*Journal of Computer Science*

Original Research Paper

# Cardiovascular Disease Prediction through Ensembled Transfer Learning on Cardiac Magnetic Resonance Imaging

**[1]Sibu Cyriac, [1]Sivakumar R. and [2]Nidhin Raju**

*[1]Department of Computer Science, CHRIST (Deemed to be University), India*
*[2]Centre for Digital Innovation, CHRIST (Deemed to be University), India*

*Article history*
Received: 15-07-2022
Revised: 07-09-2022
Accepted: 22-09-2022

Corresponding Author:
Sibu Cyriac
Department of Computer Science, CHRIST (Deemed to be University), India
Email: sibu.cyriac@christuniversity.in

**Abstract:** Cardiovascular Diseases (CVD) cause more deaths worldwide than most of the other diseases. The diagnosis of cardiovascular disease from Magnetic Resonance Imaging plays a major role in the medical field. The technological revolution contributed a lot to increase the effectiveness of CVD diagnosis. Many Artificial Intelligence methods using Deep Learning models are available to assist the cardiologist in the diagnosis of CVD from Magnetic Resonance Imaging (MRI). In this study, we leverage on the merits of deep learning, transfer learning, and ensemble voting to improve the accuracy of Artificial Intelligence-based CVD detection. VGG16, MobileNetV2, and InceptionV3, trained on ImageNet, are the models used and the dataset is the Automatic Cardiac Diagnosis Challenge dataset. We customized the classification layers of all three models to suit the CVD detection problem. The results from these models are ensembled using the soft-voting and hard-voting approaches. Test accuracies obtained are 97.94% and 98.08% from hard-voting and soft-voting respectively. The experimental results demonstrated that the ensemble of outputs from transfer learning-based Deep Learning models produces much improved results for CVD diagnosis from MRI images.

**Keywords:** Cardiovascular Disease, Deep Learning, Medical Imaging, ACDC Dataset, Transfer Learning

## Introduction

Cardiovascular Disease (CVD) is listed as the foremost cause of death by the World Health Organization (WHO), with 17.9 million people dying yearly. Lifestyle-related issues such as overweight and obesity, hypertension, hyperglycemia, and high cholesterol increase the risk of heart-related diseases (Nolf *et al*., 2003). Furthermore, the American Heart Association adds symptoms like chronic cough and high heart rate, sleep problems, weight gain, and leg swelling (Piña *et al*., 2003) into this list. A report from WHO says that 67% is the best accuracy with which medical practitioners can correctly predict heart disease from clinical tests (Organization, 2007). On the other hand, these medical practitioners have massive medical data available that can be leveraged to predict cardiovascular diseases. These data can be used in information technology-enabled medical support systems to train AI-based disease prediction models (Shalev-Shwartz and Ben-David, 2014) (Friedman, 2017; Marsland, 2011).

In developing countries, the diagnosis and treatment of cardiac diseases are very complex due to the lack of modern diagnostic equipment and the unavailability of experienced physicians. It seriously affects the timely prediction and treatment of the disease. Due to these, the cardiac diagnosing approaches result in imprecise or delayed diagnosis. Furthermore, it is costlier as such computationally demanding and thus takes more time for the diagnosis (Haq *et al*., 2018). A physician may order specific tests to inspect symptoms of suspected cardiovascular diseases. It includes chest X-ray, blood tests, blood pressure monitoring, stress tests, and Electrocardiogram (ECG). ECG is a simple and very common, yet noninvasive diagnostic test that checks the rhythm and electrical activity of the heart. However, an asymptomatic patient may be diagnosed with normal electrocardiogram rhythm by ECG and it has certain limitations as a prognostic tool for predicting the future occurrence of heart diseases. Of late, medical practitioners started using angiograms as a rule for cardiovascular disease detection and diagnosis Shu *et al*. (2017). A patient's clinical history, physical test results, and examination of connected symptoms by the physician were considered in traditional approaches. Angiography is regarded as one of the most precise procedures for identifying heart problems among traditional methods. It, on the other hand, has a few disadvantages such as it requires high cost, expert knowledge, and results in different side effects. These traditional techniques frequently result in uncertain results and take more time (Muhammad *et al*., 2020).

Magnetic Resonance Imaging (MRI) is another important approach to medical diagnosis. In common cardiovascular disease diagnosis approaches, clinicians do manual segmentation of the MRI images to diagnose cardiac problems. However, this manual segmentation requires more labor and is time-consuming. The heart MRI gives important data for CVD detection by empowering quantitative appraisal of useful parameters such as myocardium thickness, the volume of the Left Ventricle (LV) and Right Ventricle (RV), and Ejection Fraction (EF). Thus, cardiovascular MRI segmentation has turned into an arising medical imaging field (Brewer *et al*., 2015). Because of the unique qualities of cardiovascular MRI, heart segmentation is a difficult task. For example, the brightness of LV intracavity varies from time to time and sometimes it shows similarities with signals of other heart organs. These irregularities make the manual assessment of cardiac diagnosis from MRI images more complex. Many existing approaches segment either the LV or the RV, but the diagnosis of some cardiac issues requires both LV and RV. Experiments to segment both LV and RV simultaneously are being undertaken by a few researchers (Hayes *et al*., 2008).

