2025 حزيران العـــدد 17S
No.17S JUNE 2025
المجلة العراقية للبحوث الإنسانية والإجتماعية والعلمية
Iraqi Journal of Humanitarian, Social and Scientific Research
Print ISSN 2710-0952 Electronic ISSN 2790-1254

A Novel Swin Transformer Variant Without Embedding for Accurate Cardiomyopathy Classification Using CMR Imaging

Bushra Kamil Hilal [1,2] Ebtesam Najim AlShemmary [3*]
[1]Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Kufa, Iraq
[2]Department of Computer Information Systems, Faculty of Computer Science and Information Technology,
Al-Qadisiyah University Al- Qadisiyah, Iraq
{bushrak.almiahy@student.uokufa.edu.iq ; bushra.k.h@qu.edu.iq}
[3]IT Research and Development Center, University of Kufa, Kufa, Iraq
* Corresponding author's Email: dr.alshemmary@uokufa.edu.iq

Abstract :

This paper discusses and assesses diverse deep learning architectures to classify cardiac magnetic resonance (CMR) images to hypertrophic cardiomyopathy (a kind of heart disease). The study employs a relatively big sample size of 59,145 images of both the healthy persons and those diagnosed with hypertrophic cardiomyopathy and combats the class issue by applying stratified sampling and accurate assessment methods. Preprocessing to be done on the images included resizing in 224 x 224 pixel size with preservation of aspect ratios, normalizing pixel intensities to range [-1.0, 1.0], and data storage format with HDF5 to speed up the training process. Rotation, translation, zooming and horizontal flipping were performed in real-time to augment data in an attempt to generalize.Several architectures were evaluated, and they included classic CNN structures to more advanced Swin Transformer based, with some with a combination in which the Swin Transformer is integrated excluding the embedding and patch embedding layers. Findings indicated that the MobileNetV2 with Swin Transformer (without embedding and patch embedding) attained the best accuracy of 96.22 percent. This detailed paper underlines the importance of proper model selection and architectural design to enhance the accuracy of the disease classification of cardiomyopathy based on CMR data, which would finally lead to better early disease screening machines and medical decision support. The discovery sets the path to the development of more effective and reliable models with which healthcare can be advanced.

Keywords: DenseNet121, Swin Transformer(without embedding and patch), CMR, CNN, Transfer Learning.

متغير جديد لمحول سوين بدون تضمين لتصنيف دقيق لاعتلال عضلة القلب باستخدام التصوير بالرنين المغناطيسي القلبي

بشرى كامل هلال [1، 2] ابتسام نجم الشمري [3*]

قسم علوم الحاسوب، كلية علوم الحاسوب والرياضيات، جامعة الكوفة، العراق [1]
قسم نظم المعلومات الحاسوبية، كلية علوم الحاسوب وتكنولوجيا المعلومات، جامعة القادسية، العراق [2]
{bushrak.almiahy@student.uokufa.edu.iq ; bushra.k.h@qu.edu.iq}
مركز أبحاث وتطوير تكنولوجيا المعلومات، جامعة الكوفة، العراق [3]
البريد الإلكتروني للمؤلف: dr.alshemmary@uokufa.edu.iq [*]

ملخص: تناقش هذه الورقة البحثية وتُقيّم بنى التعلم العميق المتنوعة لتصنيف صور الرنين المغناطيسي للقلب (CMR)إلى اعتلال عضلة القلب الضخامي (نوع من أمراض القلب). تستخدم الدراسة عينة كبيرة نسبيًا، تبلغ 59,145 صورة، لأشخاص أصحاء ومُشخصين باعتلال عضلة القلب الضخامي، وتعالج مشكلة التصنيف من خلال تطبيق أخذ العينات الطبقي وأساليب التقييم الدقيقة. تضمنت المعالجة المسبقة للصور تغيير حجمها إلى 224 × 224 بكسل مع الحفاظ على نسب العرض إلى الارتفاع، وتطبيع شدة البكسل إلى نطاق [-1.0، 1.0]، وتخزين البيانات باستخدام HDF5 لتسريع عملية التدريب. أُجريت عمليات التدوير والتحويل والتكبير والانعكاس الأفقي في الوقت الفعلي لزيادة البيانات في محاولة لتعميمها. تم تقييم العديد من البنى، بدءًا من هياكل CNN الكلاسيكية وصولًا إلى بنى Swin Transformer الأكثر تطورًا، مع دمج بعضها مع محول Swin باستثناء طبقات التضمين والتضمين الرقعي. أشارت النتائج إلى أن نموذج MobileNetV2مع محول Swin (بدون التضمين والتضمين الرقعي) حقق أفضل دقة بنسبة 96.22 %. تُبرز هذه الورقة البحثية المُفصلة أهمية اختيار النموذج والتصميم الهيكلي المُناسب لتعزيز دقة تصنيف مرض اعتلال عضلة القلب بناءً على بيانات التصوير بالرنين المغناطيسي القلبي، مما سيؤدي في النهاية إلى تطوير أجهزة فحص مُبكر للمرض ودعم القرارات الطبية. يُمهد هذا الاكتشاف الطريق لتطوير نماذج أكثر فعالية وموثوقية تُمكّن من تطوير الرعاية الصحية.

1. Introduction

Hypertrophic cardiomyopathy (HCM) is a hereditary heart disease which involves thickening of the left ventricular myocardium resulting in poor blood flow, arrhythmia, and sudden death heart attack in certain situations. Symptoms of patients include chest pains, chest dyspnea, fainting, predisposed heart murmur, and palpitation. HCM is strongly underdiagnosed as about 80-90 percent of cases do not get noticed, which makes HCM a growing global burden because it is seen in up to 200 cases per 100,000 people[1].

An early and proper diagnosis of HCM is vital to treatment, risk avoidance, and fatalities. The echocardiography and cardiac magnetic resonance imaging (CMR) are the most widely used to diagnose myocardial morphology and functioning. Of these, CMR has become the gold standard because it has better imaging of the myocardial fibrosis, wall thickness and microscopic phenotypic characters that can be missed in other next modalities. In addition, CMR will eliminate the invasive procedures like biopsies[2].

