

An intelligent ensemble EfficientNet prediction system for interpretations of cardiac magnetic resonance images in heart failure severity diagnosis

Muthunayagam Muthulakshmi^a, Koteswaran Venkatesan^b, Balaji Prasanalakshmi^c, Rahayu Syarifah Bahiyah^{d,*}, Vijayakumar Divya^a

^a Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Chennai, 601103, India

^b Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai, 601103, India

^c Department of Computer Science, King Khalid University, Abha, Saudi Arabia

^d Faculty of Defense Science and Technology, National Defence University Malaysia, Kuala Lumpur, Malaysia

ARTICLE INFO

Keywords:

Ensemble prediction
Efficientnet
Heart failure severity
Cardiac magnetic resonance images
Deep features
Federated learning

ABSTRACT

Ensemble models as part of federated learning leverage the ability of individual models to learn unique patterns from the training dataset to make more efficient predictions than single predicting systems. This study aggregates the output of four best-performing EfficientNet models to arrive at the final heart failure severity prediction through federated learning. The seven variants of EfficientNet models (B0-B7) learn the features from the cardiac magnetic resonance images that are most relevant to heart failure severity. Further, the performance of every model variant has been analysed with three different optimizers i.e. Adam, SGD, and RMSprop. It has been observed that the developed ensemble prediction system provides an improved overall testing accuracy of 0.95. It is also worthy to note that the ensemble prediction has yielded significant improvement in the prediction of individual classes which is evident from sensitivity measure of 0.95, 0.88, 1.00, 0.93, and 0.98 for hyperdynamic, mild, moderate, normal and severe classes respectively. It is obvious from these results that the proposed ensemble EfficientNet prediction system would assist the radiologist in better interpretation of cardiac magnetic resonance images. This in turn would benefit the cardiologist in understanding the HF progress and planning effective therapeutic intervention.

1. Introduction

A recent survey carried out by the World Health Organization states that 17.9 million fatalities worldwide are attributed to cardiovascular diseases (CVD) and this is expected to reach an alarming count of 23 million by 2030 [1]. Among CVDs, heart failure (HF) claims to be the most prevalent one and contributes to most of the health burden worldwide. Recently, HF owed to most of the young age and middle age mortality, which vastly disturbs the family's psychological stamina and increases the financial burden. The pathophysiology of HF is related to the structural or functional abnormality in the heart, which directly affects the cardiac output [2]. The external symptoms of HF are not effective enough to prove the appropriate stage of HF. This poses a great challenge in the prediction of survival rate and risk grading in patients with HF. HF can be categorized as "HF with reduced ejection fraction" (HREF), "HF with preserved ejection fraction" (HPEF), and "HF with increased ejection fraction" (HIEF). The earlier one results in impaired

ventricular contraction while the later one results in impaired ventricular dilation [3]. Though in previous decades much focus was given to HREF, recent analysis of HPEF and HIEF has gained more momentum as this constitutes a heterogeneous group. Coronary microvascular dysfunction may lead to heart failure with preserved ejection fraction. The pathophysiology underlying HIEF includes elevated diastolic stiffness and smaller size in LV [4]. However, cardiologists need intelligent assistance to predict the level of damage in cardiac muscles to plan appropriate therapeutic interventions and monitor cardiac conditions.

Clinically, HF can be diagnosed using multiple imaging approaches such as echocardiography, computed tomography (CT), positron emission tomography (PET), single photon emission computed tomography (SPECT), cardiac magnetic resonance imaging (CMR), etc. Among them, CMR is considered to be the gold standard non-invasive procedure for assessing cardiac functions, tissue characterization, and anatomy [5]. Despite all these uniqueness among other imaging modalities, usage of CMR on a large scale is hindered by the limited availability of CMR

* Corresponding author.

E-mail address: syarifahbahiyah@upnm.edu.my (R. Syarifah Bahiyah).

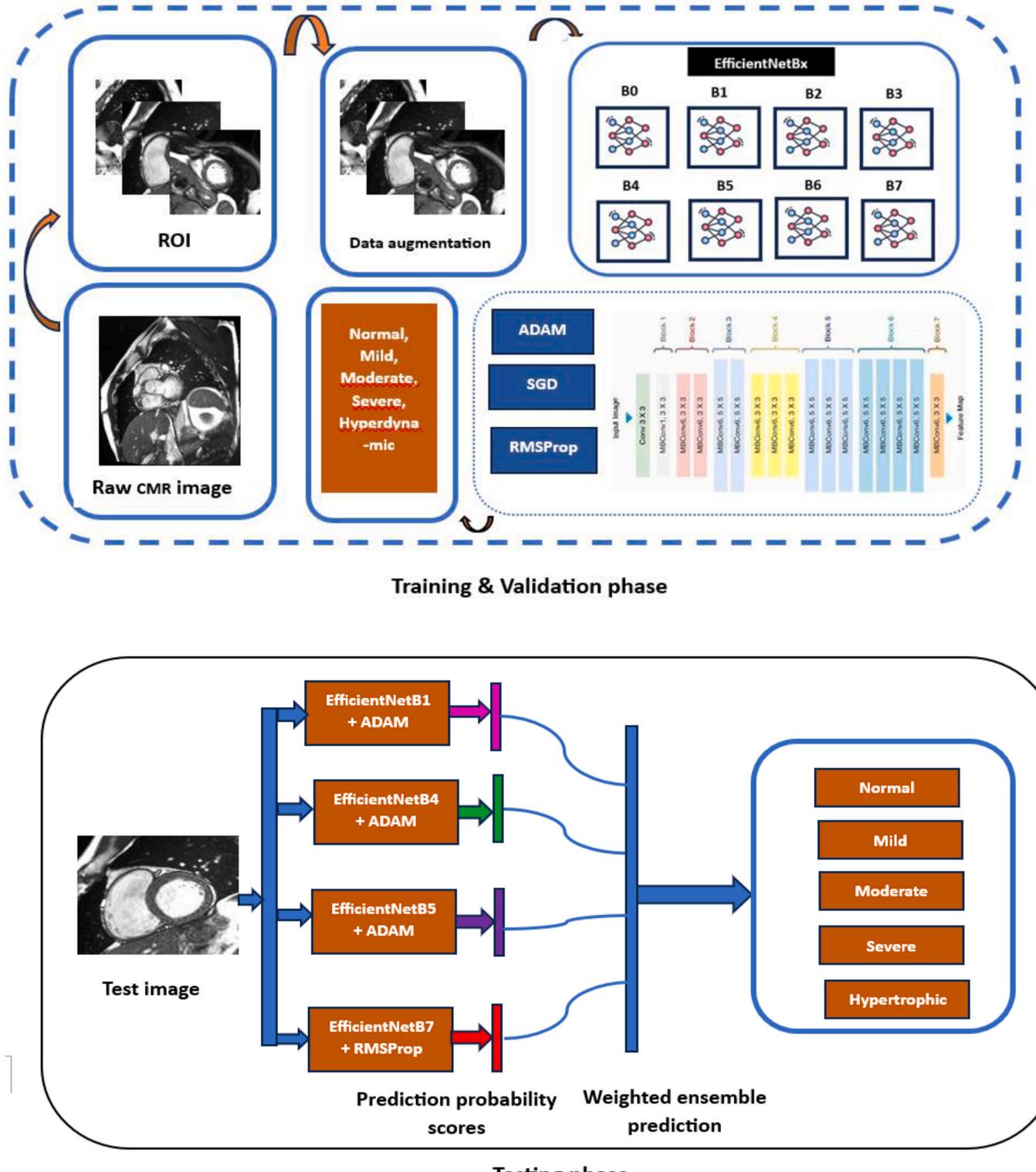


Fig. 1. Proposed workflow.

interpretation experts. Hao et al. study indicates the diagnostic significance of CMR in discriminating heart failure with preserved, mid-range, and reduced ejection fraction from healthy subjects [6]. In the initial studies, texture features, radiomic features have been extracted to predict the prognosis of HF using CMR [7]. The preliminary indicator of HF is ejection fraction which is dependent on left ventricle volume at end

diastolic and end systolic phase. Hence, there are studies on automated assessment of biventricular volumes and mass from CMR images [8]. Subsequently, the potential of artificial intelligence can be exploited for effective CMR interpretation, especially in the HF prognosis.

The ability of deep learning algorithms to identify and learn distinguishing features in medical images has paved the way for its

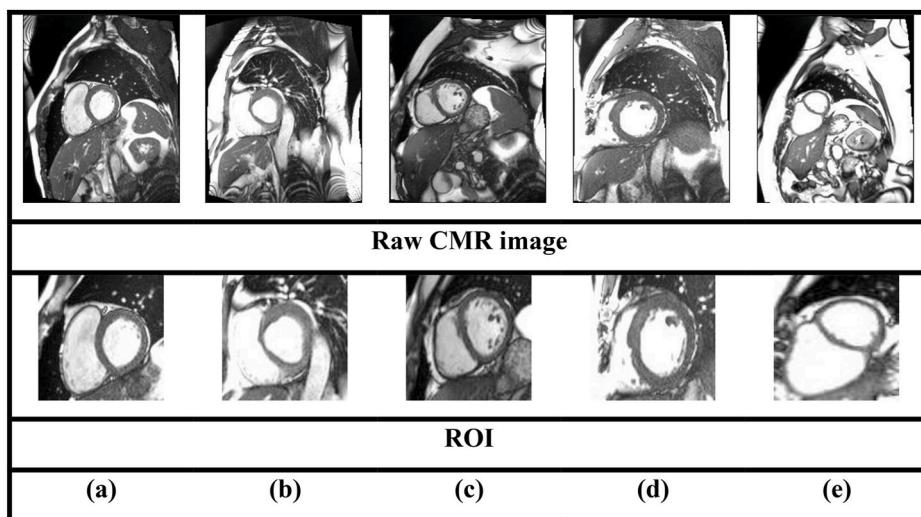


Fig. 2. Raw CMR image and ROI for different categories of subjects: (a) normal, (b) mild, (c) moderate, (d) severe and (e) hyperdynamic.

exhaustive usage in the development and enhancement of computer-assisted image-based diagnosis [9]. A study by Gao et al. considered heart failure patients with a reduced ejection fraction [10]. The denoising autoencoder model has been used to discriminate high-risk and low-risk groups with a concordance index of 0.8546. Wang et al. have worked with Resnet 32 for effective discrimination of hypertrophic cardiomyopathy and hypertensive heart disease using T1-weighted CMR images [11]. The usage of single machine learning or deep learning algorithms is sometimes considered to be biased, as their decisions are purely based on the patterns learnt by those models alone. Hence, there seems to be a strong need for a generalized model that would work effectively better on unseen data. In this fully digitalized world, intelligent computer-aided assistance for CMR interpretation will reduce the screening time greatly.