With the invention and invasion of information technology into the medical field, many researchers started looking for technology-assisted methods for more cost-effective and better methods of diagnosing cardiac diseases. These intelligent systems can be used to assist physicians with a second opinion (Tama *et al*., 2020). A multivariate regression assessment through a longitudinal study can yield a risk forecasting model for coronary disease (Muthuvel *et al*., 2018). Different data mining approaches such as association rules (Ordonez *et al*., 2001), Apriori, Predictive Apriori, Tertius (Nahar *et al*., 2013), and more were used earlier to select, explore and model a large amount of patient's data for the cardiac diagnosis (AbuKhousa and Campbell, 2012).

Many studies have started applying Machine Learning (ML) based clinical trials to diagnose and predict various cardiac problems with decent accuracy. Specifically, efficient implementation of clinically designed ML algorithms improves the efficiency of the health care system in diagnosing CVD. There are frequently numerous elements that contribute to recognizing patients who are at risk for such common illnesses. ML techniques can help determine unseen patterns in these elements that might otherwise be ignored (Dinh *et al*., 2019). The advanced ML-based CVD monitoring technologies also provide a real-time diagnosis for personalized care (Krittanawong *et al*., 2020). Similarly, many researchers use ML algorithms when the collected data are in statistical form. Such data requires extensive screening and processing to efficiently extract features (Asfi-Ar- *et al*., 2021). Nevertheless, the analysis of a huge dataset using traditional statistical approaches was impractical in most cases. Therefore, ML algorithms have emerged as one most significant tool in this modern era in analyzing and processing such statistical data for the development of the medical sector (Asfi-Ar *et al*., 2021).

Cardiovascular imaging also plays an important role in the diagnosis of heart diseases and thus Deep Learning (DL) provides another opportunity to analyze these images for cardiovascular disease prediction (Martin-Isla *et al*., 2020). DL has become a prominent technology to effectively assist in many medical problems such as diagnosis, prediction, and intervention (Bizopoulos and Koutsouris, 2019). The latest developments in DL have shown pathbreaking efficiency in image segmentation and detection. Nowadays, Convolutional Neural Networks (CNNs) are the backbone of DL because of their fastest implementation and ability to extract many features from the input data. Compared with traditional ML methods, Deep CNNs have achieved much better positive results in image segmentation, detection, and classification tasks. The CNNs are made to look like neuronal patterns of bio connectivity and this helps to extract more features from the entire image than done manually (Shivamurthy *et al*., 2014; Farag and Fakhreldin, 2012).

The main aim of our proposed work is to detect cardiovascular disease using AI-based methods. We used models trained on ImageNet weights to classify CVD from the Automated Cardiac Diagnosis Challenge (ACDC) dataset. The model's architectures are customized to suit the CVD detection problem. Another objective was to achieve good performance from a smaller dataset.

## Literature Review

According to WHO, cardiovascular diseases caused 38% of the 17 million premature deaths that happened due to non-communicable diseases in the year 2019. People from below-average and average-income families account for at least three-quarters of all CVD deaths worldwide. Alcohol consumption, tobacco use, unhealthy food, and physical inactivity are the major behavioral risk factors for CVD. A heart attack or stroke could be the initial symptom of a more serious cardiovascular disease. Heart attack symptoms include pain in the center of the chest area and joins. Furthermore, the individual may experience vomiting, shortness of breath, and more. These are all signs of heart attack, stroke, heart failure, and other cardiac consequences which can be detected in clinics. It is known that people with CVDs and other noncommunicable illnesses in these countries have reduced access to effective and equitable health care services. As a result, many people in these nations die from CVDs and other non-communicable diseases at a younger age, sometimes in their prime working years (Martin-Isla *et al*., 2020).

Physical examination of the cardiac involves noticing the movements of the heart, auscultation of the cardiac, palpation and percussion, and the evaluation of the arterial pulse and venous pulsations. The palpation can aid in the identification of heaves and lifts thrills, impulses, and the first (S1) and second (S2) heart sounds. In inspection, the patients are required to show the whole chest to the doctor. A close inspection of the patient can reveal ventricle movements or the point of maximal impulse which helps to identify certain CVD conditions. Even if the palpation is accurate, the percussion can be used to estimate the size of the heart (Neary and Pinson, 2015; Nicholas Zakov, 2021). A doctor can listen for dullness in the heart by tapping the spaces between the ribs from the left side of the chest. Listening to all four areas of the heart such as aortic, pulmonic, tricuspid, and mitral with a stethoscope is important to record any murmurs, rubs, or gallops. A pulmonary examination can also aid in the diagnosis of cardiac diseases such as the auscultation of specific lung sounds or pleural effusions. An unusual sound detected during a pulmonary examination may indicate a negative heart condition. However, it is neither well organized nor conclusive to make a person go through all these examinations. (UCSD, 2021; Holler *et al*., 2015).