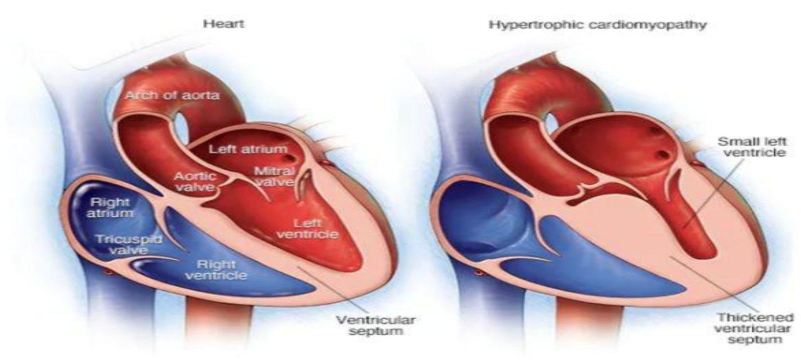
These developments notwithstanding, the manual interpretation of the CMR images is still time-consuming and subjective. Recently, artificial intelligence (AI) and deep learning (DL) in specific have transformed the way medical imaging can be analyzed and have introduced automated methods of assessing, segmenting, and

Print ISSN 2710-0952 Electronic ISSN 2790-1254

classifying cardiac diseases. Transformer and convolutional neural network (CNN) designs have proven to be exceptionally effective when dealing with the identification of pathological features directly in the imagery data, increasing the findings that can be prosecuted in diagnosis and clinical decision-making[3].

With the complexity and the heterogeneity of the cardiomyopathies that are divided into dilated, hypertrophic and restrictive cardiomyopathies, objective based image-based classification is a critical problem. The proposed study will fill the gap by comparing and contrasting the effectiveness of different deep learning models, namely, CNN, DenseNet121, MobileNetV2, and Xception along with their hybrid versions that incorporate the Swin Transformer (no embedding or patch embedding) and binary CMR dataset of large scale. The issue examined in the study is the influence of model architecture and transformer design on improving classification accuracy to enhance early identification of patients with HCM and cure their condition better[4,5].

To enhance the clinical and visual comprehension of HCM, the anatomical distinctions of a healthy heart as compared to a heart with hypertrophic cardiomyopathy will be described and detailed in figure -1. The figure clearly shows the thickened ventricular septum and reduced left ventricular chamber size in the HCM heart, which are among the key features targeted by deep learning models in CMR image classification[6].



Hypertrophic cardiomyopathy

Illustrations of a regular heart (left) and a heart with hypertrophic cardiomyopathy. Note that the heart walls (muscles) are much thicker (hypertrophied) in the heart with hypertrophic cardiomyopathy.

Figure 1: How the heart with Hypertrophic Cardiomyopathy (HCM) and a healthy heart differ anatomically.

2. Related Work

Kolluri and Hathwar (2024) suggested that two deep learning models were produced to aid hypertrophic cardiomyopathy (HCM) diagnosis with cardiac magnetic resonance (CMR) and electrocardiogram (EKG) data. Both convolutional neural network (CNN) and bidirectional long short-term memory (LSTM) networks were suggested as the models to be applied to the automatic classification of CMR images and EKG signals, respectively. They have a high performance measure in the CNN model with an accuracy of 94.71% and a precision of 96.97% with a recall of 91.21 and the F1-score of 94.85%[7].

Agibetov et al. (2021) suggested completely automatic diagnostic system to identify cardiac amyloidosis (CA) based on the convolutional neural network (CNN) learning the cardiac magnetic resonances (CMR) images. Three deep-learning approaches were tested: feature learning, performing training without any initializations, and fine-tuning of pretrained VGG16 model. The outcomes indicated that the trained CNN with LGE images had the highest performance with ROC AUC of 0.96, sensitivity of 94% and the specificity of 90 which further showed that it had high potential on clinical use[8].

The study by Pu et al. (2023) compared five models of radiomics-based machine learning used to detect myocardial fibrosis in patients with hypertrophic cardiomyopathy (HCM) based on cine cardiac magnetic resonance (CMR) image data. among five models, ICMR+R2 model presented the best quality, the model had AUC of 0.898, accuracy of 89.02 percent, sensitivity of 92.54 percent, and F1-score of 93.23 The authors have come to a conclusion that this method can be used in order to identify the patients with fibrosis efficiently[9].

In a recent analysis by Jacob et al. (2024), a deep-learning model was designed with the Cine CMR images as the target in identifying four conditions of a heart, including hypertrophic cardiomyopathy (HCM). Even though the research takes multiclass classification, specific to HCM measures are applicable to our binary research. The model attained an AUC of 0.908 in the detection of HCM, which is notable. The work presents a helpful point of reference against which other applications of deep learning to the classification of cardiomyopathy could be compared[10].

TransMed is a CNN-Transformer hybrid model proposed by Dai et al. (2021) and it aims to improve the classification performance in the multi-modal medical images. CNN and Transformer with low-level feature extraction methods and long-range dependence capture was also used, which showed the effectiveness of the hybrid architecture from low-level information with limited data in the analysis of medical images in achieving 10.1 percent and 1.9 percent accuracy margins over the baseline techniques[11].

3. Dataset Description

This study involved the cardiac magnetic resonance (CMR) dataset obtained at the Kaggle platform that was specially tailored towards the diagnosis of the hypertrophic cardiomyopathy (HCM) [12]. The data comprises 59, 145 grayscale images, which are separated into two folders, which are illustrated in Figure 2.

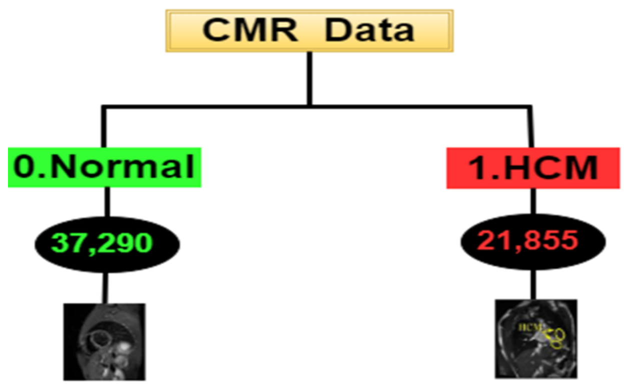


Figure 2: Distribution of CMR images data into two classes: normal and HCM.

