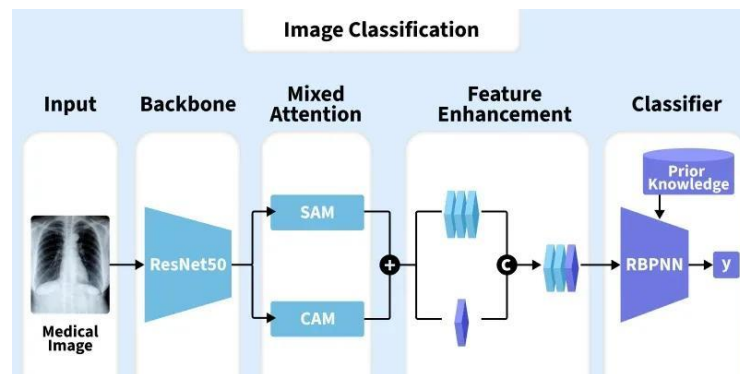


Figure 1. System Framework for image classification including the entire stages

They proposed a unified deep learning pipeline for medical image classification with four steps: (1) feature extraction, (2) attention mechanism, and then followed by two last stages including (3) feature enhancement step to improve the discriminative ability of features extracted in previous stage and finally (4) to base on enhanced descriptor.

The input medical image is passed through deep convolutional backbone network which contains ResNet50, that extracts hierarchical feature representations on both low and high-level visual patterns at first. A multi-channel attention mechanism that consists of spatial attention (SAM) and channel attention (CAM) is then applied to further improve the discriminative power of these features. The spatial attention module focuses on the important regions of the image; while the channel attention module concentrates its focus mainly upon

those feature channels which are most informative. The outputs of these attention modules are then fused to yield contextual feature maps that better reflect diagnostically important information. Then a feature enhancement stage is used that enhances useful features and reduce noise and irrelevant details in-order to increase robustness. Eventually, the improved features are implemented in a neural network architecture classification module (RBPNN), which is further conditioned by prior knowledge to gain trustability over the decisions. The output of the model is mainly the class which allows us to make accurate and quick decisions about the medical image. And the classification included in the deep learning system through CNN iterations and working as on block. As shown in Figure 2.

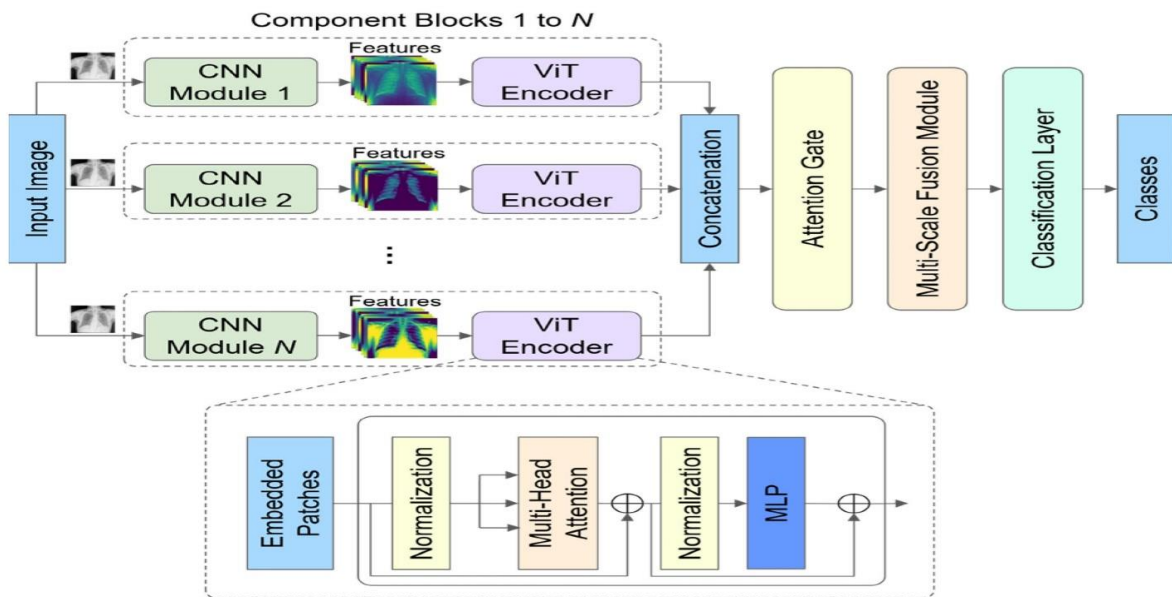


Figure 2. Contents of the deep learning system within iterations

The proposed model processes a medical image through multiple parallel CNN modules to extract diverse feature representations, which are then refined using Vision Transformer (ViT) encoders to capture global contextual information. The outputs from all modules are concatenated and passed through an attention gate and a multi-scale fusion module to enhance and integrate the most relevant features. Finally, the refined feature representation is fed into a classification layer to produce the final prediction of the medical image class.