The current ultrasound technologies can provide an immediate answer to many cardiac-related problems, but with certain limitations. Based on Echocardiography (ECG), even to the professional, much of cardiovascular disease is neither observable nor precisely measurable on the physical verification. For example, irregularities around the heart valves are frequently overlooked; the pumping of oxygen-rich blood to the aorta may be significantly reduced without any observable anomaly. As a result, it is better to think of the physical checkup and the echo as a supplementary method. In contrast, with a negative cardiac physical checkup and a normal ECG, an echo need not be performed in many cases (Neary and Pinson, 2015). Breath holds test length, contrast medium reactivity, and relatively high cost are the main drawbacks of cardiac MRI sequences. New real-time MRI sequences alleviate some of these drawbacks (Saeed *et al*., 2015).

Khaleel Faieq and Mijwil (2022), researchers used Support Vector Machine (SVM) and Artificial Neural Networks (ANN) to predict heart disease using body parameter data and achieved the highest accuracy of 85.8%. The authors (Qian *et al*., 2022) used L1 regularized logistic regression (L1-LR), SVM, and AdaBoost algorithms for heart disease prediction with a cumulative incidence of 9.26%. The SVM, K-Nearest Neighbor (KNN), Decision Trees (DT), and ANN used for cardiac prediction using the heart attack dataset achieved the highest test accuracy of 85.24% with the ANN model (Pasha *et al*., 2020). Many researchers (Shorewala, 2021; Radhakrishnan *et al*., 2021; Alqahtani *et al*., 2022) have stated that ML algorithms can be used as a good tool for diagnosing CVD using statistical data. They achieved results accurately and rapidly using various ML algorithms. Meda and Bhogapathi (2022) developed the fuzzy neural-genetic algorithm-based model to classify and categorize cardiac diseases using UCI Cleveland Heart Disease (UCI) Dataset. Santhi and Renuka (2020) achieved 96% classification accuracy for five classes of cardiac using various ML algorithms using the UCI dataset. Channabasavaraju and Vinayakamurthy (2020) used Random Forest Feature Selection (RFS) strategy to extract features from the UCI dataset to improve the prediction accuracy of heart disease. Santhi and Renuka, 2020 used Cluster-based Disease Diagnosis (CDD) with different ML classifiers and UCI datasets to predict cardiac diseases (Mohan *et al*., 2020). Many other research works (Rajalakshmi and Madhav, 2019; Elsayad and Fakhr, 2015) used various ML techniques and datasets to detect CVD with limited accuracy.

A novel contingent Generative Adversarial Networks (GAN) model was proposed by Xia *et al*. (2021) to empower high-resolution imaging technology, 3 Dimensional (3D) isotropic heart Magnetic Resonance (MR) reproductions, utilizing single image stacks. Bernard *et al*. (2018) showed a detailed description of different classification and segmentation research done using the ACDC dataset in their survey published in 2018. They noticed that the highest accuracy from the available research work was 96% obtained using a Random Forest (RF) architecture and the second-highest was 92% which was obtained from both RF and SVM models. Ammar *et al*. (2021) suggested a DL network called UNet for both cardiac segmentation and diagnosis. The study was conducted on a dataset of 150 patients from Dijon Hospital, France in the context of the Medical Image Computing and Computer Assisted Intervention (MICCAI) conference 2017. The ACDC dataset with the 5 classes of the data was used for training the model. They have used an SVM model, an RF model, and a Multilayer Perceptron model (MLP) in their work which was ensembled using the soft-voting approach. The training set contained 100 patients' data, that is each pathology category with 20 samples. The test set contained 50 patients' data (10 samples for each pathology category) and they achieved 92% test accuracy.

Yang *et al*. (2018) used an ensemble of UNet-based architectures to segment cardiac elements. They trained an RF classifier and regularized MLP model and ensembled the results to predict the pathologic target class. The model achieved a cross-validation training accuracy of 94% and 92% of test accuracy. Baumgartner *et al*. (2017) investigated different 2-Dimensional (2D) and 3-

Dimensional (3D) convolutional neural network models for segmentation from the ACDC dataset. The experiments revealed that the performance of 3D U-Net models is not as good as that of 2D U-Net models. Khened *et al*. (2017) proposed a combination of Inception and DenseNet blocks for cardiac MRI image segmentation. This model achieved an accuracy of 96% with RF as the classifier. Mijwil and Shukur (2022) in their review paper summarized 20 research works using various ML and DL methods. They tabulated the techniques used and the accuracy achieved. According to them, the highest accuracy of 95% was achieved by a CNN model using intracardiac voltage time-series data.