1.1. Related work

Petmezas et al. have conducted an exhaustive review of different deep-learning algorithms employed for heart failure diagnosis [12]. Zhang et al. study suggests that CMR features based on maximum 2D diameter slice, least axis length, minor axis length, median, voxel volume, sphericity, flatness, Kurtosis, and Skewness gave an accuracy of 91.2 % in discriminating hypertrophic cardiomyopathy and dilated cardiomyopathy patients from healthy ones using random forest classifier [13]. In another study, Xie et al. have attempted a framework that leverages the Siamese network and bidirectional LSTM to obtain spatial and temporal information from 4D CMR to improve the HF classification [14]. Atrial size measurement, myocardial deformation analysis, and strain rate measurement made from CMR sequences have proved to be clinically significant in HPEF clinical diagnosis [15].

You et al., fed features extracted from two CNN structures as input to recurrent network to understand the temporal deformation in myocardium as an investigation on hypertrophic obstructive cardiomyopathy [16]. In the evaluation of HIEF, 3D convolution-based ResNet has been used to extract features from short-axis view, 2-chamber, 3-chamber, and 4-chamber long-axis view [17]. Budai et al., achieved an F1-score and recall of 92 % and 97 % respectively in HIEF detection. Jiang et al. have attempted to differentiate cardiac amyloidosis and hypertrophic cardiomyopathy using radiomics, conventional CMR metrics, and support vector machine [18]. The features extracted from CNN are further fed to a random forest classifier for assessment of coronary artery disease [19].

Recently different versions of EfficientNet have been utilized for intelligent image-based diagnosis framework development for various

diseases. Efficientnet is a CNN version with compound scaling capability that provides high accuracy with reduced computational and memory costs [20]. The depth, width, and resolution of baseline EfficientNet-B0 architecture are scaled up in subsequent versions of EfficientNet to improve model performance proportionately by striking an effective trade-off between computational efficiency and accuracy. The baseline architecture includes 2D depthwise convolutional blocks, that consists of mobile inverted bottleneck convolution layers with squeeze and excitation optimization. Here, the optimal depth, width, and resolution are discovered automatically. Hence, the aim of this study is to provide an intelligent EfficientNet model based comprehensive assistance to cardiologist for accurate HF severity grading and prognosis. This would help us in usage of federated learning from raw CMR image acquired at different scan centres.

The major contributions of the proposed framework are as follows:

- **EfficientNetB0-B7 models:** EfficientNetB0-B7 models have been studied for the prediction of HF severity stages (mild, moderate, severe, hypertrophic)
- **Evaluation of performance of three different optimizers (SGD, RMSProp, Adam):** All the considered EfficientNet versions have been trained using optimizers such as SGD, RMSProp and Adam. The sensitivity and specificity obtained for each class using each learning algorithm have been analysed. The best-performing EfficientNetBx + Optimizer pair has been chosen based on individual class sensitivity.
- **Ensemble Prediction:** The weighted aggregation of probability scores of best performing EfficientNetBx + Optimizer pair has been considered for final prediction.

Sections 2, 3, 4, and 5 elaborately explain the materials and methods, results, discussions & future scope and conclusion respectively.

2. Materials and methods

The process flow adopted in the proposed work is shown in Fig. 1. This study utilizes the T1-weighted CMR images of people with normal, reduced (mild, moderate, severe) and hyperdynamic ejection fractions. The ROI region comprising of right ventricle cavity, left ventricle cavity and myocardium is delineated from the raw CMR cine-sequence. The ROI image is fed as input to different versions of the EfficientNet model, i.e. EfficientNetB0-B7. All the models have been trained with three different optimizers namely, stochastic gradient descent (SGD), RMSprop and Adam in the training and validation phase. The best

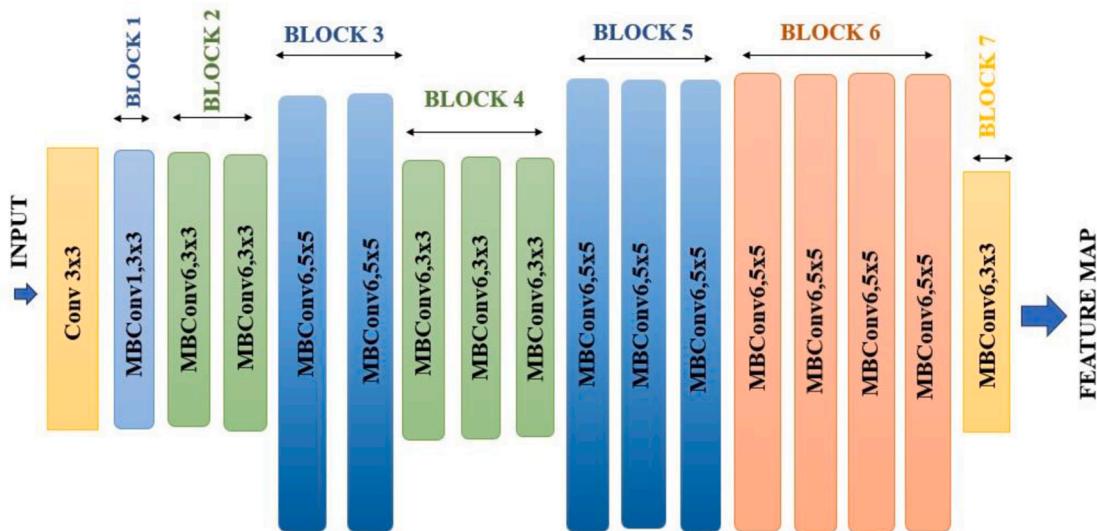


Fig. 3. EfficientNet base architecture.

performing EfficientNetBx + Optimizer combination is selected among the twenty-four combinations of EfficientNetBx + Optimizer in the testing phase based on the individual class sensitivity. The ensemble prediction is carried out by considering the weighted aggregation of probability scores obtained from the models having the best individual class sensitivity.

2.1. Dataset

The proposed study employs a dataset comprising of 2141, T1 weighted short-axis view CMR cine-sequences of 890 subjects, consolidated from Kaggle's challenge database (<https://www.kaggle.com/c/second-annual-data-science-bowl>). The CMR images considered in the current analysis pertains to healthy subjects and people with various HF severity levels (mild, moderate, severe, and hyperdynamic). Here, mild, moderate, and severe stages correspond to HREF and hyperdynamic corresponds to HIEF. The dataset involves both the gender and it consists of people in the age group of 20 years–60 years. Here, 9 to 23 CMR frames have been scanned to cover entire longitudinal length of heart. However, our current evaluation framework has been limited to mid-cavity frames of people imaged at end diastole phase of the cardiac cycle. Other metadata information about the CMR is as follows: slice thickness - 6 to 8 mm, pixel spacing - 0.6490–1.7969 mm/pixel.

2.2. Data pre-processing

The CMR frames can be categorized as basal, mid-cavity and apical slices based on the position across longitudinal length of the heart. This study is totally based on mid-cavity slices of each patient, that was acquired during end-diastole. The raw CMR images acquired include left ventricle, right ventricle, chest wall, stomach and liver.

However, for this study the region surrounding ventricle alone is of prominent interest. Hence, region of interest that accommodates left ventricle blood pool, myocardium and right ventricle is delineated as shown in Fig. 2. The HREF (mild, moderate and severe) patients show an impaired systolic function, while on contrary HIEF patients show a decline in diastolic function.

2.3. EfficientNet models

The proposed framework utilizes eight variants of EfficientNet model i.e. B0 to B7 [20]. The EfficientNet base architecture (EfficientNetB0) has been shown in Fig. 3. The further variants of EfficientNet are

designed through appropriate scaling of architecture in terms of width, depth and resolution of input as given in Fig. 4. The major building block of EfficientNet is the mobile inverted bottleneck convolution (MBConv) block, squeeze and excitation optimization. Each layer utilizes MBConv block of different kernel sizes. MBConv block performs depth wise separable convolution. Unlike traditional convolution layer, where single filter process all the input channels, in MBConv block there are separate filters for each channel. Then point wise convolution is performed to consolidate output across all channels. The kernel size in each layer is chosen considering trade-off between learning spatial information and computational cost. The squeeze and excitation layers in MBConv provide better feature learning through recalibration of channel wise responses.

All the EfficientNet model utilizes RELU activation function in the intermediate layers and Softmax activation function in the final layer. The L1 regularization parameter, L2 regularization parameter, learning rate, and dropout is chosen as 0.01, 0.01, 0.0001 and 0.4 respectively. The categorical loss function has been adopted for all the models. This study explores the performance of all the EfficientNet variants with three different optimizers such as Adam, RMSprop and SGD.

Let θ denote the model parameters, t the current iteration and $E(\theta)$ be a function dependant on model parameters. Here, let α be the learning rate, β_1, β_2 be the exponential decay rates for momentum estimates. SGD uses a fixed learning rate, while RMSprop and Adam uses an adaptive learning rate for each parameter. The equation for the three learning algorithms is given as follows:

SGD:

$$\theta_{t+1} = \theta_t - \alpha \cdot \nabla_{\theta_t} E(\theta) \quad (1)$$

RMSprop:

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{v_{t+1}}} \cdot \nabla_{\theta_t} E(\theta) \quad (2)$$

$$v_t = \nabla_{\theta_t} E(\theta)^2 \quad (3)$$

$$v_t = \beta_1 \cdot v_{t-1} + (1 - \beta_1) \nabla_{\theta_{t-1}} E(\theta)^2 \quad (4)$$

Adam:

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{\hat{v}_{t+1}} + \epsilon} \cdot \widehat{\nabla_{\theta_t} E(\theta)} \quad (5)$$