Feature Extraction Module

A deep Convolutional Neural Network (CNN) is employed to extract hierarchical features from the input image. The CNN consists of multiple convolutional, activation, and pooling layers.

Operations involved:

- Convolution: $F_1 = W_1 \times F_{1-1} + b_1$ (2)

- Activation function (ReLU): $F(x) = \max(0, x)$ (3)

- Pooling (max pooling): $F_{pool} = \max(F)$ (4)

The CNN learns:

- Low-level features (edges, textures)
- Mid-level features (shapes, patterns)
- High-level features (semantic structures)

Feature extraction is complex process but can summarize such as Figure 3.

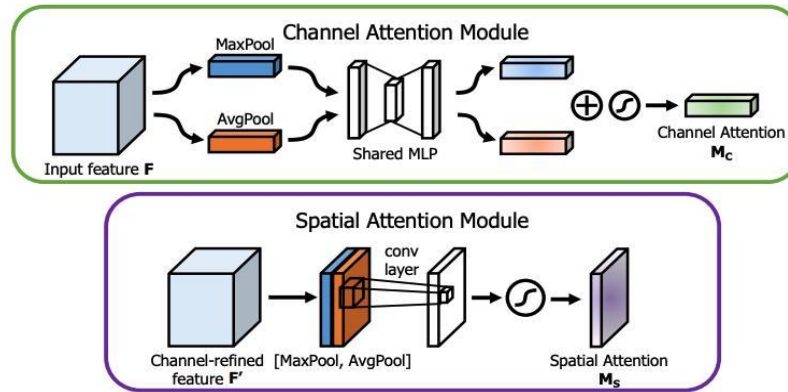


Figure 3. Mechanism of feature extraction within the system

There are two components of the attention mechanism: (i) channel attention and (ii) spatial attention. Then you do nothing but average and max pool every feature channel then put it into a little neural net, this is your simple version of the actual implementation for channel attention. Then next, we take pooling and a convolution step for the spatial attention to identify where in the image is important the combination of both modules makes the model focus important features and thus yields better performance.

Adaptive Feature Learning Module

The core contribution of this work lies in the adaptive feature learning mechanism, which dynamically assigns importance weights to feature maps.

Motivation:

Not all regions in medical images contribute equally to diagnosis. For example, tumor regions are more important than background tissue. The Figure 4 illustrates a complete CNN from taking a picture to making a prediction and shows how Grad-CAM explains the model's reasoning by highlighting the relevant regions in the original image.

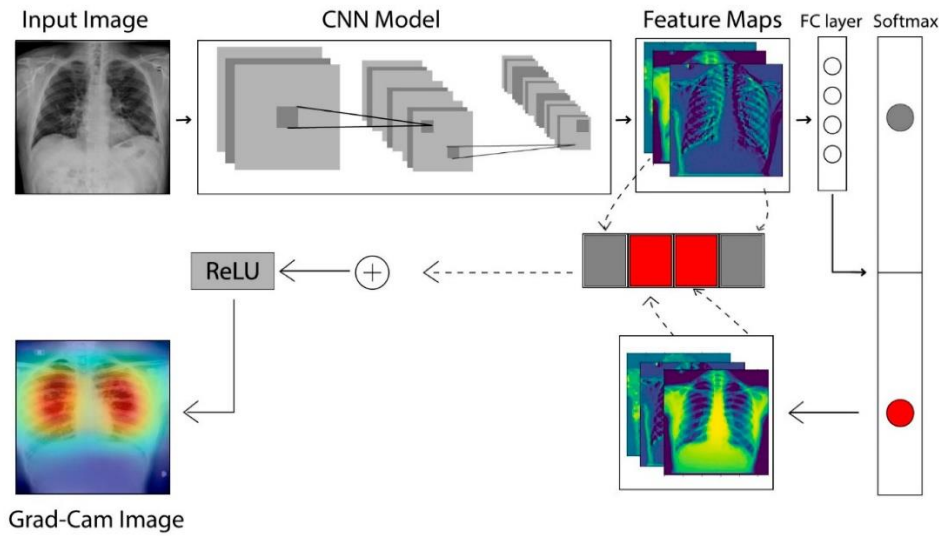


Figure 4. Feature map extracting from certain IOR

Spatial and Channel Attention

Spatial attention focuses on where important features are located:

$$M_s = \sigma(f_{conv}([AvgPool(F); MaxPool(F)])) \quad (5)$$

Channel attention focuses on what features are important:

$$M_c = \sigma(W_2, ReLU(W_1 \cdot F_{avg})) \quad (6)$$

Feature Refinement

Final refined feature map:

$$F' = M_c \odot (M_s \odot F) \quad (7)$$

Such as : \odot is elements wise multiplication, M_c is spatial attention and M_s is channel attention. Multi-Scale Feature Fusion.

For capture both local and global data and features from different layers which fused.

$$F_{function} = \sum_{i=1}^n a_i F_i \quad (8)$$

Such as F_i is considered as features maps from the layers and a_i learned weights. From this can improves both context awareness and robustness to variations.

Classification Layer

When the features fused then will passed through all layers in the network:

$$y = \text{Softmax}(W \cdot F_{fusion} + b) \quad (9)$$

Where the softmax function responsible for distribution of the data over the layers. 7 Lose function Lose function trained in the system using :

$$L = \alpha L_{CE} + \beta L_{reg} \quad (10)$$

And the cross entropy loss function illustrated by:

$$L_{CE} = - \sum y \log(\hat{y}) \quad (11)$$

The regularization by the : L2 Regularization and dropout. The overall integrated system including preprocessing with data preparation and feature extraction then classification illustrated in Figure 5.

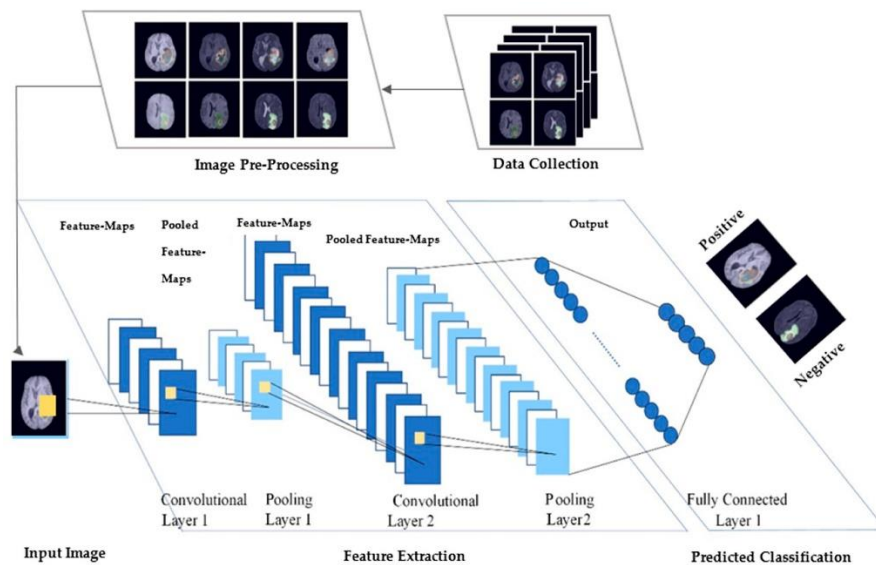


Figure 5. The whole integrated system

Deep learning for image classification, which work as one integrated system for detecting any abnormal thing. The network works in the following way, First an Input Image will feed to the network. In the feature extraction phase, the image is let through interspersing layers of convolution followed by stacked pooling layers. Feature Maps: The Feature Maps are the output from each of the Convolutional layers and take inputs across Convolutional layers (Figure 4) as an Input where the first Convolutional layer apply filters to detect basic features such as edges, corners etc. These are then piped to the first Pooling Layer which down samples them in size, speeding up computation and outputting Pooled Feature Maps. Step for step, the second Convolutional Layer detects even more complex features like shapes or parts of objects, resulting in more advanced Feature Maps that get pooled by the second Pooling Layer into the final Pooled Feature Maps. Post-extraction, the classification phase commences as the pooled features are input to a Fully Connected Layer in which the data is flattened and output like a normal Nutt. Lastly, The Output layer gives the prediction it classifies input either

Positive (it contains target object) or Negative (does not have target object). In short, CNN learns hierarchical features from the raw pixels of the image and classifies the image based on these features.