Jang *et al*. (2017) proposed a Fully Convolutional Network (FCN) architecture based on M-Net for CVD detection from MRI images. This architecture, based on U-Net architecture, has the same layers as that of M-Net except for the 3D convolution filter. They obtained average dice scores, a measurement used to match DL output with ground truth annotation, of 94%, 89%, and 88% respectively for LV, RV, and myocardium segmentation. Patravali *et al*. (2017) developed 2D and 3D convolutional segmentation pipelines for cardiac MRI image segmentation. The Deep CNN-based models were trained on the ACDC dataset. Both 2D and 3D segmentation model architectures were adapted from U-Net and trained with a Stochastic Gradient Descent optimizer for 300 epochs. They achieved dice scores of 95%, 90%, and 86% for LV, RV, and myocardium from the 2D pipeline and 95%, 91%, and 85% for LV, RV, and myocardium respectively from the 3D pipeline. Abdeltawab *et al*. (2020) suggested a new DL framework for detecting LV function and mass quantification using the ACDC dataset. Using the grid search method, they arrived at the optimal values of the hyperparameters to obtain the best segmentation results. The proposed approach achieved a dice score of 95% for LV and 88% for the myocardium. Murugesan *et al*. (2020) proposed a context-based cross-entropy loss for U-Net and GAN-based network, Seg-Global Local GAN (Seg-GLGAN), to reduce the class imbalance problem in segmentation. The U-Net-based model achieved a dice score of 85% and Seg-GLGAN achieved a dice score of 85.6%.

Paranthaman *et al*. (2022) proposed a heart attack forecasting system using DL techniques and an MLP-based model to estimate the probability of occurring heart disease in each patient. Isensee *et al*. (2021) proposed U-Net-based 3D cardiac segmentation and classification models using an ensemble of MLP models and an RF classifier. They used the Adam optimizer to train the 3D model for 300 epochs and achieved a training accuracy of 94% and 92% of test accuracy.

We found that most of the existing DL-based works used U-Net based model with CVD images. Among them, only a few are for CVD classification. To the best of our knowledge, the highest accuracy achieved from similar existing works was 96% (Bernard *et al*., 2018; Khened *et al*., 2017).

## Materials and Methods

In this study, we propose a model to predict the occurrence of cardiovascular disease from magnetic resonance imaging images. We leverage the effectiveness of deep learning, transfer learning, and ensemble techniques to achieve optimum results.

### *The ACDC Dataset*

The ACDC dataset (Janik *et al*., 2021) was created from genuine clinical tests obtained at the University Hospital of Dijon, France. The collected data were completely anonymized and organized inside the guidelines set by the moral advisory group of the hospital. The dataset was made from 150 tests conducted on various patients. This is divided into 5 subgroups of which 4 are pathological classes and 1 for normal class. Additionally, body parameter values such as diastolic and systolic phase instances, weight, and height are also part of this dataset.

### *Data Pre-Processing*

The original slice thickness of MR images varied from 5 to 10 mm. The spatial resolution varied between 1.34 and 1.68 $mm^2$/pixel (Bernard *et al*., 2018). In the first pre-processing stage, we resized all the images into 299 × 299 × 3. The original images were 4 Dimensional (4D) in Nifti format (Janik *et al*., 2021). We used a third-party software called Medcon (Nolf *et al*., 2003) to convert images in Nifti format to Digital Imaging and Communications in Medicine (DICOM) format since it is the most common format for medical images. For example, the first frame with nine slices (1 to 9) from a 4D image of a patient, Patient number 15, is demonstrated in Fig. 1. The original image was made up of a total of 21 frames and each frame contained 9 slices. Therefore, a total of 189 slices are available in this 4D image. The center-center slice separation mechanism was used to generate DICOM images. The width of a new slice is equal to the slice spacing value of two slices ((X)MedCon|Docs/SliceSeparation). The slices are generated by manipulating their volume in the transverse, coronal, or sagittal re-slicing manner (Nolf *et al*., 2003). The last frame, frame number 21, of the same patient's 4D image is shown in Fig. 2. The differences between each slice of the same 4D image are minimal. And the slices from the same position of two different frames of the same 4D image have more similarities. For example, slice 5 from frame 1(demonstrated in Fig. 3) and slice 185 from frame 21(demonstrated in Fig. 4) look similar. Therefore, we have chosen slices randomly from the extracted frames of these 4D images, which also helped to extend the dataset size. We stored these extracted slices in DICOM format, thus creating a dataset of 5600 images in the training set, 1200 images in the validation set, and 280 images in the test set. In our work, these DICOM format images are loaded and later converted into a NumPy array for training with the DL models.