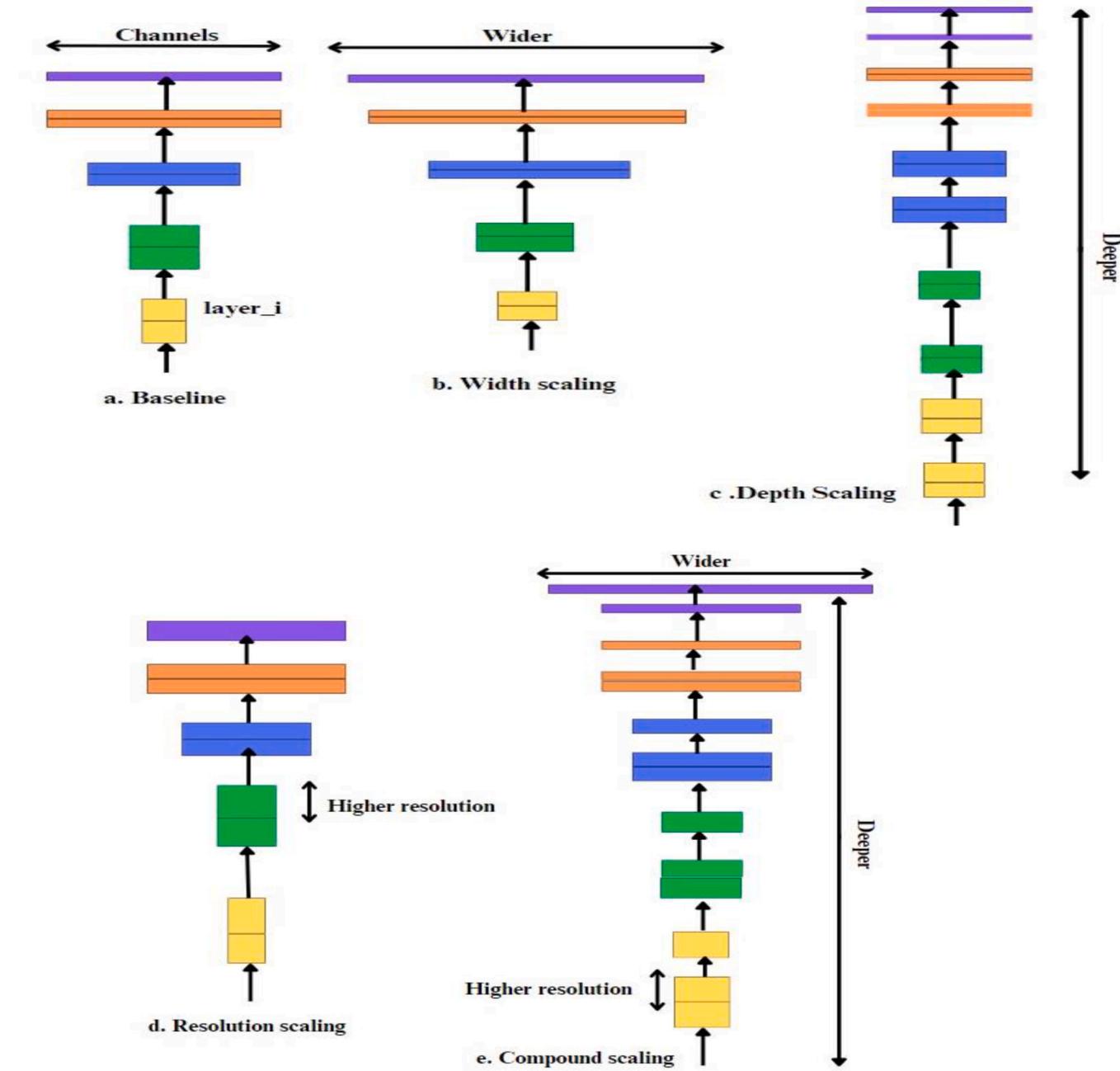


Fig. 4. EfficientNet architecture scaling.

$$\widehat{\nabla_{\theta_t} E(\theta)} = \frac{\nabla_{\theta_t} E(\theta)}{1 - \beta_1^{t+1}} \quad (6)$$

$$\widehat{v_{t+1}} = \frac{v_{t+1}}{1 - \beta_2^{t+1}} \quad (7)$$

2.4. Ensemble prediction

In the training, validation and testing phase, all 24 combinations of EfficientNetBx + optimizers are exhaustively explored. It is observed that some versions of EfficientNet are accurate in predicting particular categories of images better than all other versions. Hence, based on the sensitivity obtained for individual classes in testing phase, the prediction probability scores of a model are appropriately weighted. Let us assume that, Model₁, Model₂, Model₃, Model₄ and Model₅ are producing good

sensitivity for normal, mild, moderate, severe and hyperdynamic classes respectively. The output layer of the model will have 5 nodes, where each node pertains to one class. Hence, the model output will be a probability score vector of size 5. The probability scores produced by one model is given as an example below:

$$\text{Probability scores of Model}_1 = [\text{PS}_{11}, \text{PS}_{12}, \text{PS}_{13}, \text{PS}_{14}, \text{PS}_{15}] \quad (8)$$

Based on the sensitivity towards a class of images, the probability scores from individual model are weighted. For example, if Model₂ is performing better for mild classes. Hence the probability score PS₂₂ from Model₂ is given higher weightage in prediction of mild cases while performing the ensemble prediction. The weighted ensemble prediction process is given as below:

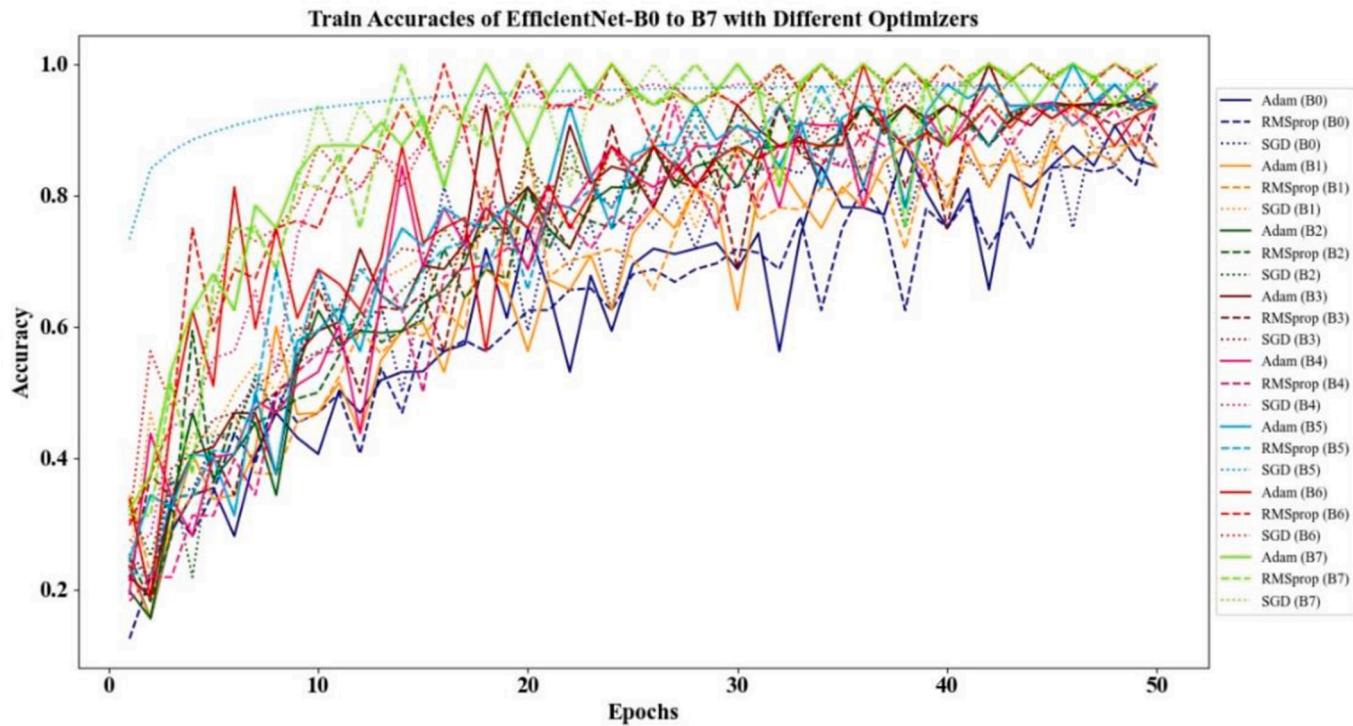


Fig. 5. Training accuracy vs Epochs for all models and optimizers.

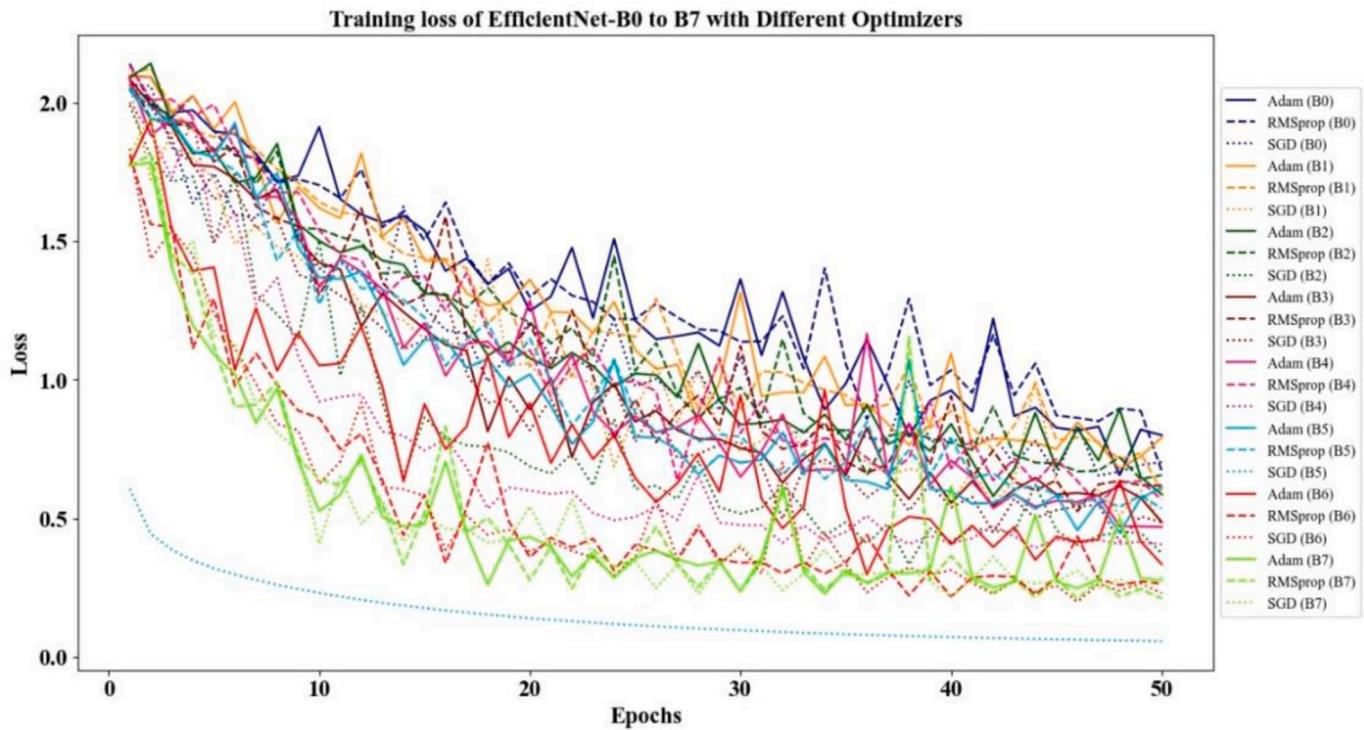


Fig. 6. Training loss vs Epochs for all models and optimizers.