The data has no special label files, instead the folder names are used directly as a source of label creation at the preprocessing stage. A major problem of medical imaging data collection, especially advanced methods, such as cardiac CMR, MRI and CT, is their cost, ethical, legal and resource logistics in the developing world.

3.1 Preprocessing

The preprocessing stage consisted of a few different steps to get the image data into a standard, but also effective, form to become compatible with deep learning models without losing the detail and balance of the classes. The procedures that were used were the following:

1. **Resizing:**
   All the images were changed in size to 224x224 pixels which was retained at its original aspect ratio. As required, black padding was inserted to limit distortion as well as to center it well.

2. **Normalization:**
   The pixel intensities were transformed so that they were limited to original grayscale range [0, 255] down to the range [-1.0, 1.0], as shown in this relationship:

$$\text{Normalized Value} = \frac{\text{Pixel Value}}{127.5} - 1.0$$

3. This normalization was useful towards stabilising the model and accelerating convergence in training.

4. **HDF5 Storage:**

   All the images and their labels were saved in HDF5 format, hierarchical, and efficient medium of storing large-scale datasets, to speed up the training process.

5. **Data Splitting:**
   The data was stratified into training and testing with the proportion being 80/20 and the classes percentages in the split matchedusing stratified split . All the experiments were made reproducible with a fixed random seed.

6. **Data Augmentation:**
   Going forward to enhance generalization and avoid overfitting, real-time data augmentation was used in training using a custom generator, and this consisted of:

   - Random horizontal flipping
   - Random rotations within ±10°
   - Translations up to ±10% of the image dimensions
   - Random zoom and cropping

The preprocessing procedures provided a better and diverse training data, which had a beneficial impact on the efficiency and generalizability of the deep learning.

3.2 Model Architecture Design

The proposed model adheres to a hybrid structure combining the Convolutional Neural Networks (CNNs) and Swin Transformer framework to use the benefits of both local and larger-scale features extraction in cardiac magnetic resonance

(CMR) images. The most basic components of the model, as shown in Figure 3-35, include the following:

- CNN Backbone (Feature Extraction):

  Extraction of low level visual features of the input images e.g. edges, textures/spatial patterns.

- Swin Transformer Block 1 – Local Window Attention:

  Uses self-attention on the data of non-overlapping local windows making it combine the information of localized areas of the image.

- Swin Transformer Block 2 – Shifted Window Attention:

  Rotates the windows to allow cross-window connections, which improve the model capacity to observe the global structure and long term dependencies.

- Classification Head:

  - Global Average Pooling (GAP): decrease feature maps to a constant-size vector.
  - Dense Layer: A learner of nonlinear feature combinations.
  - Dropout Layer: It prohibit overfitting, by randomly turning off neurons in the process of training.
  - Softmax Layer: product the probabilities for binary classification (healthy vs. sick).

This architecture is an effective integration of spatial sensitivities of CNNs and contextual strength of Swin Transformer and is effective in classification of binary cardiomyopathy.

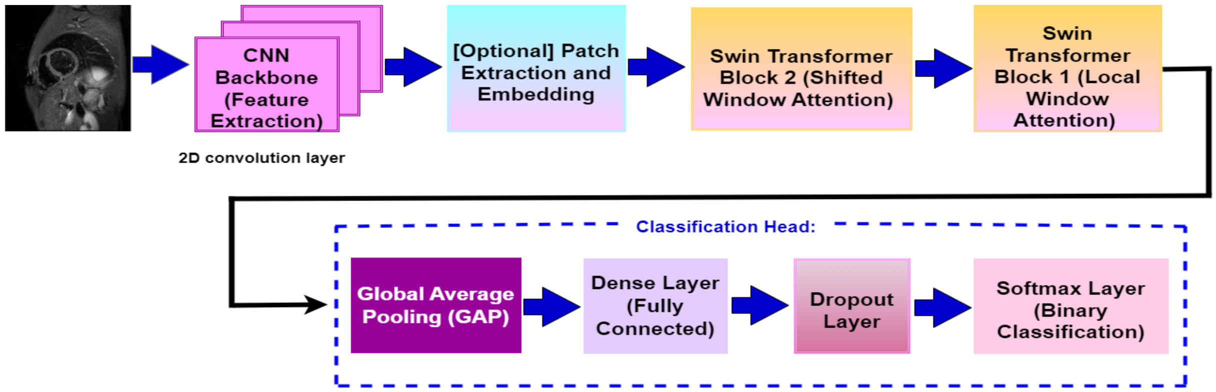


Figure 3: Hybrid Model Architecture Diagram.

3.3 CNN Backbones

In order to find out which convolutional feature extractor is more impotent in the classification of hypertrophic cardiomyopathy on the CMR images, four diverse CNN-based backbones were benchmarked: the Classic CNN, MobileNetV2, DenseNet121, and Xception, which each of them provided a different architectural benefit in a hybrid design modality.

1. **Classic CNN:**

This study developed a special architecture which was designed. It is constituted of sequence of 2D convolutional blocks, sequence of batch normalization and Leaky ReLU activations with max pooling operations. To support gradient flow and maintain spatial information in higher levels identity blocks with skip connections were used. Figure 4 illustrates the structural design of this classic CNN that comes with skip-connected identity blocks.