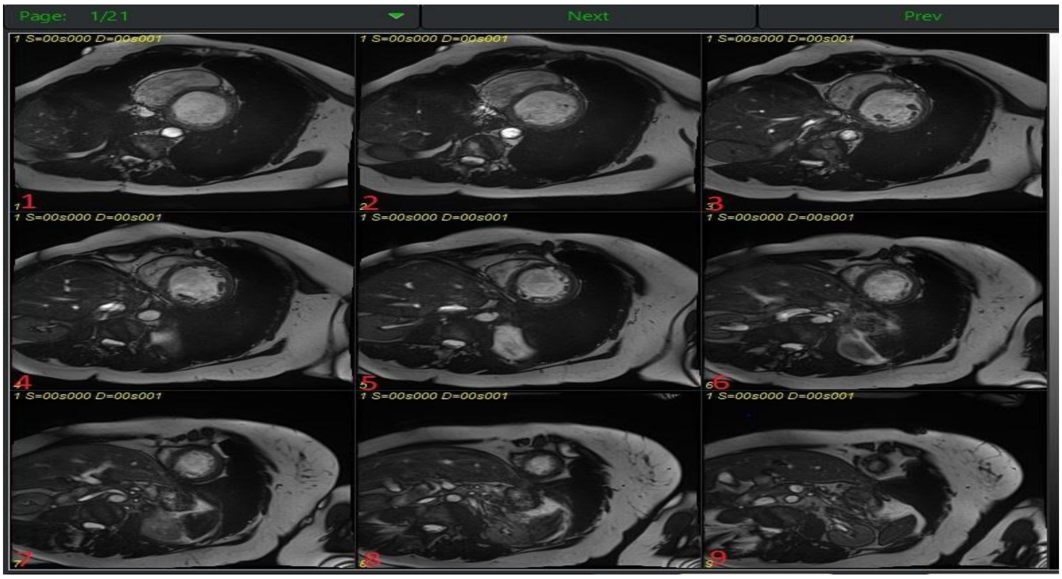



**Fig. 1:** Frame number 1 of patient number 1

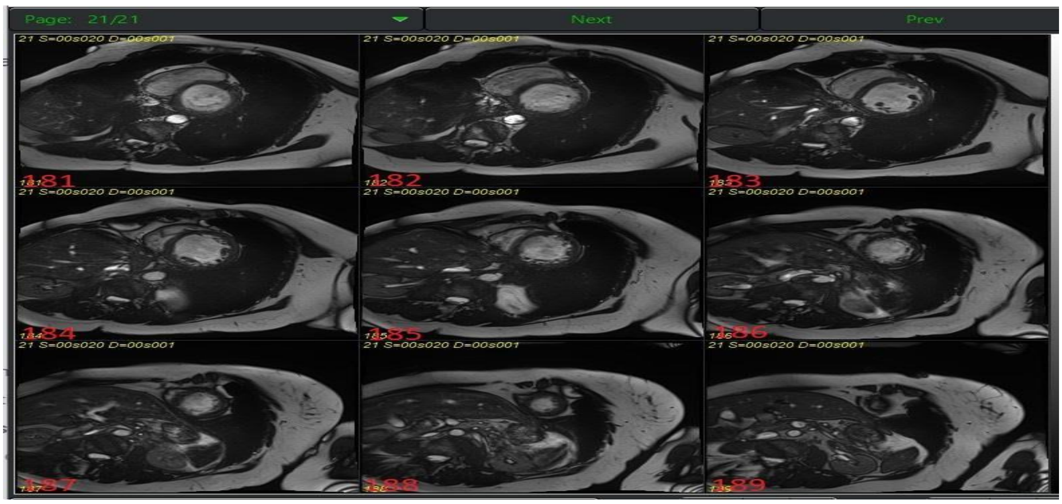



**Fig. 2:** Frame number 21 of patient number 15

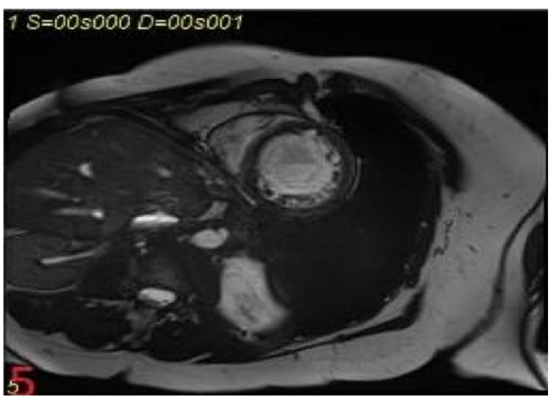



**Fig. 3:** Slice number 5 of patient number 15

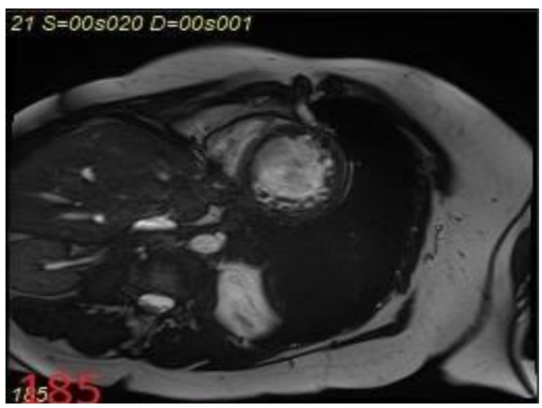



**Fig. 4:** Slice number 185 of patient number 15

The DICOM images contain details about the patient and the imaging techniques used, as displayed in Fig. 5 and 6. Such information from DICOM images helps the model extract more features easily during the training. Each DICOM image is standalone; the header contains all the information required to identify the file. Patient, study, series, and instance are the four levels of hierarchy in which this data is arranged. The person who is being examined is called a patient. The study is an imaging process that is done in the hospital at a specific date and time. There are several series in each study. A series can reflect the patient being physically scanned many times in one study or it can indicate the patient being scanned once and the data is reconstructed in various ways. Every slice of a three-dimensional image is handled as its instance. The DICOM file itself is referred to as an instance (Understanding, 2022; DICOM. How to read, write and organize… | by Alexander Weston, Ph.D. | Towards Data Science).

### VGG16 Model

The VGG-16 is one of the most used pre-trained image classification models. It was developed in Oxford University's Visual Graphics Group. This VGG16 is trained on ImageNet weights, featuring 13 convolutional layers, 5 pooling layers, and 3 fully connected layers. It has several 3×3 filters with 1 PX as a stride on each layer of convolution as shown in Fig. 7. The last layer used for classification is a SoftMax layer. The ReLU method was used as the activation function in each block. The most distinctive feature of VGG16 is that it prioritized convolution layers of a 3x3 filter with stride 1 rather than many hyper-parameters and consistently employed the same padding and maxpool layer of a 2×2 filter with stride 2. There are 64 filters in the first convolutional layer, 128 filters in the second, 256 filters in the third, and 512 filters in the fourth and fifth convolutional layers (Han *et al*., 2015) (Rezende *et al*., 2018).