2.4.1. Probability scores with ensemble prediction =

$$[a_{11} * PS_{11} + a_{21} * PS_{21} + a_{31} * PS_{31} + a_{41} * PS_{41} + a_{51} * PS_{51}, a_{12} * PS_{12} + a_{22} * PS_{22} + a_{32} * PS_{32} + a_{42} * PS_{42} + a_{52} * PS_{52}, a_{13} * PS_{13} + a_{23} * PS_{23} + a_{33} * PS_{33} + a_{43} * PS_{43} + a_{53} * PS_{53}, a_{14} * PS_{14} + a_{24} * PS_{24} + a_{34} * PS_{34} +$$

$$a_{44} * PS_{44} + a_{54} * PS_{54}, a_{15} * PS_{15} + a_{25} * PS_{25} + a_{35} * PS_{35} + a_{45} * PS_{45} + a_{55} * PS_{55}] \quad (9)$$

The values for the ensemble weight parameters $\{a_{11}, a_{21}, a_{31}, a_{41}, a_{51}\}$, $\{a_{12}, a_{22}, a_{32}, a_{42}, a_{52}\}$, $\{a_{13}, a_{23}, a_{33}, a_{43}, a_{53}\}$, $\{a_{14}, a_{24}, a_{34}, a_{44}, a_{54}\}$, $\{a_{15}, a_{25}, a_{35}, a_{45}, a_{55}\}$ are chosen in the range of {0,1} based on

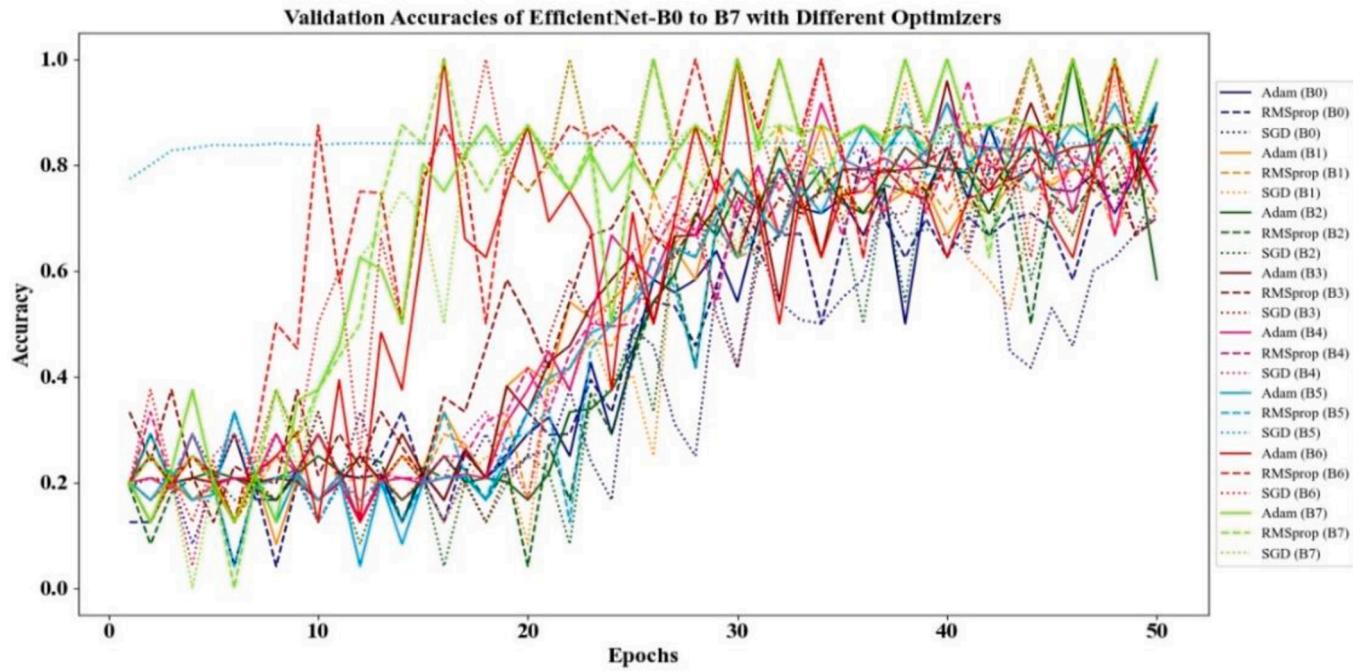


Fig. 7. Validation accuracy vs Epochs for all models and optimizers.

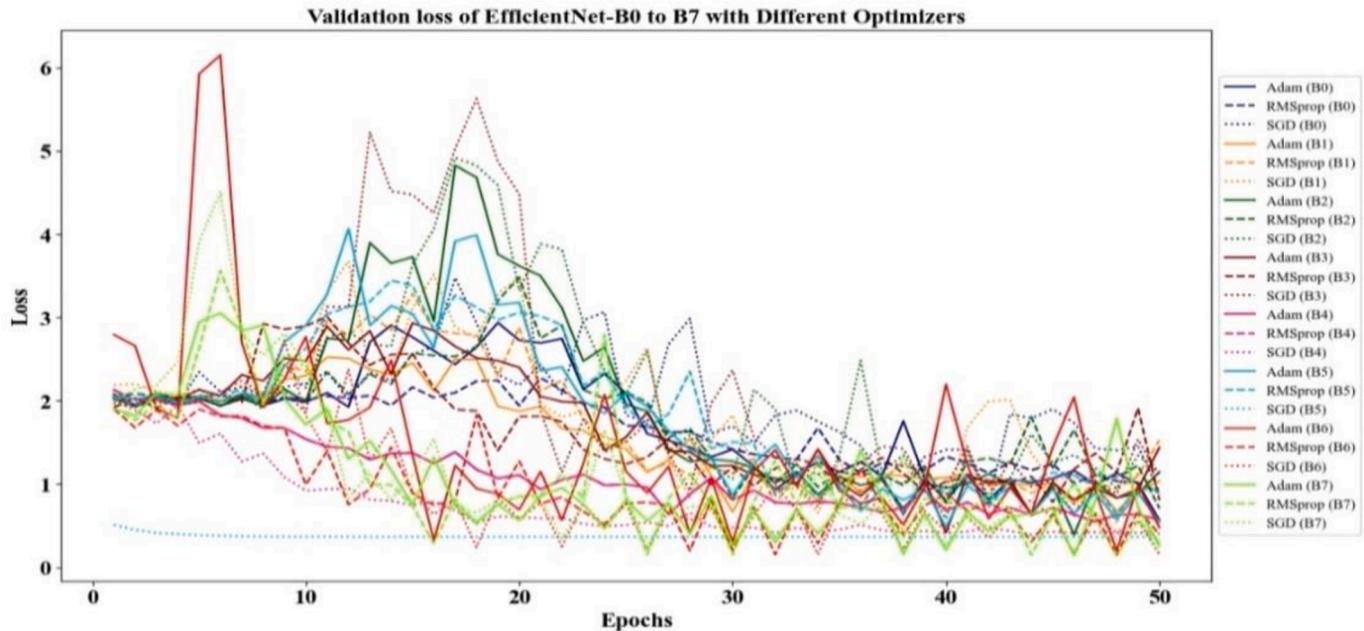


Fig. 8. Validation loss vs Epochs for all models and optimizers.

trial and error method after several trials. Finally, the HF severity stage of a patient is predicted by choosing the class of maximum probability score.

8. Results

The system used in this study has a 11th gen Intel i5-1135G7 processor operating at a speed of 2.40 GHz, with a RAM of 8 GB and 64-bit windows operating system. The codes were developed using Python 3.10.13 and Tensorflow 2.15.0 framework and executed in online platforms such as Google Collab and Kaggle.

The data has been split into 80:10:10 for training, validation and

testing. This study completely utilizes the different versions of EfficientNetB0-B7. Here, validation and testing dataset are completely unseen data for the model. After several trial and error, the hyperparameters for the EfficientNet models have been chosen as: learning rate - 0.0001, L1 regularization - 0.01, L2 regularization - 0.01, activation function - ReLU, drop out - 0.4, last layer activation function - softmax and loss function - categorical cross entropy. The optimizer for the models has been varied among SGD, RMSProp and Adam.

Figs. 5 and 6 show the Training accuracy and training loss curves for all combinations of EfficientNetBx and optimizers respectively. Similarly Figs. 7 and 8 represents validation accuracy and validation loss curves for all combinations of EfficientNetBx and optimizers

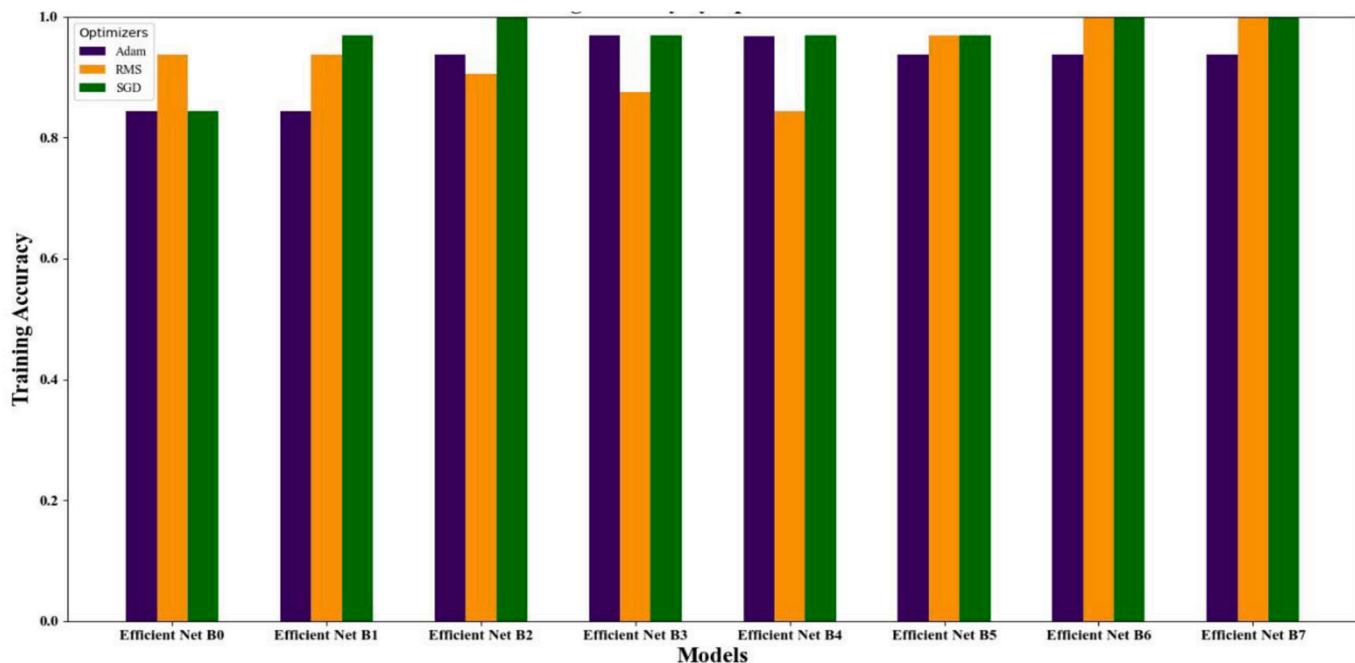


Fig. 9. Training accuracy for different models w.r.t Adam, RMSprop and SGD optimizers.