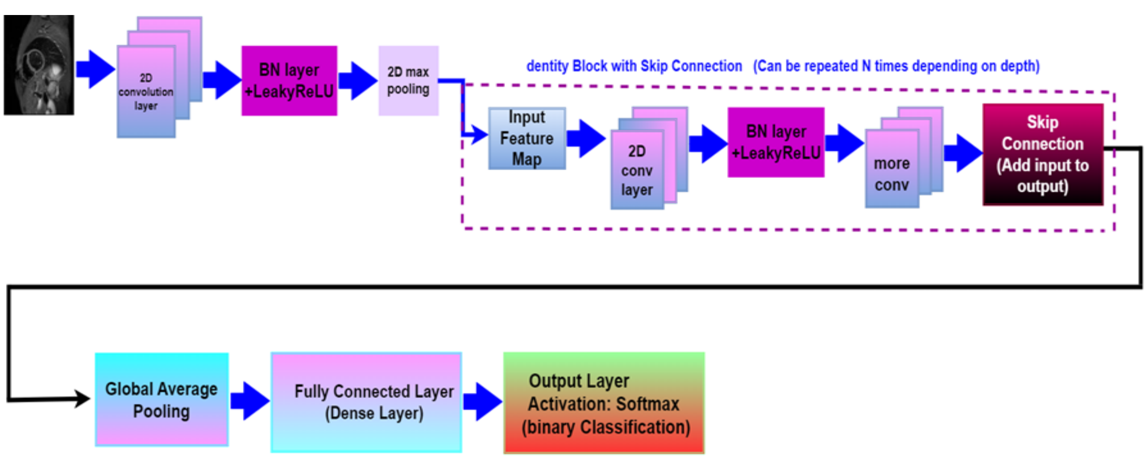


Figure 4: network contains Identity Blocks and Skip Connections to improve the performance of its classic CNN architecture for the classification of CMR images.

2. **MobileNetV2:**

It is a lightweight architecture that incorporates inverted residuals and linear bottlenecks, which is quite appropriate to deploy in resource-limited cases. The one involved in this work produced on pretrained Keras weights and comprised global average pooling and dense classification head.

3. **DenseNet121:**

This model allows the reuse of its features and it resolves the problem of the vanishing gradient due to its dense connected layers. It is based on the one taken out of the Keras library and optimized to support binary classification to enable the transfer of features via layers.

4. **Xception:**

Relaying on depthwise separable convolutions, Xception has fewer trainable parameters, but high accuracy. It was picked because of its high representational power and harmonization to hybrid structure as an up-to-date alternative to conventional CNNs.

All these CNN backbones were subsequently incorporated with Swin Transformer modules and a classification head to become the entire hybrid architecture applied in the current paper.

3.4 Swin Transformer – Traditional Version

The standard Swin Transformer implementation was absorbed into the selective hybrid model variations in this work, as it shows an advantage in terms of capturing local patterns and long-range contextual aspects in medical images.

3.4.1 Key Components

1. **Patch Extraction & Embedding:**

   The original image is split into the same sized patches. The patches are then run through an embedding layer that could transform it into a numerical vector representation[13].

2. **Position Embedding:**

   Each patch gets additional positional encodings made to maintain the spatial order in the image.

3. **Swin Transformer Block 1 – Local Window Attention:**

Print ISSN 2710-0952 Electronic ISSN 2790-1254

The self-attention operates in each of the non-overlapping windows and the local context can be understood[14].

4. **Swin Transformer Block 2 – Shifted Window Attention:**

In the next block, the windows are moved a little to enable exchange of information between the windows to facilitate global context modeling[15].

This strategy offers an efficient tradeoff that can be identified as local detail extraction and global awareness, and thus it becomes an ideal combination to apply high-resolution medical image classification, including cardiomyopathy detection within CMR scans.

3.5 Swin Transformer – Simplified Version (No Windows / No Embedding)

A simplification variant or variant of Swin Transformer was also discussed in this study. In contrast to the classical Swin, none of the components of the window partitioning/patch embedding/position embedding is used and touches upon a more computationally efficient and general design[16].

3.5.1 Reason to Simplify

The purpose of this version is to minimize the level of architectural complexity and evading the limitation brought about by image patching, but remain the strength of self-attention to exploit global image context.

3.5.2 Key Characteristics of the Simplified Version:

1. **Full-image processing:**

   Instead of partitioning the image to fixed-size windows, self-attention blocks are directly consuming the feature maps produced by the CNN backbone. This makes the model to process through the complete context of the image with no segment boundaries.

2. **No Patch Embedding Layer:**

   The model does not transform images to the patch tokens. It does this by using the raw CNN feature maps as inputs but this has the advantage of fewer parameters and flexibility[17].

3. **No Position Embedding**

   They do not perform spatial encoding, but are instead based on the spatial structure naturally inherent in the feature maps of CNNs, that are naturally spatially encoded[18].

Print ISSN 2710-0952 Electronic ISSN 2790-1254

Such reduced architecture saves on computation expense and matches well comparative in performance to CNNs. It is suitable especially in the medical field where speed and efficiency is needed, like in the image classification of cardiomyopathy.

3.6 Training & Experimental Setup

All the values of the main hyperparameters that were used to train all types of CNN and hybrids are shown in Table 1 as a consistency point of comparison in the experiments.

Table 1: summarizing the hyperparameters used in training.

Hyperparameter	Value	Description
Batch Size	32	Number of samples per training batch
Learning Rate	0.001	Initial step size for the Adam optimizer
Number of Epochs	40	Total passes over the entire training dataset
Dropout Rate	0.3	Dropout probability applied in classification head
Optimizer	Adam	Adaptive moment estimation optimizer
Loss Function	Categorical Cross-Entropy	Loss function for two-class softmax classification
Data Split Ratio	80% Train / 20% Test	Stratified sampling to preserve class balance
Early Stopping	Patience = 5	Stops training if no validation loss improvement for 5 epochs

To be able to compare architectures and achieve reproducibility in training and testing results, all the settings were standardized when configuring the training and evaluation environment.

3.7 Evaluation Metrics

Six evaluation metrics were used to thoroughly gauge the model performance in classification of images of cardiomyopathy. All these measures indicate a level of accuracy in making a prediction, the discrimination between the classes, and the strength of the model to process medical data[19].