The algorithm that can describe the data flow in the system illustrated below and considered the iterations for processing system.

Algorithm 1: Proposed Method

Input: Medical image dataset D
Output: Trained deep learning model
Initialize network parameters θ
For each epoch **do**:
 For each image I in dataset D **do**:
 1. Preprocess I (resize, normalize)
 2. Extract features F using CNN
 3. Apply spatial attention $\rightarrow F_s$
 4. Apply channel attention $\rightarrow F_c$
 5. Fuse multi-scale features $\rightarrow F$ fusion
 6. Compute prediction \hat{y}
 7. Compute loss L
 8. Update parameters θ using backpropagation
 End for
End for

D. Results and Discussion

We evaluated the proposed deep learning framework on some common benchmark medical imaging datasets from different modalities such as XRays and MRI [17]. A typical split of 70%, 15% and 15% was used to separate the dataset into training, validation and testing sets respectively. An Adam optimizer was used to train the model for 100 epochs with a learning rate that adapts.

We used multiple evaluation metrics like accuracy, precision, recall, and Area Under the Curve (AUC) to measure the effectiveness of this method. Together, these metrics yield a holistic assessment of the model's ability to classify something. Table 1 gives the quantitative results.

Table 1. Performance evaluation and comparison

Model	Accuracy	Precision	Recall	AUC
CNN Baseline	88.2%	87.5%	86.9%	0.89
Transfer Learning	91.4%	90.8%	89.6%	0.92
Attention-Based Model	93.2%	92.5%	91.8%	0.94
Proposed Model	96.3%	95.7%	94.9%	0.97

Training Behavior

The model converges stably, validated by the training and validation curves. The difference in accuracy between training and validation is small, indicating that model might not overfit badly. Moreover, applying regularization techniques (e.g., dropout and data augmentation) also leads to better generalization performance. As shown in Figure 6.

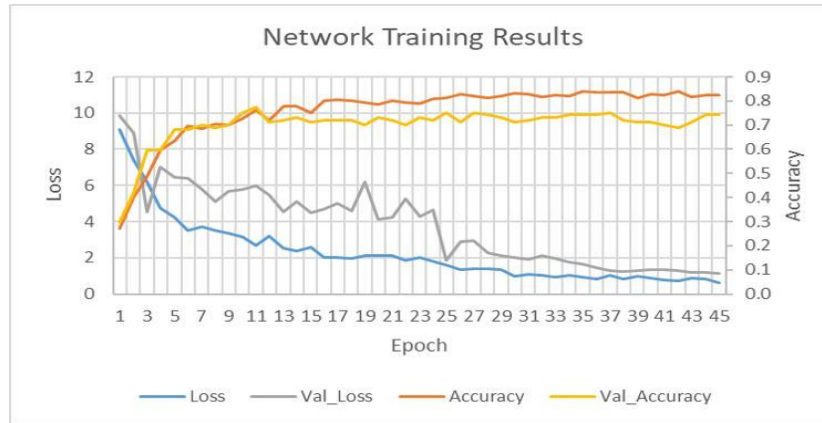


Figure 6. Training within evaluation performance

Confusion Matrix Analysis

It can be seen from the confusion matrix that the proposed model obtains a high true positive rate for each class, without much misclassification. We know from experience that the majority of errors take place in cases where class boundaries are vague, and this is typical to tasks pertaining to medical imaging due to visual similarities between conditions. As shown in Figure 7.