### InceptionV3

The InceptionV3 model, developed by Google, has 312 layers in a total of 10 blocks. This model consists of 3 inception blocks, 13 convolutional layers, and 2 pooling layers as shown in Fig. 8. Each convolution layer contains several 3×3 filters with 2 PX as a stride. The number of output nodes in the last layer is identical to the number of categories in the dataset. In each convolution block, the SoftMax layer is used as the classification layer and ReLU as the activation function. By calculating 1×1, 3×3, and 5×5 convolutions inside the same network module, the inception module aims to serve as a multi-level feature extractor. The name of this architecture's first iteration, GoogLeNet, has now been dropped in favor of just Inception vN, where N is the version number released by Google (Wang *et al*., 2019).

### MobileNetV2

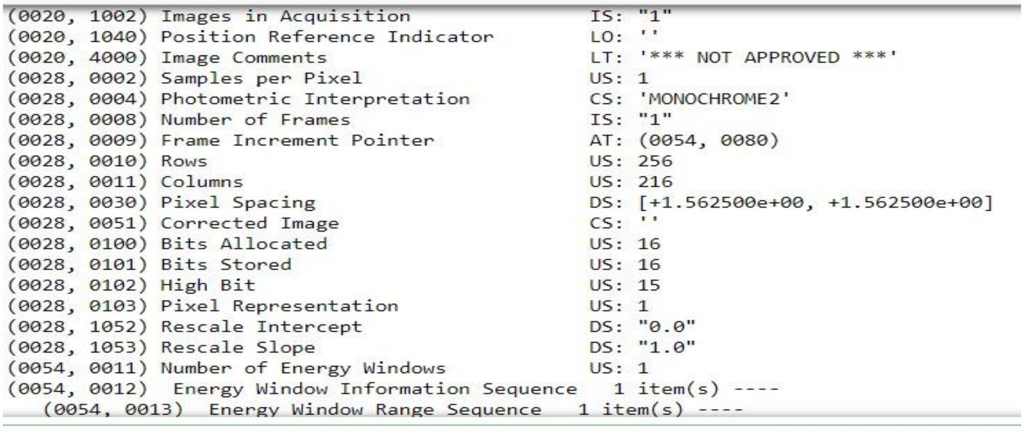
The MobileNetV2 model contains 16-layer blocks with 3×3 filters and 1 PX as a stride in each layer of convolution. This model is also developed by Google. The only distinction between MobileNet and other CNNs is the use of a thorough convolutional division, which divides the convolution into a 3×3-depth and 1x1-pointwise convolutions respectively, as shown in Fig. 9. MobileNetV2 also used SoftMax for classification and ReLU as activating function. The MobileNetV2 model contains two types of blocks. They are residual blocks with stride as 1 and downsizing blocks with stride as 2. Three types of layers are constructed for these two blocks. The ReLU6 is used in 1×1 convolution as the initial layer with non-linearity. The second layer was made for depth-wise convolution. The third layer is a 1×1 convolution with no non-linearity (Patel and Chaware, 2020).