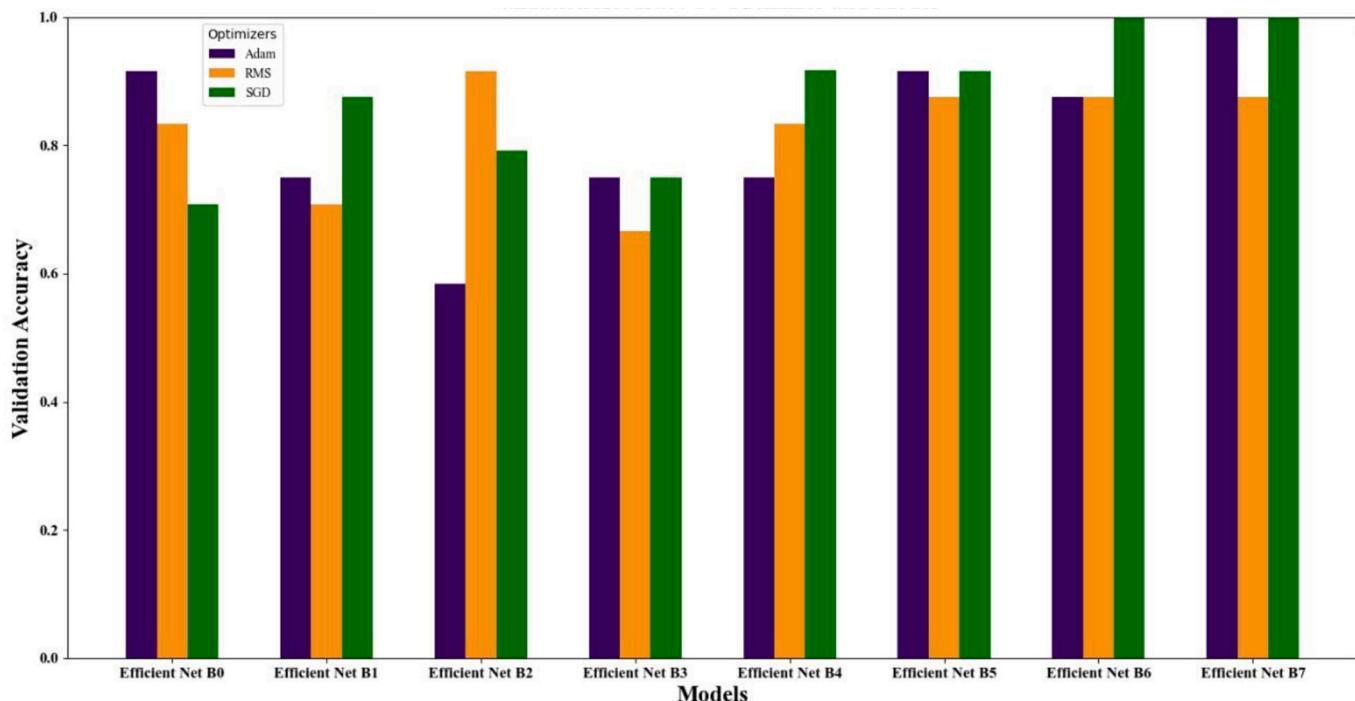


Fig. 10. Validation accuracy for different models w.r.t Adam, RMSprop and SGD optimizers.

respectively. It is observed that though Adam optimizer has dynamic curve, the final accuracy obtained through Adam optimizer is better than other optimizers with majority of the models. A steep increase in performance has been observed with EfficientNetB5 and EfficientNetB7 with validation dataset.

Figs. 9, Fig. 10, & Fig. 11 represents the accuracy obtained with different combinations of EfficientNetBx + optimizer algorithms for training and validation dataset, during the training phase. From Fig. 9 it is observed that while training EfficientnetB0 model, RMSProp has given a better accuracy of 0.94 compared to other optimizers.

On the other hand, SGD optimizer has given a better training accuracy of 0.91 and 1.0 with EfficientnetB1 and EfficientnetB2 respectively, Adam and SGD have given high comparable training accuracy of 0.97 with EfficientnetB3 and EfficientnetB4, RMSprop and SGD have given similar high training accuracy of 0.97, 1.0 and 1.0 with EfficientnetB5, EfficientnetB6 and EfficientnetB7 respectively. It is worthy to note that the highest training accuracy of 1.00 has been obtained with EfficientnetB2+SGD, EfficientnetB6+RMS/SGD and EfficientnetB7+RMS/SGD.

From, Fig. 10 it is observed that in the validation phase, overall

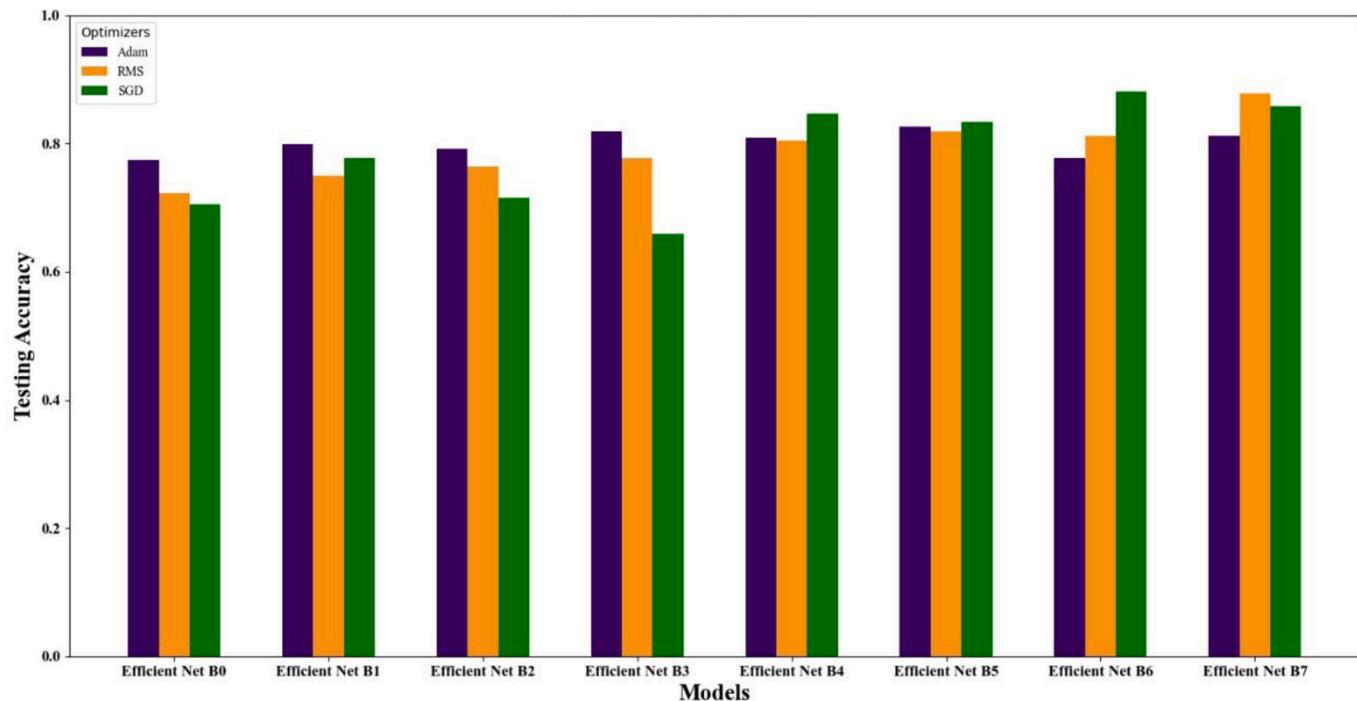


Fig. 11. Testing accuracy for different models w.r.t Adam, RMSprop and SGD optimizers.

Table 1

Sensitivity and specificity for different combinations of EfficientNetBx + Optimizer in the validation phase.

EfficientNetBx + Optimizer	Sensitivity					Specificity				
	Nor	Mil	Mod	Sev	Hyp	Nor	Mil	Mod	Sev	Hyp
EfficientNetB0 + Adam	0.66	0.80	0.75	0.93	0.82	0.95	0.90	0.96	0.96	0.97
EfficientNetB0 + RMSprop	0.70	0.67	0.74	0.93	0.87	0.91	0.95	0.96	0.98	0.93
EfficientNetB0 + SGD	0.74	0.43	0.70	0.92	0.60	0.76	0.99	0.96	0.91	0.97
EfficientNetB1 + Adam	0.83	0.71	0.73	0.91	0.86	0.86	0.96	0.98	0.99	0.96
EfficientNetB1 + RMSprop	0.80	0.74	0.63	0.79	0.88	0.86	0.93	0.97	0.99	0.95
EfficientNetB1 + SGD	0.70	0.80	0.61	0.93	0.76	0.92	0.86	0.98	0.96	0.98
EfficientNetB2 + Adam	0.75	0.70	0.89	0.89	0.86	0.93	0.96	0.95	0.97	0.96
EfficientNetB2 + RMSprop	0.78	0.58	0.81	0.81	0.79	0.87	0.98	0.94	0.97	0.94
EfficientNetB2 + SGD	0.78	0.69	0.89	0.93	0.72	0.89	0.99	0.95	0.94	0.98
EfficientNetB3 + Adam	0.73	0.71	0.89	0.97	0.82	0.93	0.95	0.93	0.99	0.98
EfficientNetB3 + RMSprop	0.81	0.48	0.88	0.94	0.84	0.88	0.99	0.93	0.99	0.95
EfficientNetB3 + SGD	0.62	0.58	0.54	0.99	0.63	0.94	0.98	0.95	0.73	0.99
EfficientNetB4 + Adam	0.91	0.75	0.78	0.97	0.80	0.88	0.97	0.97	0.99	0.99
EfficientNetB4 + RMSprop	0.87	0.70	0.80	0.92	0.84	0.89	0.98	0.95	0.99	0.97
EfficientNetB4 + SGD	0.40	0.39	0.35	0.69	0.88	0.90	0.94	0.94	0.96	0.69
EfficientNetB5 + Adam	0.73	0.77	0.91	0.95	0.83	0.94	0.96	0.97	0.98	0.95
EfficientNetB5 + RMSprop	0.73	0.76	0.88	0.94	0.88	0.94	0.97	0.97	0.99	0.94
EfficientNetB5 + SGD	0.77	0.84	0.87	0.94	0.88	0.94	0.95	0.97	0.99	0.97
EfficientNetB6 + Adam	0.48	0.72	0.86	0.96	0.88	0.97	0.96	0.91	0.94	0.94
EfficientNetB6 + RMSprop	0.80	0.77	0.87	0.98	0.88	0.93	0.97	0.98	0.98	0.96
EfficientNetB6 + SGD	0.86	0.83	0.88	0.93	0.92	0.95	0.98	0.96	0.99	0.97
EfficientNetB7 + Adam	0.79	0.75	0.91	0.96	0.95	0.96	0.98	0.97	0.97	0.96
EfficientNetB7 + RMSprop	0.79	0.83	0.93	0.94	0.88	0.95	0.96	0.98	0.98	0.97
EfficientNetB7 + SGD	0.80	0.79	0.88	0.93	0.91	0.95	0.97	0.99	0.99	0.96

*Normal-Nor, Mild-Mil, Moderate-Mod, Severe-Sev, Hyperdynamic-Hyp.

different optimizer gives better accuracy with each model. EfficientNetB0 model gives a better validation accuracy of 0.92 with Adam optimizer. On the other hand, EfficientNetB1, EfficientNetB4, and EfficientNetB6 gives a better validation accuracy of 0.88, 0.92 and 1.00 respectively with SGD optimizer. In contrary, RMSprop optimizer and EfficientNetB2 achieved a better validation accuracy of 0.92. Both Adam and SGD optimizer with EfficientNetB3, EfficientNetB5, and EfficientNetB7 yielded a better validation accuracy of 0.75, 0.92 and 1.00 respectively. To summarize the validation phase, both EfficientNetB6, and EfficientNetB7 achieved the highest accuracy of 1.00 with SGD

optimizer.