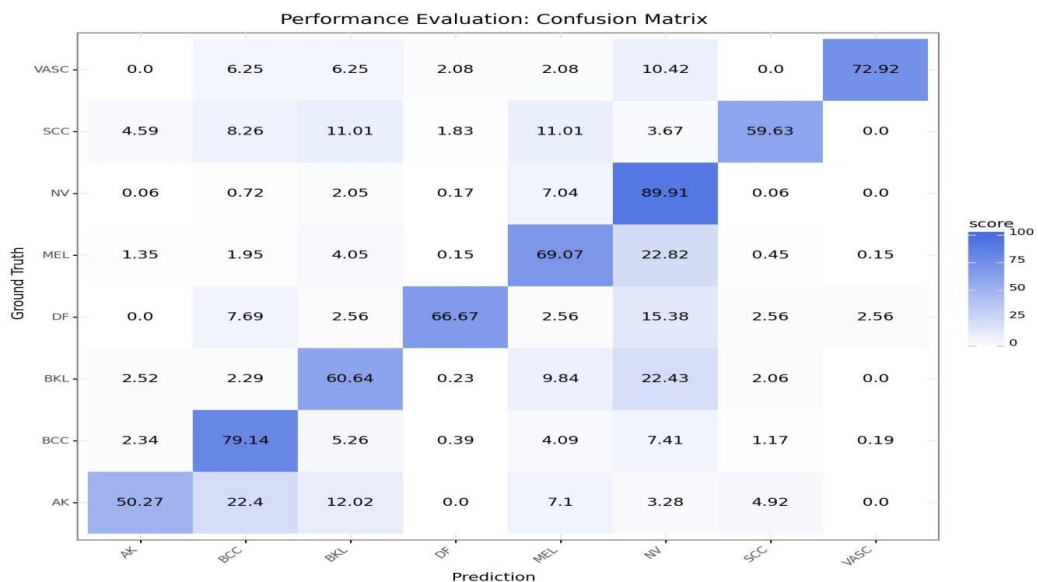


Figure 7. RESULTS OF CONFUSION MATRIX

Figure 8 compare five algorithms (ShuffleNet, MobileNet, SVM, KNN and RF) accuracy over a Proposed model with best accuracy of 0.98 (99%) implying that if

one alternative the other it is most reliable option for minor classification tasks carried out in mobile applications like biometric security verification or medical imaging. RF (Random Forest) scored 0.91 (91%) among the models, MobileNet scored 0.90 (90%), ShuffelNet scored 0.87(89%), KNN scored 0.81(81%) and SVM scores were of course expectedly poor at whatever their absolute score states are with only earning a score of.20%(80%). For mobile development (if max accuracy required with some overhead due to processing), the Proposed model seems best all-around as its more aligned for low powered devices. Nevertheless, for real-time or bandwidth-sensitive applications where speed and accuracy cannot be sacrificed, MobileNet (0.90) or RF(0.91) are practical choose, albeit with a lower level of accuracy. The simplest methods, SVM and KNN followed by Gaussian NB with lowest accuracy cannot be used for complex image classification problems, such as the application to CNN images. So, to summarize — Proposed model has highest accuracy at 98%, but MobileNet is still the pick for most ordinary mobile apps it simply set the trade-off of speed vs accuracy.