Accuracy: $ACC = \frac{TP + TN}{TP + FN + FP + TN}$ (1)

Precision: $Prec. = \frac{TP}{TP + FP}$ (2)

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

F1 Score: $F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recal}}$ (4)

Area under the ROC curve, calculated as:

$$ROC\ AUC = \int_0^1 TPR\ (FPR) \quad (5)$$

Where: $TPR = \frac{TP}{TP+FN}$, $FPR = \frac{FP}{FP+TN}$

Matthews Correlation Coefficient:

$$MCC = \frac{TN \cdot TP - FN \cdot FP}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (6)$$

Confusion Matrix:

The overall layout of a binary classification confusion matrix giving relation of the actual and predicted labels (true/false positives and negatives)is given in figure 5[20].

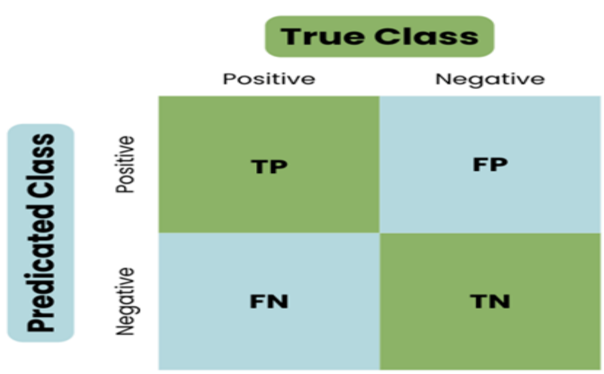


Figure 5: The Basic Structure of a Confusion Matrix.

4. Results and analysis

The chapter finds an in-depth analysis of the findings achieved with the proposed models, applying a complete range of assessment measures widely used in the medical imagery classification. The performance of each model is measured with regards to accuracy, precision, the recall, F1 score, AUC, and Matthews Correlation Coefficient (MCC). Training-validation plots and confusion matrices are given too to provide a visual aid to the evaluation. The last stage is to compare the performance of the best working models to identify the conclusions of how well each architecture can classify cardiomyopathy.

4.1 CNN Model (Original &Hybrid)

The CNN model is an from scratch , and its basic version provided background to the current study. It returned great results, it showed an accuracy of 93.60%, precision of 94.11%, recall of 92.32% and an F1 Score of 93.01. It also recognized a high Area Under Curve (AUC) of 98.90 % and Matthews Correlation Coefficient (MCC) of 86.32 % which signifies a balanced and strong classifier in any of the classes.

The results were worsened when the standard Swin Transformer was attached to the CNN. The accuracy of the model reduced to 64.40%, and F1 Score reduced to 48.11%, and MCC is very low 12.31%. This drop indicates that there is an architectural mismatch between CNN backbone and complete Swin attention. But applying Swin Transformer simplified version (without patching and embedding), the performance of the model recovered. It obtained 91.91 accuracy, 91.52 F1 Score, an AUC of 98.30 and an MCC of 83.20 per cent accuracies, as indicated on Table 2.

Table 2: The analysis results of the proposed CNN model.

Model	Accuracy	Pre.	Recall	F1 Score	AUC	MCC
✔ **CNN**	**93.60**	**94.11**	**92.32**	**93.01**	**98.90**	**86.32**
CNN + Swin	64.40	61.50	53.32	48.11	57.50	12.31
CNN + Swin (no embedding /patch)	91.91	90.90	92.32	91.52	98.30	83.20

Comparison among the three variants based on CNN indicated that the original model performed better than the other two modelsAs illustrated in Figure 6,

Print ISSN 2710-0952   Electronic ISSN 2790-1254

confusion matrix reveals that the model has very good capacity to differentiate normal and HCM cases with a few misclassifications. Additional evidence of the steady learning behavior and a continuous decrease in loss over time is provided by the training and validation curves (Figures 7).

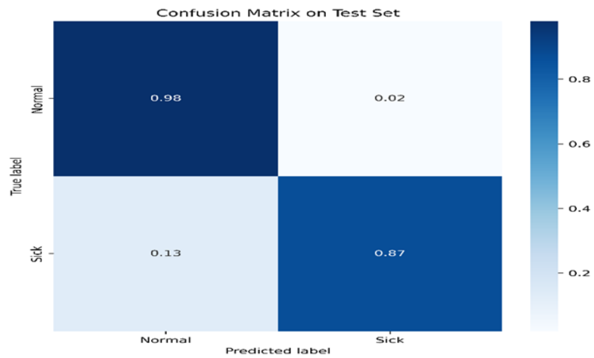
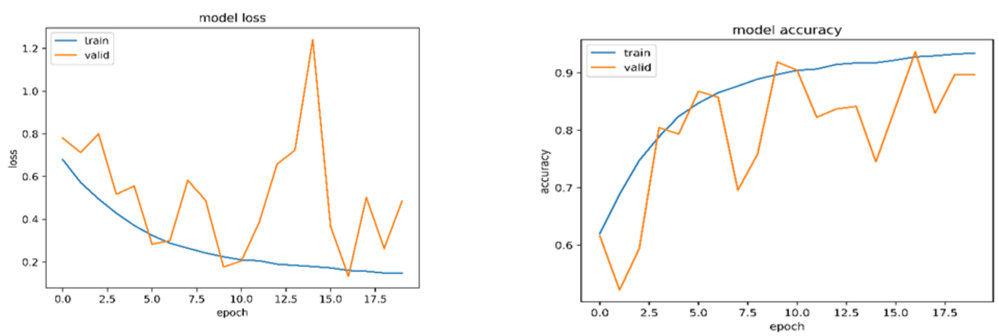


Figure 6: confusion matrix for CNN model.



(a)   (b)

Figure 7: (a) Training and validation accuracy and (b) loss curves for CNN model.

4.2 DenseNet121 Model (Original &Hybrid)

DenseNet121, due to dense connections of its layers and feature re-usage, have performed fairly well in their original variants, reporting 87.51accuracy, and F1 Score of 86.20. The AUC of the model was 94.50% showing that it was also a trusted architecture in extracting fine-grained features in the heart. Introducing the classic Swin Transformer into DenseNet121 had a rather demonstrable effect positively. The accuracy increased to 91.20 percent, the values of the remaining performance measures also increased, as indicated in Table 3.