In Fig. 11, testing phase shows the better generalization ability of EfficientNetBx + optimizer combination to unseen data. Adam optimizer with EfficientNetB0, EfficientNetB1, EfficientNetB2 and EfficientNetB3 have resulted in better testing accuracy of 0.77, 0.80, 0.79, and 0.82 respectively. Subsequently, EfficientNetB4, EfficientNetB5 and EfficientNetB6 produced a better testing accuracy of 0.85, 0.83 and 0.88 respectively with SGD optimizer. Finally, the highest testing accuracy of 0.88 has been obtained with EfficientNetB6 + SGD and EfficientNetB7 + RMSprop. Further, the sensitivity and specificity obtained with these

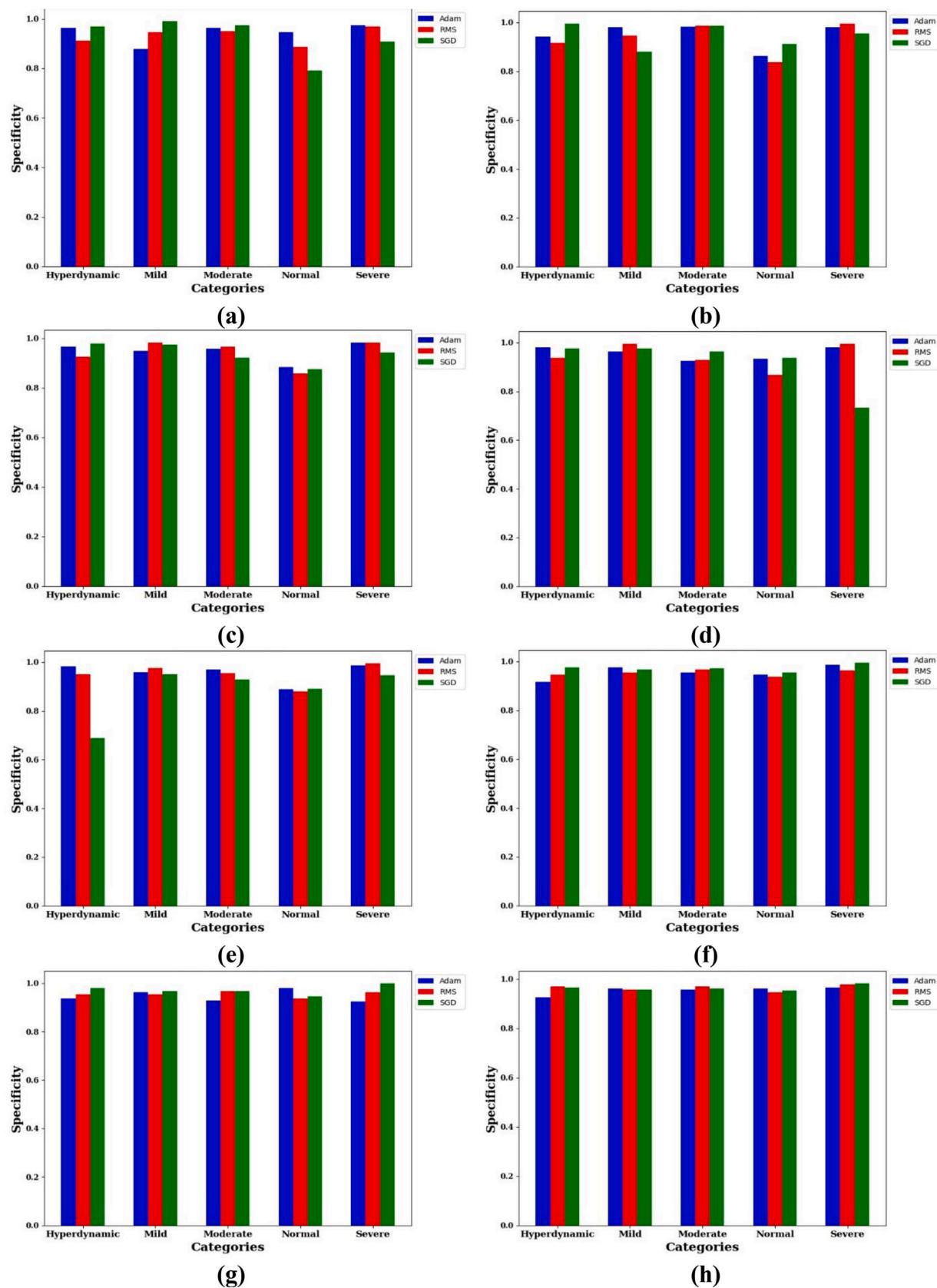


Fig. 12. Specificity for different combinations of EfficientNetBx + Optimizer in testing phase: a) EfficientNetB0, b) EfficientNetB1, c) EfficientNetB2, d) EfficientNetB3, e) EfficientNetB4, f) EfficientNetB5, g) EfficientNetB6 and h) EfficientNetB7.

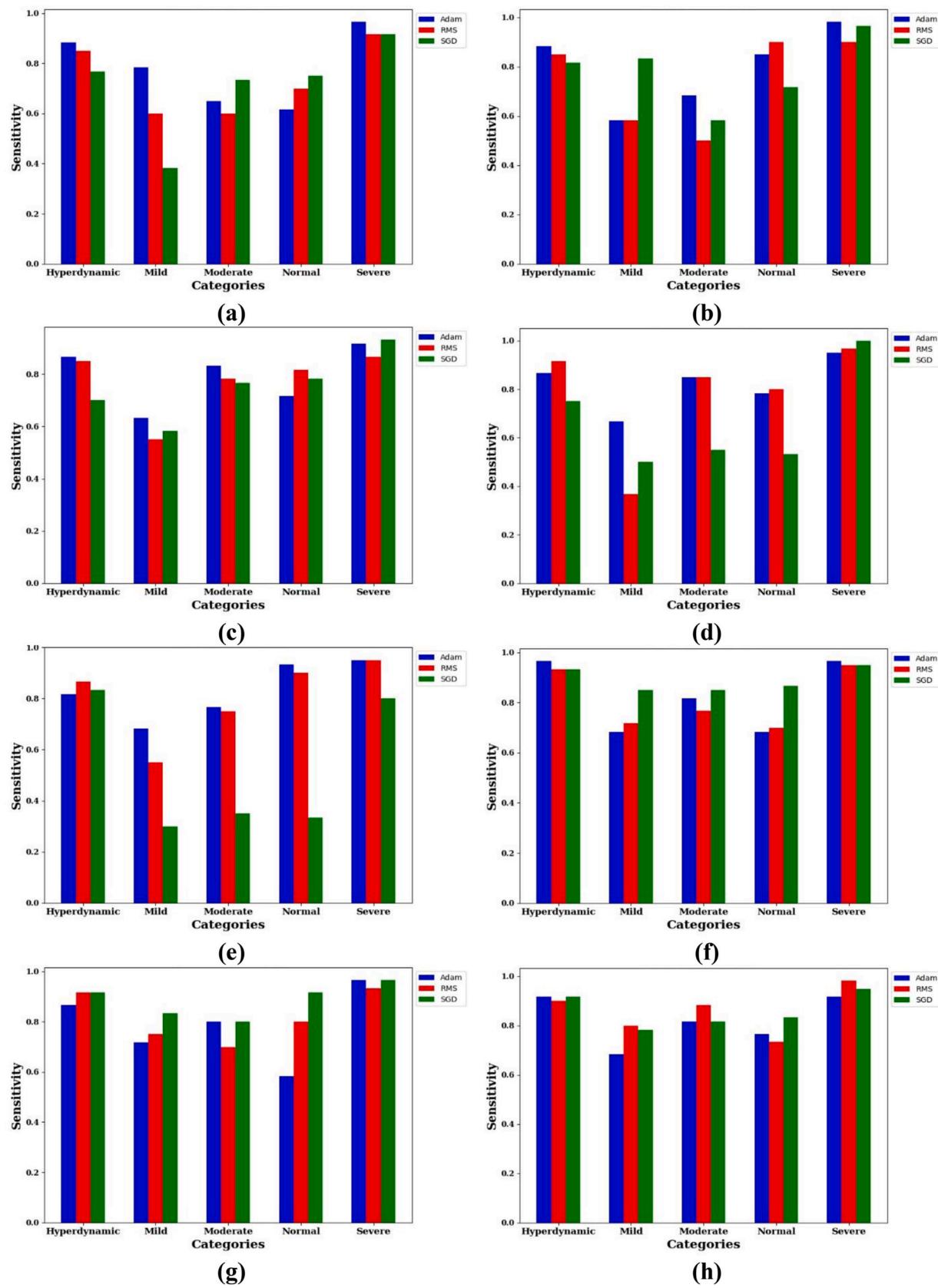


Fig. 13. Sensitivity for different combinations of EfficientNetBx + Optimizer in testing phase: a) EfficientNetB0, b) EfficientNetB1, c) EfficientNetB2, d) EfficientNetB3, e) EfficientNetB4, f) EfficientNetB5, g) EfficientNetB6 and h) EfficientNetB7.