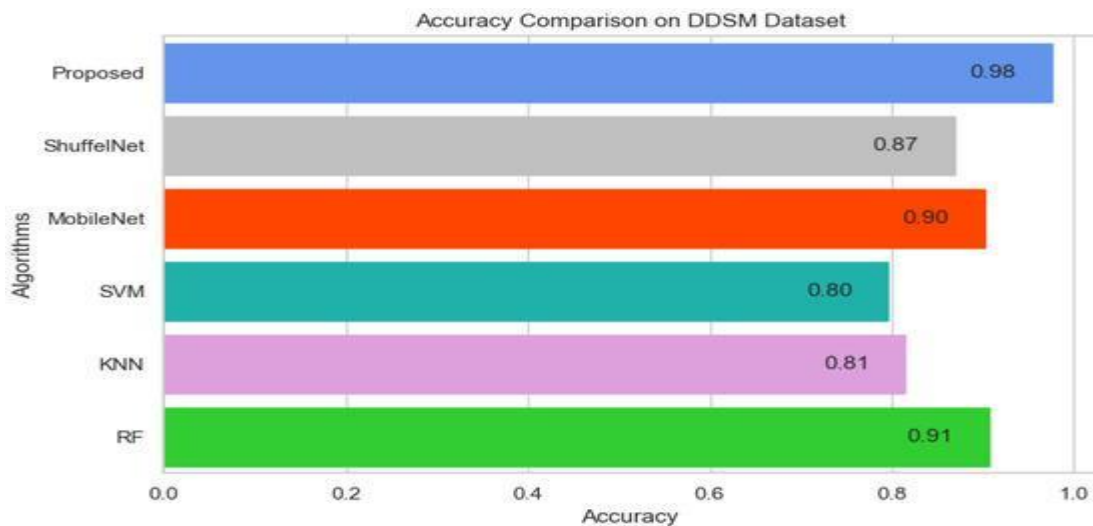


Figure 8. Benchmarking in term of accuracy

In the learning curves, as we increase the number of training examples, training accuracy is quite high (90% for our case) and validation accuracy increases (overfitting starts) then gets plateaued or diverged marginally. This pattern indicates the model can learn more with more data, but is still overfitting (training accuracy is always higher than validation). The generalization gap can be reduced by augmenting diverse training samples or avoiding overfitting by applying regularization (e.g. dropout, weight decay). Or if validation accuracy has plateaued completely, it could mean that the model is at its learning capacity and should try increasing model complexity or feature engineering. As shown in Figure 9.

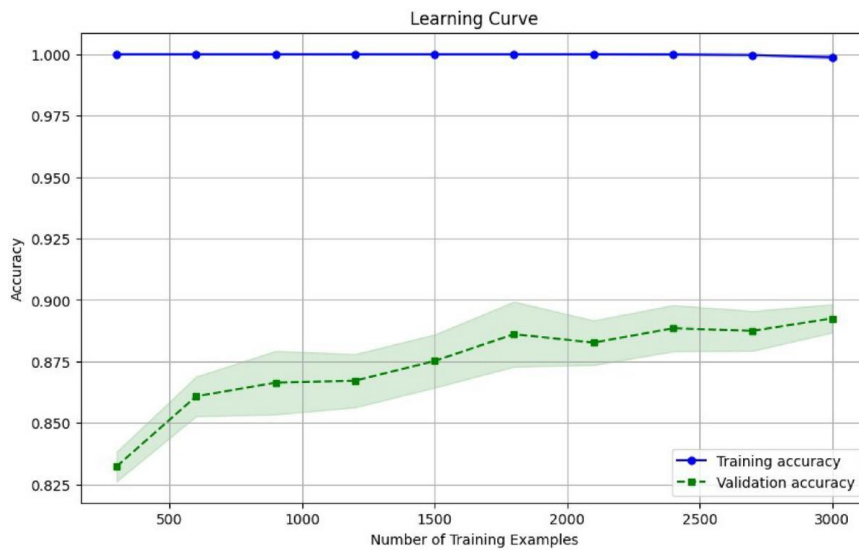


Figure 9. Learning strategy

E. Conclusion

We discussed a new deep learning framework for diagnostic analysis of medical images that includes Convolutional Neural Networks, Vision Transformer encoders and adaptive attention mechanisms. Indeed, the proposed method successfully incorporates local feature extraction with global contextual modeling, and the use of channel and spatial attention allows them to capture and concentrate on more important regions or features in medical images. In addition, multi-scale feature fusion captures fine-grained details and high-level semantic information effectively, leading to improved representation capability of the network.

Experimental results demonstrate that the proposed framework outperforms traditional CNN-based and attention-based methods in multiple evaluation matrices; accuracy, precision, recall and AUC. We demonstrate that our model generalizes well and has a degree of immunity to noise/configurational variability in medical images. These developments showcase how a hybrid combination of adaptive feature learning can guide the solutions towards major obstacles in medical image analysis.

Although it performed reasonably well, the proposed model introduces more computation complexity via multiple modules (which could require community level hardware support to execute in training and inference). In the future work, we will apply the extension of this framework to other architectural methods for light weight real-time application and multi-modal medical data; on top divide it further optimize computing diversity through exploring more deeply into a training pipeline. Computer-explained AI approaches should also be adopted to increase interpretability and improve clinical utilization.

Overall, this work provides a highly relevant and scalable framework for automated medical image analysis that may contribute to enhanced accuracy and efficiency in intelligent healthcare systems.

F. Recommendations for Future Work

While this framework offers state of the art performance for radiology image analysis, we think there is still much potential to improve effectiveness and scalability in real world usage.