Surprisingly, the strongest were the results of the DenseNet121 + Swin (no embedding/patch) version. In this hybrid, the accuracy was 96.02 percent and F1 Score was 95.70 percent. This model also reported AUC of 99.22% that ranks among the high-performance models in the whole experiment. As Figure 8 shows (the confusion matrix), it was correctly classified precisely well with the least error, and the training plot (Figures 9) shows steady progress with little overfitting.

Table 3: The analysis results of the proposed DenseNet121 model.

Model	Accuracy	Precision	Recall	F1 Score	AUC	MCC
DenseNet121	87.51	87.70	85.22	86.20	94.50	72.91
DenseNet121 +Swin	91.20	92.21	88.91	90.22	97.90	81.12
✔ DenseNet121+Swin (no embedding/patch)	**96.02**	**95.90**	**95.51**	**95.70**	**99.22**	**91.50**

The integration and the ability to increase performance with hybridization were high in DenseNet121. Although the initial variant has been strong, the classical Swin helped improve the performance, and simplified Swin version provided excellent outcomes. This implies that designed with architectural simplicity in mind, dense connections in DenseNet121 can complement attention.

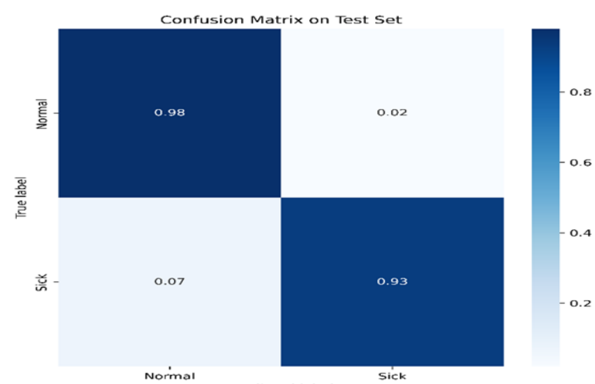


Figure 8: confusion matrix for DenseNet121+ Swin (No Embedding / Patch Splitting) model.

(a) (b)

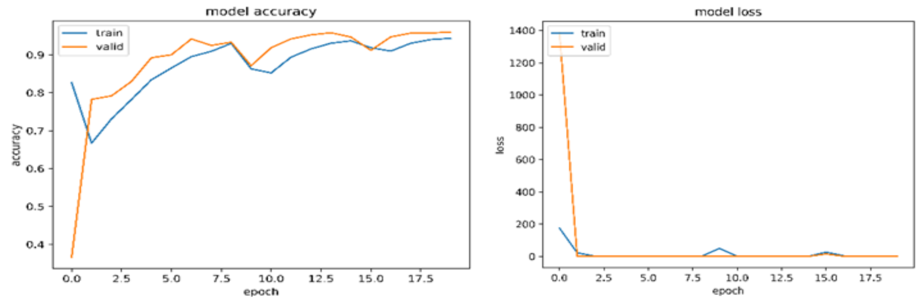


Figure 9: (a) Training and validation accuracy and (b) loss curves for DenseNet121+ Swin (No Embedding / Patch Splitting)model.

4.3 MobileNetV2 Model (Original & Hybrid)

Lightweight MobileNetV2 with an efficient architecture demonstrated a rather low level of performance with its standalone implementation. It had an accuracy of 63.41%, . The model had 1.22% Matthews Correlation Coefficient (MCC), which was hardly decent (poor balance and less discriminating power between the classes). Combining MobileNetV2 and complete Swin Transformer generated a significant increase in performance. The hybrid model performed at 92.90 % accuracy and the MCC is at 86.30%. Such combination proved that Swin Transformer attention mechanism can be useful supplement to the feature extraction of MobileNetV2 even though being lightweight.

The improvement became even greater when the simplified version of Swin Transformer, namely the one which does not include an embedding or patch procedure, was used. This version of the model was the highest performer of the models based on MobileNetV2, with accuracy of 96.22%,the AUC of 99.23%, table 4 depicts this. ,documenting great balance and durability in the model with regards to classification. Figure 10, shows the confusion matrix of MobileNetV2 + Swin (no embedding / patch) which depicts few misclassification errors. On training and validation accuracy and loss curves (Figure 11) there was a stable convergence and improving the performance was consistent throughout the epochs.

Table 4: The analysis results of the proposed MobileNetV2 model.

Model	Accuracy	Precision	Recall	F1 Score	AUC	MCC
MobileNetV2	63.41	81.60	50.02	38.80	61.01	01.22
MobileNetV2 +Swin	92.90	93.41	92.91	93.12	98.20	86.30

Print ISSN 2710-0952 Electronic ISSN 2790-1254

✔ **MobileNetV2 +Swin (no embedding/patch)**	**96.22**	**96.51**	**95.40**	**95.90**	99.23	**91.80**

Compared with the original standalone version of the MobileNetV2, the hybrid variants outperformed it by far, which confirmed the importance of attention mechanism of the Swin Transformer in improving feature representation. The best overall performance was obtained with the Swin which lacked the embedding and patching, meaning there is, possibly, greater synergies between the base CNN and the transformer block to achieve performance in this scenario.

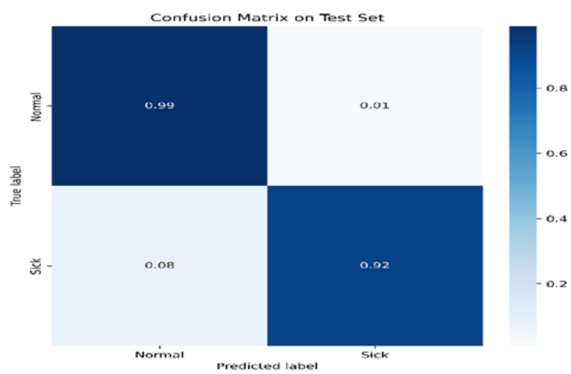


Figure 10: confusion matrix for MobileNetV2+ Swin (no embedding/patch) model.