### Transfer Learning

Transfer learning is an advanced technique in deep learning that involves training a CNN model on a similar problem to the one being solved. It is a method for feature representation from a previously trained model that saves us from having to train a new model from scratch. It helps in transferring the problem knowledge from one source to another, as shown in Fig. 10.

```
(0010, 0030) Patient's Birth Date              DA: ''
(0010, 0032) Patient's Birth Time              TM: ''
(0010, 0040) Patient's Sex                     CS: 'O'
(0010, 1020) Patient's Size                    DS: "0.0"
(0010, 1030) Patient's Weight                  DS: "0.0"
(0018, 0015) Body Part Examined                CS: 'Unknown'
(0018, 0050) Slice Thickness                   DS: "10.0"
(0018, 0070) Counts Accumulated                IS: "0"
(0018, 0088) Spacing Between Slices            DS: "10.0"
(0018, 5100) Patient Position                  CS: ''
(0020, 000d) Study Instance UID                UI: 777.777.0.0.0.1629796096.1823994544
(0020, 000e) Series Instance UID               UI: 777.777.0.0.0.1629796096.1823994544.3606581155
(0020, 0010) Study ID                          SH: 'Unknown'
(0020, 0011) Series Number                     IS: "0"
(0020, 0012) Acquisition Number                IS: "0"
(0020, 0013) Instance Number                   IS: "0"
(0020, 0052) Frame of Reference UID            UI: 777.777.0.0.0.1629796096.1823994544
(0020, 1002) Images in Acquisition             IS: "1"
(0020, 1040) Position Reference Indicator      LO: ''
```

**Fig. 5:** The extracted information about the patient from the DICOM image of patient number 15-Part 1



```
(0020, 1002) Images in Acquisition           IS: "1"
(0020, 1040) Position Reference Indicator    LO: ''
(0020, 4000) Image Comments                  LT: '*** NOT APPROVED ***'
(0028, 0002) Samples per Pixel               US: 1
(0028, 0004) Photometric Interpretation      CS: 'MONOCHROME2'
(0028, 0008) Number of Frames                IS: "1"
(0028, 0009) Frame Increment Pointer         AT: (0054, 0080)
(0028, 0010) Rows                            US: 256
(0028, 0011) Columns                         US: 216
(0028, 0030) Pixel Spacing                   DS: [+1.562500e+00, +1.562500e+00]
(0028, 0051) Corrected Image                 CS: ''
(0028, 0100) Bits Allocated                  US: 16
(0028, 0101) Bits Stored                     US: 16
(0028, 0102) High Bit                        US: 15
(0028, 0103) Pixel Representation            US: 1
(0028, 1052) Rescale Intercept               DS: "0.0"
(0028, 1053) Rescale Slope                   DS: "1.0"
(0054, 0011) Number of Energy Windows        US: 1
(0054, 0012)  Energy Window Information Sequence   1 item(s) ----
   (0054, 0013)  Energy Window Range Sequence   1 item(s) ----
```