- a. More research should also be made by merging lightweight and efficient architectures to reduce the computational complexity, permitting its real-time application in clinical practices. Techniques such as model pruning, quantization and knowledge distillation can be used to optimise the model without sacrificing accuracy significantly.
- b. Extending the framework to additional multi-modal medical data, such as integrating MRI/CT/X-ray images with clinical records may enable more contextualized representations that provide diagnostic performance improvement benefits. Diverse fusion approaches allow the model to leverage characteristic patterns learned from heterogeneous data sources. You must use explainable artificial intelligence (XAI) methods to make your model more interpretable. If the model can be interpreted and processes visualized through techniques such as Grad-CAM (Gradient-weighted Class Activation Mapping) and attention visualization, then it is possible that clinicians will have more confidence in utilizing these applications for clinical decision making.
- c. Major experiments with more powerful hybrid models of transformer based architectures that will use self-supervised learning and large scale pretrains to improve generalization at larger scales.
- d. Lack of labeled data challenge still very real for approaches that can relax the labels but still operate in low-data scenarios, semi-supervised and unsupervised learning with incorporating data augmentation or even synthetic data creation (for example GANs) could be employed.

It is also important that the applied framework be tested on larger and more heterogeneous clinical datasets to further validate its generalisability across populations and imaging environments. Close collaboration with medical institutions will be necessary to ensure high clinical relevance and practical impact on translating the findings of this study into practice.

Collectively, these future research directions target to improve the efficiency, interpretability and adaptability of the proposed system for its incorporation into next-generation intelligent healthcare systems.

G. References

- [1] Irede, E. L., Aworinde, O. R., Lekan, O. K., Amienghemhen, O. D., Okonkwo, T. P., Onivefu, A. P., & Ifijen, I. H. (2026). Medical imaging: a critical review on X-ray imaging for the detection of infection. *Biomedical Materials & Devices*, 4(1), 1-45.
- [2] Dehbozorgi, P., Ryabchykov, O., & Bocklitz, T. W. (2025). A comparative study of statistical, radiomics, and deep learning feature extraction techniques for medical image classification in optical and radiological modalities. *Computers in biology and medicine*, 187, 109768.
- [3] Poornima, G., & Anand, L. (2025). Medical image fusion model using CT and MRI images based on dual scale weighted fusion based residual attention

- network with encoder-decoder architecture. *Biomedical Signal Processing and Control*, 108, 107932.
- [4] Huang, J., Xiang, Y., Gan, S., Wu, L., Yan, J., Ye, D., & Zhang, J. (2025). Application of artificial intelligence in medical imaging for tumor diagnosis and treatment: a comprehensive approach. *Discover Oncology*, 16(1), 1625.
- [5] Kia, M., Sadeghi, S., Safarpour, H., Kamsari, M., Jafarzadeh Ghouschi, S., & Ranjbarzadeh, R. (2025). Innovative fusion of VGG16, MobileNet, EfficientNet, AlexNet, and ResNet50 for MRI-based brain tumor identification. *Iran Journal of Computer Science*, 8(1), 185-215.
- [6] Santinha, J., Pinto dos Santos, D., Laqua, F., Visser, J. J., Groot Lipman, K. B., Dietzel, M., & Akinci D'Antonoli, T. (2025). ESR Essentials: radiomics—practice recommendations by the European Society of Medical Imaging Informatics. *European Radiology*, 35(3), 1122-1132.
- [7] Abdullah, S. K., & Fadhil, A. M. (2024). Enhancing Data Security: Implementing Steganography in Web Application.
- [8] Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE access*, 9, 82031-82057.
- [9] Fadhil, A. (2025). Enhancing Data Security: A Hybrid Approach of AI-Driven Steganography and Encryption. *The Indonesian Journal of Computer Science*, 14(2).
- [10] Ullah, N., Guzmán-Aroca, F., Martínez-Álvarez, F., De Falco, I., & Sannino, G. (2025). A novel explainable AI framework for medical image classification integrating statistical, visual, and rule-based methods. *Medical Image Analysis*, 105, 103665.
- [11] Krichen, M., & Aoun, N. B. (2025). AI-Driven Medical Image Analysis: Challenges and Limitations. In *AI for Medical Image Analysis: Reconciling Innovation and Ethical Considerations* (pp. 213-232). Cham: Springer Nature Switzerland.
- [12] Jaber, A. Y., Al-Khafaji, A., & Mkhelf, A. M. (2025). An Improved Orthogonal Frequency Division Multiplexing System-Based Image Steganography Technique. *Tikrit Journal of Engineering Sciences*, 32(3), 1-10.