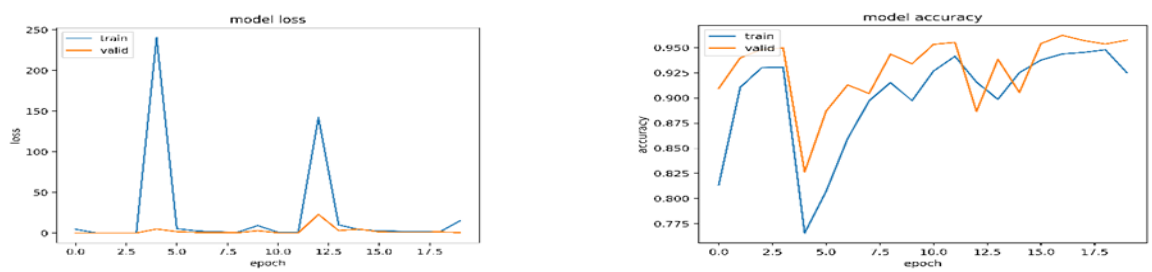


Figure 11: (a) Training and validation accuracy and (b) loss curves for MobileNetV2+ Swin (no embedding/patch) model.

(a) (b)

4.4
Xception Model (Original & Hybrid)

Xception model is noteworthy of its depthwise separable convolutions and powerful feature extraction, and thus showed good baseline performances. The single Xception model had the accuracy of 93.90%. It also showed high AUC of 98.90% which implies that it is a very reliable and balanced classifier. However, when used together with the full Swin Transformer, its performance dragged. The hybrid had an accuracy rate of 66.30 with the MCC lower at 19.9 %. The drop

Print ISSN 2710-0952 Electronic ISSN 2790-1254

shows the architectural incompatibility or inefficiencies of using this configuration of the full Swin attention mechanism and Xception backbone.

Applying the simplified version of the Swin Transformer, without using embedding or patching led to better performance as compared to the original version of Swin. The model had an 80.60 percent accuracy, 84.90 percent precision, and the AUC increased to 91.80 percent, as shown in Table 5. Whereas this is still not good enough compared to the standalone Xception model. Xception-based In the variants using Xception, the original standalone model is still the best performing configuration, and this alone demonstrates the power of the convolutional design in this architecture alone is very good.

Table 5: The analysis results of the proposed Xception model.

Model	Accuracy	Precision	Recall	F1 Score	AUC	MCC
✔ **Xception**	**93.90**	**94.51**	**93.42**	**93.91**	**98.90**	**87.92**
Xception + Swin	66.30	65.02	56.61	54.12	62.30	19.90
Xception + Swin (no embedding/patch)	80.60	84.90	74.41	76.21	91.80	58.30

4.5 Models performance comparison

This part gives a comparative review of the model whose performance was the best within each of the categories of architectural designs that are used during the study. The algorithm comparison is on the six main evaluation metrics which are Accuracy, Precision, Recall, F1 Score, AUC, and Matthews Correlation Coefficient (MCC) parameters. When applying a simplified Swin Transformer (without patch embeddings), the hybrid model that uses MobileNetV2 as the backbone had the best scores according to all the metrics, as it can be seen in the table 6 below. This means that it is superior to the other settings and it is the best model of binary classification of cardiomyopathy in this study.

Table 6: Comparison results of the proposed models.

Model	Accuracy	Precision	Recall	F1 Score	AUC	MCC
CNN	93.60	94.11	92.32	93.01	98.90	86.32
DenseNet121+Swin (no embedding/patch)	96.02	95.90	95.51	95.70	99.22	91.50
✔ **MobileNetV2 +Swin (no embedding/patch)**	**96.22**	**96.51**	**95.40**	**95.90**	**99.23**	**91.80**
Xception	93.90	94.51	93.42	93.91	98.90	87.92

4.6 Analysis and comparison with other work

It is a comparison of the proposed models with a few related studies in the area of cardiac disease classification with CMR and similar medical imaging data in the past. Some of the recent articles that applied both conventional and hybrid deep learning models have been provided as a reference. A more detailed model comparison, over architecture, datasets and performance indicators is shown in Table 7. This discussion will aid in pointing out the bright and weaknesses of the suggested models, especially the simplified hybrid architecture comprising MobileNetV2 and Swin Transformer.

Table 7: The comparison between the proposed classification models and previous works.

Ref.	Year	Classifier	Dataset	Accuracy	Precision	Recall	F1 Score	AUC	MCC
[7]	2024	CNN (CMR)	CMR images (Hypertrophic Cardiomyopathy Dataset (Omid Hospital, Iran))	94.710	96.971	91.210	94.850	/	/
[8]	2021	Fine-uned VGG16 CNN	CMR images (LGE protocol) – Cardiac Amyloidosis	/	/	94.0	/	0.96	/
[9]	2023	ICMR + R2	Cine CMR (HCM patients, fibrosis detection)	89.02	/	92.54	93.23	0.898	/
[10]	20	VAE + Deep	Cine	77.8	/	/	/	0.90	/

2025 حزيران العـــدد 17S المجلة العراقية للبحوث الإنسانية والإجتماعية والعلمية
No.17S JUNE 2025 Iraqi Journal of Humanitarian, Social and Scientific Research
Print ISSN 2710-0952 Electronic ISSN 2790-1254

	24	CNN	CMR (Omid Hospital, USA)					8 (HCM)	
[11]	2021	TransMed (CNN + Transformer)							
		├─TransMed-S	PGT	88.9	88.3	/	/	/	/
		├─TransMed-S	MRNet (ACL Tear)	94.9	/	/	/	/	/
		├─TransMed-S	MRNet (Abnormality)	91.8	/	/	/	/	/
		├─TransMed-B	MRNet (Meniscus Tear)	85.3	/	/	/	/	/
Current Authors	2025	Proposed Model							
		CNN	CMR	93.6	94.1	92.2	93.0	98.9	86.3
		Xception	CMR	93.9	94.5	93.4	93.9	98.9	87.9
		DenseNet121 + Swin (no embed)	CMR	96.0	95.9	95.5	95.7	99.2	91.5
		MobileNetV2 + Swin (no embed)	**CMR**	**96.2**	**96.5**	**95.4**	**95.9**	**99.2**	**91.8**

5. Conclusions

Following the results of the analysis period during the evaluation stage, the models that performed better and were used during all evaluation metrics are hybrid models with only the combination of MobileNetV2 with simplified Swin Transformer (which does not use patch embedding method) providing superior performance on all evaluation metrics. Those configurations managed to take effective advantage of the spatial sensitivity of CNNs and of the contextual power of attention mechanisms. The traditional models like CNN and Xception did not outperform the hybrid models; nevertheless, they were performing decently. These findings lay emphasis on the role of an architectural synergy in improving the medical image classification by providing potent and precise tools to help clinicians in an earlier diagnosis and in a better patient management.