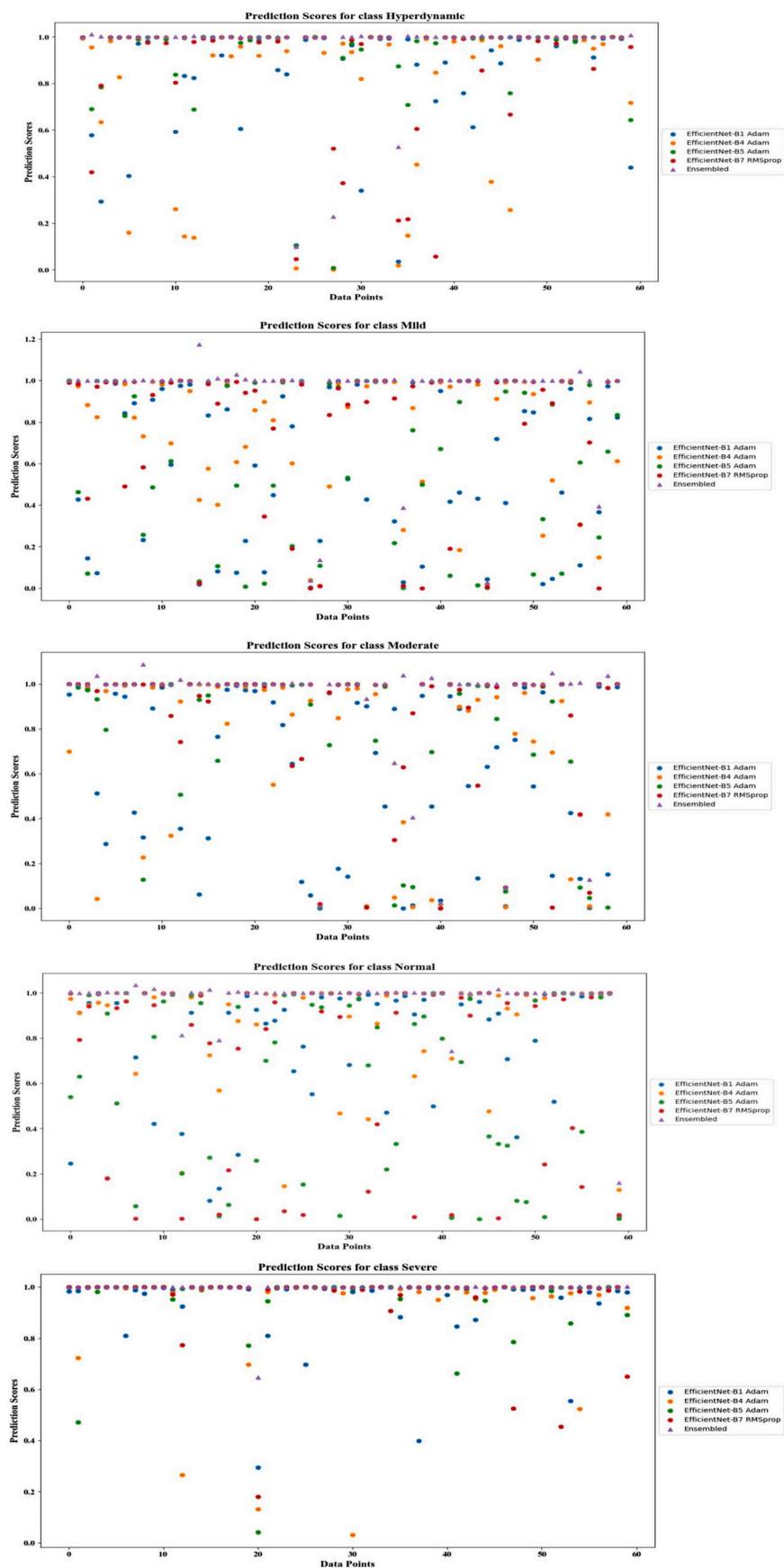


Fig. 14. Probability scores for different category of subjects.

Table 2

Performance comparison of the proposed method with representative algorithms in terms of training, validation, and testing accuracy.

Model	Training Accuracy (%)	Validation Accuracy (%)	Testing Accuracy (%)
DenseNet 121	98	80	78
DenseNet 201	97	80	81
XceptionNet	100	100	80.9
Inception V1	96	79	78
MobileNet V1	100	79.2	76.4
EfficientNetB0	93.8	91.7	77.4
EfficientNetB1	96.9	95.8	79.9
EfficientNetB2	100	100	79.2
EfficientNetB3	100	95.9	81.9
EfficientNetB4	100	100	84.7
EfficientNetB5	100	91.7	83.3
EfficientNetB6	100	100	88.2
EfficientNetB7	100	100	87.9

models need to be analysed in detail to understand the performance of the model towards each individual class of data.

Specificity is an indicator of a model's ability to reject the inputs which doesn't pertain to that specific class. The sensitivity and specificity for each considered sub-category achieved with different combinations of EfficientNetBx + Optimizer for the validation phase has been tabulated in [Table 1](#). After tuning the model parameters during validation phase, the highest sensitivity has been acquired by EfficientNetB4+Adam(0.91), EfficientNetB5+SGD(0.84), EfficientNetB7 + RMSprop(0.93), EfficientNetB3 + SGD(0.99), and EfficientNetB7 + Adam(0.95) for normal, mild, moderate, sever and hyperdynamic classes.

[Fig. 12](#) portrays the specificity achieved for individual classes with EfficientNetBx + Optimizer combinations for the unseen test data. Highest specificity is produced by EfficientNetB1 + SGD (1.00), EfficientNetB0 + SGD (0.99), EfficientNetB1 + SGD (0.99), EfficientNetB6 + Adam (0.98) and EfficientNetB6 + SGD (1.00) for hyperdynamic, mild, moderate, normal and severe categories respectively. EfficientNetB0 + SGD is better than other models in rejecting the input images that doesn't belong to mild class. It is evident that EfficientNetB1 + SGD is the least prone to give positive results for hyperdynamic and moderate stages, in their absence. Similarly, EfficientNetB6 + Adam and EfficientNetB6 + SGD better rules out normal and severe categories than other models.

The sensitivity achieved for individual classes with EfficientNetBx + Optimizer combinations for the test data is visually shown in [Fig. 13](#). The greatest sensitivity is obtained by EfficientNetB5 + Adam (0.97), EfficientNetB1 + SGD (0.83), EfficientNetB7 + RMSprop (0.88), EfficientNetB4 + Adam (0.90), and EfficientNetB1 + Adam (0.98) for hyperdynamic, mild, moderate, normal and severe categories

respectively compared to other models. This shows the superiority of aforesaid models in truly detecting the particular class. Hence the probability scores obtained from EfficientNetB5 + Adam, EfficientNetB7 + RMSprop, EfficientNetB4 + Adam, and EfficientNetB1 + Adam are weighted aggregated to obtain the final prediction. The weight parameters for ensemble has been chosen based on trial and error method.

In order to showcase the efficacy of the proposed ensemble model, the probability scores obtained from top performing individual models and ensemble model has been plotted in [Fig. 14](#). The probability scores computed by EfficientNetB5 + Adam, EfficientNetB7 + RMSprop, EfficientNetB4 + Adam, and EfficientNetB1 + Adam models are plotted as circles in different colours. The probability score generated by the ensemble model has been plotted as triangle. It is evident from the figure that the ensemble model provides the best probability scores in majority of the cases than individual models for all the classes.

4. Discussion

The representative algorithms such as DenseNet 121, DenseNet 201 [21], MobileNet v1, XceptionNet and Inception v1 [22] have been compared with proposed EfficientNet variants for heart failure severity diagnosis using cardiac MRI.

The [Table 2](#) indicates the accuracy obtained during training, validation and testing phase with the representative algorithms and different versions of EfficientNet. It is observed that the EfficientNetB4, EfficientNetB6 and EfficientNetB7 perform better than other representative algorithms with high training, validation and testing accuracy. This highlights their superior performance compared to other state of art methods in the study. Further in order to examine the individual class prediction performance of various methods, [Tables 3 and 4](#) displays the individual class sensitivity and specificity achieved with all approaches for training and validation phase respectively.

It is noted that during training, EfficientNet variants are performing only comparable to DenseNet 201, XceptionNet and MobileNet v1. This shows the efficacy of Efficient models in capturing the complex patterns similar to representative algorithms. As an add-on, the compound scaling of network depth, width and resolution results in optimal utilization of computational resource by EfficientNet. However, during validation, individual class sensitivity and specificity of the EfficientNet variants is better than other representative algorithms. This evidently proves the generalization capability of EfficientNet variants. The use of squeeze and excitation blocks, optimized feature extraction layers has allowed EfficientNet variants to generalize across validation.

Though CMR is considered to be the gold standard for CVD analysis. The major road blockers in wide usage of CMR in clinical practise is the lack of experts for CMR interpretation. Hence, an automated CMR interpretation tool would be a great assist to cardiologist in especially

Table 3

Comparative analysis of individual class sensitivity and specificity during training.

Model	Sensitivity					Specificity				
	Nor	Mil	Mod	Sev	Hyp	Nor	Mil	Mod	Sev	Hyp
DenseNet 121	0.95	0.98	0.99	0.99	1	1	0.98	0.99	0.97	0.98
DenseNet 201	0.98	0.99	0.91	1	0.99	1	0.98	1	0.97	0.93
XceptionNet	0.99	0.99	1	1	1	0.99	0.99	0.99	0.99	0.99
Inception V1	0.99	0.99	0.92	0.93	0.95	0.95	0.91	0.98	0.97	0.98
MobileNet V1	0.96	0.93	0.97	1	0.97	1	0.99	1	0.87	1
EfficientNetB0	0.91	0.97	0.90	0.99	0.93	0.98	0.97	0.99	0.99	0.99
EfficientNetB1	0.98	0.96	0.95	0.98	0.97	0.98	0.99	0.99	0.99	0.99
EfficientNetB2	0.97	0.97	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
EfficientNetB3	0.99	0.86	0.99	0.99	1	0.98	0.99	0.99	0.99	0.99
EfficientNetB4	1	0.99	0.99	1	0.98	0.99	0.99	1	0.99	1
EfficientNetB5	1	1	0.99	1	1	0.99	1	1	1	1
EfficientNetB6	1	1	0.99	1	1	0.99	1	1	1	0.99
EfficientNetB7	1	1	1	1	1	1	1	1	1	1

*Normal-Nor, Mild-Mil, Moderate-Mod, Severe-Sev, Hyperdynamic-Hyp.

Table 4

Comparative analysis of individual class sensitivity and specificity during validation.

Model	Sensitivity					Specificity				
	Nor	Mil	Mod	Sev	Hyp	Nor	Mil	Mod	Sev	Hyp
DenseNet 121	0.78	0.71	0.83	0.99	0.67	0.72	0.84	0.79	0.76	0.94
DenseNet 201	0.87	0.80	0.65	0.89	0.77	0.62	0.81	0.87	0.86	0.93
XceptionNet	0.70	0.80	0.61	0.93	0.76	0.92	0.86	0.98	0.96	0.98
Inception V1	0.64	0.84	0.71	0.85	0.89	0.79	0.68	0.87	0.89	0.76
MobileNet V1	0.93	0.67	0.71	0.84	0.70	0.54	0.81	0.84	0.98	0.94
EfficientNetB0	0.74	0.43	0.70	0.92	0.60	0.76	0.99	0.96	0.91	0.97
EfficientNetB1	0.83	0.71	0.73	0.91	0.86	0.86	0.96	0.98	0.99	0.96
EfficientNetB2	0.78	0.58	0.81	0.81	0.79	0.87	0.98	0.94	0.97	0.94
EfficientNetB3	0.81	0.48	0.88	0.94	0.84	0.88	0.99	0.93	0.99	0.95
EfficientNetB4	0.91	0.75	0.78	0.97	0.80	0.88	0.97	0.97	0.99	0.99
EfficientNetB5	0.77	0.84	0.87	0.94	0.88	0.94	0.95	0.97	0.99	0.97
EfficientNetB6	0.86	0.83	0.88	0.93	0.92	0.95	0.98	0.96	0.99	0.97
EfficientNetB7	0.79	0.83	0.93	0.94	0.88	0.95	0.96	0.98	0.98	0.97

*Normal-Nor, Mild-Mil, Moderate-Mod, Severe-Sev, Hyperdynamic-Hyp.