References

[1] Q. Liu, Y. Chai, J. Meng, and C. Hu, "Convolutional neural networks to differentiate hypertrophic cardiomyopathy from hypertensive heart disease based on cardiac cine imaging", In: Proc. of International Society for Magnetic Resonance in Medicine ... Scientific Meeting and Exhibition , 2023, doi:10.58530/2022/4123.

[2] V. Sangha et al., "Identification of hypertrophic cardiomyopathy on electrocardiographic images with deep learning", medRxiv, 2023.

[3] H. Wu, Z. Huang, J. Liu, J. Dai, Y. Zhao, and W. Luo, "The predictive value of deep learning-based cardiac ultrasound flow imaging for hypertrophic cardiomyopathy complicating arrhythmias", *Eur. J. Med. Res.*, Vol. 28, No. 1, pp. 36, 2023.

[4] R. K. Hughes et al., "Improved diagnostic criteria for apical hypertrophic cardiomyopathy", *Cardiovasc. Imaging*, Vol. 17, No. 5, pp. 501-512, 2024.

[5] Yilmaz, A., Bauersachs, J., Bengel, F., Büchel, R., Kindermann, I., Klingel, K., ... & Frey, N. (2021). Diagnosis and treatment of cardiac amyloidosis: position statement of the German Cardiac Society (DGK). *Clinical Research in Cardiology*, *110*, 479-506.

[6] Das, S., Sultana, M., Bhattacharya, S., Sengupta, D., & De, D. (2023). XAI–reduct: accuracy preservation despite dimensionality reduction for heart disease classification using explainable AI. The Journal of Supercomputing, 79(16), 18167-18197.

[7] Kolluri, S., & Hathwar, S. (2024, July 29). Diagnosing hypertrophic cardiomyopathy using machine learning models on CMRs and EKGs of the heart. Journal of Emerging Investigators, 7(1). https://doi.org/10.59720/23-209.

[8] Agibetov, A., Kammerlander, A., Duca, F., Nitsche, C., Koschutnik, M., Donà, C., Dachs, T.-M., Rettl, R., Stria, A., Schrutka, L., Binder, C., Kastner, J., Agis, H., Kain, R., Auer-Grumbach, M., Samwald, M., Hengstenberg, C., Dorffner, G., Mascherbauer, J., & Bonderman, D. (2021). Convolutional neural networks for fully automated diagnosis of cardiac amyloidosis by cardiac magnetic resonance imaging. Journal of Personalized Medicine, 11(12), 1268. https://doi.org/10.3390/jpm11121268.

[9] Pu, C., Hu, X., Lv, S., Wu, Y., Yu, F., Zhu, W., Zhang, L., Fei, J., He, C., Ling, X., Wang, F., & Hu, H. (2023). Identification of fibrosis in hypertrophic cardiomyopathy: A radiomic study on cardiac magnetic resonance cine imaging. European Radiology, 33, 2301–2311. https://doi.org/10.1007/s00330-022-09217-0.

[10] Jacob, A. J., Chitiboi, T., Schoepf, U. J., Sharma, P., Aldinger, J., Baker, C., Lautenschlager, C., & others. (2024). Deep-learning-based disease classification in patients undergoing cine cardiac MRI. Journal of Magnetic Resonance Imaging. https://doi.org/10.1002/jmri.30071.

[11] Dai, Y., Gao, Y., & Liu, F. (2021). TransMed: Transformers advance multi-modal medical image classification. Diagnostics, 11(8), 1384. https://doi.org/10.3390/diagnostics11081384.

[12] D. Sharifrazi and R. Alizadeh, "Hypertrophic Cardiomyopathy Dataset," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/danialsharifrazi/hypertrophic-cardiomyopathy-datase.

[13] R. A. Dihin, W. A. M. Al-Jawher, and E. N. AlShemmary, "Diabetic retinopathy image classification using shift window transformer," Baghdad Sci. J., vol. 21, no. 8, pp. 2741-2756, 2024.

[14] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in Proc. IEEE/CVF Int. Conf. Comput. Vis., pp. 10012-10022, 2021.

[15] Y. Sun, L. Ning, B. Zhao, and J. Yan, "Tomato Leaf Disease Classification by Combining EfficientNetv2 and a Swin Transformer," Appl. Sci., vol. 14, no. 17, p. 7472, 2024.

[16] Z. Li, G. Xie, G. Jiang, and Z. Lu, "ShadowMaskFormer: Mask Augmented Patch Embeddings for Shadow Removal," arXiv preprint arXiv:2404.18433, 2024.

[17] M. Kumar, U. Mehta, and G. Cirrincione, "Enhancing neural network classification using fractional-order activation functions," AI Open, vol. 5, pp. 10-22, 2024.

[18] E. U. Henry, O. Emebob, and C. A. Omonhinmin, "Vision transformers in medical imaging: A review," arXiv preprint arXiv:2211.10043, 2022.

المجلة العراقية للبحوث الإنسانية والإجتماعية والعلمية
Iraqi Journal of Humanitarian, Social and Scientific Research
Print ISSN 2710-0952 Electronic ISSN 2790-1254

[19] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," BioData Mining, vol. 16, no. 1, p. 4, 2023.

[20] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," BMC Genomics, vol. 21, pp. 1-3, 2020.