HF prognosis studies, as it demands the close monitoring of cardiac functional improvements over the timeline. The heart undergoes structural changes such as ventricular dilation, ventricular myocardium thickening, myocardium fibrosis and heart chamber shape changes with HF progression. Understanding the functional and structural changes in heart is crucial in determination of targeted therapies for effective management of patients. Hence, automated CMR interpretation will assist the physicians in a better way to understand the underlying HF etiology. The inclusion of federated learning will enable the healthcare sector to provide effective solutions by harnessing the power of collective data without affecting data privacy standards, as the learning will be happening at the place of data acquisition itself.

5. Conclusion

In this study, a framework based on ensemble prediction of EfficientNet has been proposed to identify heart failure severity through CMR images. The developed framework takes full advantage of the different patterns learnt by four different EfficientNet variants. Each EfficientNet variant is better in detecting specific class of the disease. Hence, the weighted ensemble of multiple prediction results in better overall prediction. Compared to the traditional texture based learning methods and state of the art deep learning models, the ensemble prediction is able to achieve a better test accuracy for good count of unseen data.

The authors have already explored the efficacy of different EfficientNet models in identifying type of diabetic foot ulcer in their previous study. Hence, it is evident that the variants of EfficientNet model could be a better choice for any automated image-based diagnosis framework. However, still there is scope of localization of damaged myocardium or anatomical cavity change to give precise interpretation to cardiologist for effective planning of treatment.

CRediT authorship contribution statement

Muthunayagam Muthulakshmi: Writing – original draft, Validation, Software, Methodology, Investigation, Conceptualization. **Kottesarwan Venkatesan:** Writing – review & editing, Visualization, Validation, Formal analysis. **Balaji Prasanalakshmi:** Supervision, Resources. **Rahayu Syarifah Bahiyah:** Validation, Project administration, Funding acquisition. **Vijayakumar Divya:** Software, Data curation, Conceptualization.

Consent to participate

This study includes the images from publicly available database (<https://www.kaggle.com/c/second-annual-data-science-bowl>) and the database has been cited.

Ethics approval

This article does not contain any studies with animals performed by any of the authors.

Data availability statement

Data sharing is not applicable to this article because no new data were created or analysed in this study.

Funding

This research was supported and financially funded from the National Defence University Malaysia (UPNM) - UPNM/2023/GPPP/ICT/1CT/1.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was conducted at the Cyber Security & Digital Industrial Revolution Centre, National Defence University Malaysia (UPNM), with support and financial funding from the National Defence University Malaysia (UPNM) - UPNM/2023/GPPP/ICT/1CT/1.

References

- [1] Luengo-Fernandez R, Little M, Gray A, Torbica A, Maggioni AP, Huculeci R, Timmis AD, Vardas P, Leal J. Cardiovascular disease burden due to productivity losses in European Society of Cardiology countries. Eur Heart J Qual Care Clin Outcomes 2024;10:36–44. <https://doi.org/10.1093/ejqcco/qcad031>.
- [2] Mandoli GE, Spaccaferri L, Carluccio E, Inciardi RM. Editorial: methods in diagnosing heart failure. Front Cardiovasc Med 2024;11. <https://doi.org/10.3389/fcvm.2024.1365006>.
- [3] Barison A, Aimo A, Todiere G, Grigoratos C, Aquaro GD, Emdin M. Cardiovascular magnetic resonance for the diagnosis and management of heart failure with preserved ejection fraction. Heart Fail Rev 2022;27:191–205. <https://doi.org/10.1007/s10741-020-09998-w>.
- [4] Popovic D, Alogna A, Omar M, Sorimachi H, Omote K, Reddy YNV, Redfield MM, Burkhoff D, Borlaug BA. Ventricular stiffening and chamber contracture in heart failure with higher ejection fraction. Eur J Heart Fail 2023;25:657–68. <https://doi.org/10.1002/ejhf.2843>.
- [5] Wang Y-R, Yang K, Wen Y, Wang P, Hu Y, Lai Y, Wang Y, Zhao K, Tang S, Zhang A, Zhan H, Lu M, Chen X, Yang S, Dong Z, Wang Y, Liu H, Zhao L, Huang L, Li Y, Wu L, Chen Z, Luo Y, Liu D, Zhao P, Lin K, Wu JC, Zhao S. Screening and diagnosis of cardiovascular disease using artificial intelligence-enabled cardiac magnetic resonance imaging. Nat Med 2024;30:1471–80. <https://doi.org/10.1038/s41591-024-02971-2>.

- [6] Hao Y, Zhang R, Chen L, Fan G, Liu B, Jiang K, Zhu Y, Zhang M, Guo J. Distinguishing heart failure subtypes: the diagnostic power of different cardiac magnetic resonance imaging parameters. *Front Cardiovasc Med* 2024;11. <https://doi.org/10.3389/fcvm.2024.1291735>.
- [7] Zhang H, Zhao L, Wang H, Yi Y, Hui K, Zhang C, Ma X. Radiomics from cardiovascular MR cine images for identifying patients with hypertrophic cardiomyopathy at high risk for heart failure. *Radiol Cardiothorac Imaging* 2024;6. <https://doi.org/10.1148/rct.230323>.
- [8] Zhang Q, Fotaki A, Ghadimi S, Wang Y, Doneva M, Wetzl J, Delfino JG, O'Regan DP, Prieto C, Epstein FH. Improving the efficiency and accuracy of CMR with AI – review of evidence and proposition of a roadmap to clinical translation. *J Cardiovasc Magn Reson* 2024;10:1051. <https://doi.org/10.1016/j.jcmr.2024.101051>.
- [9] Boribalburephan A, Treewaree S, Tantisiriwat N, Yindeengam A, Achakulvisut T, Krittayaphong R. Myocardial scar and left ventricular ejection fraction classification for electrocardiography image using multi-task deep learning. *Sci Rep* 2024;14:7523. <https://doi.org/10.1038/s41598-024-58131-6>.
- [10] Gao Y, Zhou Z, Zhang B, Guo S, Bo K, Li S, Zhang N, Wang H, Yang G, Zhang H, Liu T, Xu L. Deep learning-based prognostic model using non-enhanced cardiac cine MRI for outcome prediction in patients with heart failure. *Eur Radiol* 2023;33:8203–13. <https://doi.org/10.1007/s00330-023-09785-9>.
- [11] Wang Z, Fan Z, Liu X, Zhu M, Jiang S, Tian S, Chen B, Wu L. Deep learning for discrimination of hypertrophic cardiomyopathy and hypertensive heart disease on <sc>MRI</sc> native <sc>T1</sc> maps. *J Magn Reson Imag* 2023. <https://doi.org/10.1002/jmri.28904>.
- [12] Petmezas G, Papageorgiou VE, Vassilikos V, Pagourelas E, Tsaklidis G, Katsaggelos AK, Maglaveras N. Recent advancements and applications of deep learning in heart failure: A systematic review. *Comput Biol Med* 2024;176:108557. <https://doi.org/10.1016/j.combiomed.2024.108557>.
- [13] Zhang X, Cui C, Zhao S, Xie L, Tian Y. Cardiac magnetic resonance radiomics for disease classification. *Eur Radiol* 2022;33:2312–23. <https://doi.org/10.1007/s00330-022-09236-x>.
- [14] Xie Y, Zhong H, Wu J, Zhao W, Hou R, Zhao L, Xu X, Zhang M, Zhao J. Automatic classification of heart failure based on Cine-CMR images. *Int J Comput Assist Radiol Surg* 2023;19:355–65. <https://doi.org/10.1007/s11548-023-03028-4>.
- [15] Ng M-Y, Kwan CT, Yap PM, Fung SY, Tang HS, Tse WWV, Kwan CNF, Chow YHP, Yiu NC, Lee YP, Fong AHT, Hwang S, Fong ZFW, Ren Q-W, Wu M-Z, Wan EYF, Lee KCK, Leung CY, Li A, Montero D, Vardhanabhoti V, Hai J, Siu C-W, Tse H-F, Pennell DJ, Mohiaddin R, Senior R, Yiu K-H. Diagnostic accuracy of cardiovascular magnetic resonance strain analysis and atrial size to identify heart failure with preserved ejection fraction. *European Heart J Open* 2023;3. <https://doi.org/10.1093/ehjopen/oead021>.
- [16] You Y, Viktorovich LA, Qiu J, Nikolaevich KA, Vladimirovich BY. Cardiac magnetic resonance image diagnosis of hypertrophic obstructive cardiomyopathy based on a double-branch neural network. *Comput Methods Progr Biomed* 2021;200:105889. <https://doi.org/10.1016/j.cmpb.2020.105889>.
- [17] Budai A, Suhai FI, Csorba K, Dohy Z, Szabo L, Merkely B, Vago H. Automated classification of left ventricular hypertrophy on cardiac MRI. *Appl Sci* 2022;12:4151. <https://doi.org/10.3390/app12094151>.
- [18] Jiang S, Zhang L, Wang J, Li X, Hu S, Fu Y, Wang X, Hao S, Hu C. Differentiating between cardiac amyloidosis and hypertrophic cardiomyopathy on non-contrast cine-magnetic resonance images using machine learning-based radiomics. *Front Cardiovasc Med* 2022;9. <https://doi.org/10.3389/fcvm.2022.1001269>.
- [19] Khozeimeh F, Sharifrazi D, Izadi NH, Joloudari JH, Shoeibi A, Alizadehsani R, Tartibi M, Hussain S, Sani ZA, Khodatars M, Sadeghi D, Khosravi A, Nahavandi S, Tan R-S, Acharya UR, Islam SMS. RF-CNN-F: random forest with convolutional neural network features for coronary artery disease diagnosis based on cardiac magnetic resonance. *Sci Rep* 2022;12:11178. <https://doi.org/10.1038/s41598-022-15374-5>.
- [20] Talukder MdA, Layek MdA, Kazi M, Uddin MdA, Aryal S. Empowering COVID-19 detection: optimizing performance through fine-tuned EfficientNet deep learning architecture. *Comput Biol Med* 2024;168:107789. <https://doi.org/10.1016/j.combiomed.2023.107789>.
- [21] Mofrad FB, Valizadeh G. DenseNet-based transfer learning for LV shape Classification: introducing a novel information fusion and data augmentation using statistical Shape/Color modeling. *Expert Syst Appl* 2023;213:119261. <https://doi.org/10.1016/j.eswa.2022.119261>.
- [22] Ho N, Kim YC. Evaluation of transfer learning in deep convolutional neural network models for cardiac short axis slice classification. *Sci Rep* 2021;11:1839. <https://doi.org/10.1038/s41598-021-81525-9>.