**Fig. 6:** The extracted information about the imaging from the DICOM image of patient number 15-Part 2

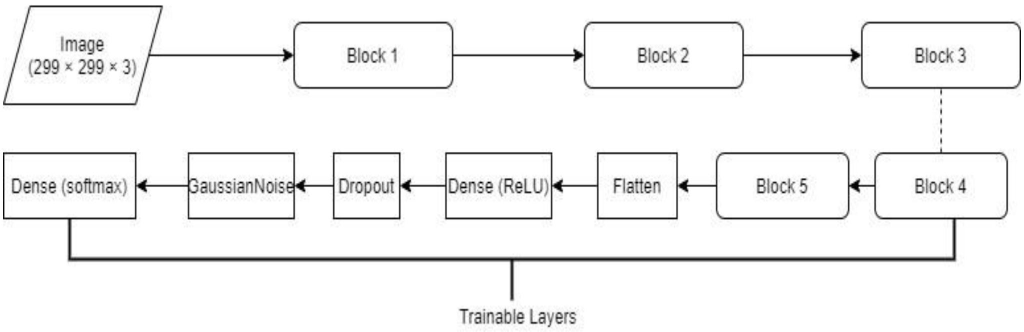



**Fig. 7:** Flow diagram of customized VGG16 model

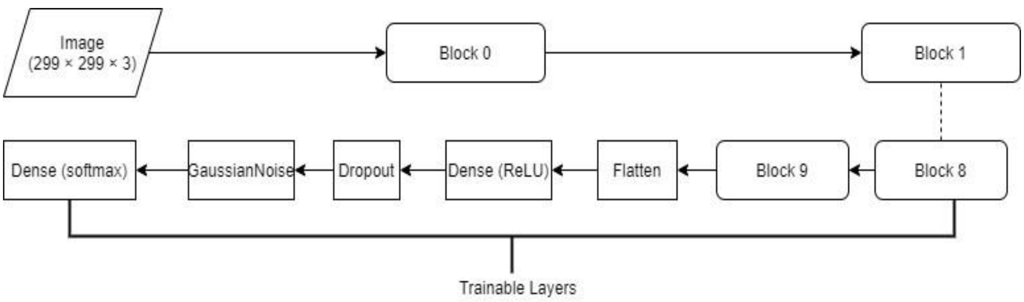



**Fig. 8:** Flow diagram of customized InceptionV3 model

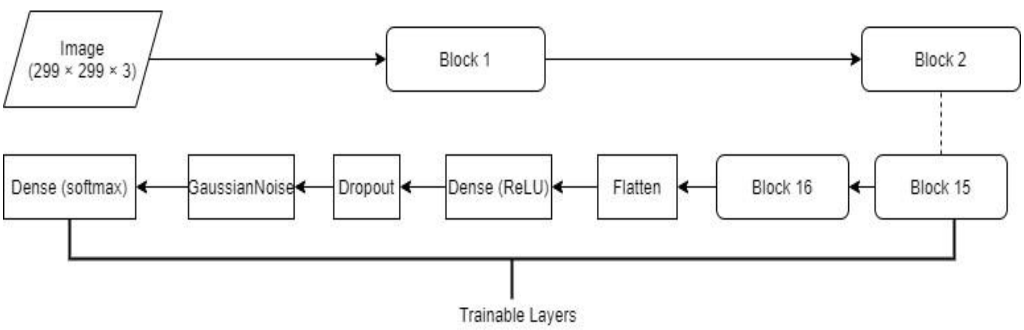



**Fig. 9:** Flow diagram of customized MobileNetV2 model

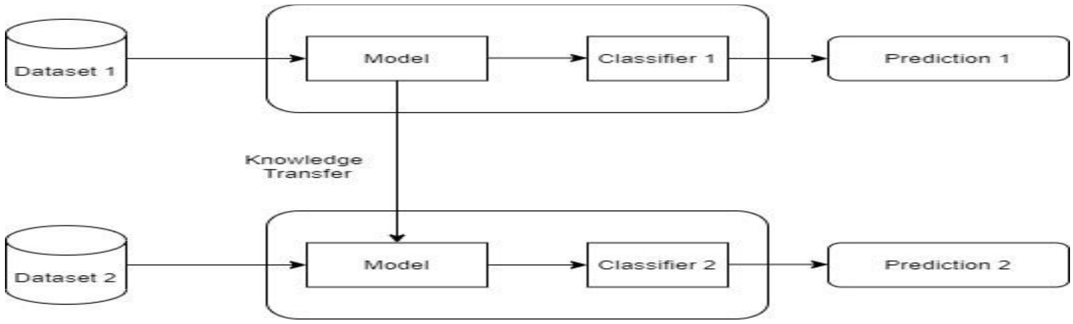



**Fig. 10:** General architecture diagram for transfer learning

A pre-trained model is typically trained on a huge dataset such as ImageNet and the weights obtained from the trained model can be used with the custom neural network for any other related problem (*Transfer Learning for Image Classification using Tensorflow |Towards Data Science*). The re-utilization of weights through this mechanism helps to effectively train a model in less time and produce output with minimum generalization error. These newly constructed models can be used directly for predictions on relatively new tasks or in training processes for related applications. The last layer contains many parameters about the original dataset since all the pre-trained models are already trained with a huge dataset. Therefore, replacing the last layer of the pre-trained models with a new classification layer is required. The model's performance depends on the similarity between the source and target data (Patel and Chaware, 2020).

### *ImageNet*

It is a massive dataset of annotated photographs designed for computer vision research work. This dataset comprises around 14 million images with over 21000 classes. In a deep learning context, the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) was specially developed for image classification problems with transfer learning. This challenge act as a benchmark in transfer learning-based image classification problems (Krizhevsky *et al*., 2017).

### *Ensemble Methods*

In general, ensemble learning is training many networks on the same dataset, then utilizing each of the trained models to predict and aggregate the predictions in some fashion to provide an outcome or prediction (Opitz and Maclin, 1999). In this proposed work, we used hard-voting and soft-voting ensemble approaches, and more detailed descriptions of the used ensemble techniques are given below.

### *I. Hard-Voting*

Hard voting is an ensemble technique in which the class that gets the maximum votes from all classifiers is chosen. Just assuming classifiers are independent, the ensemble will perform better compared to the individual low-performing classifiers. However, they are trained on the same data. The final prediction in hard voting is made through a majority vote in which the aggregator chooses the class prediction that appears repeatedly among the base models (Kumari *et al*., 2021). This approach is appropriate for models that predict separate class labels when classifier outputs are not independent, as well as for binary class issues where the number of included classifiers is not odd. It uses the predictions from each classifier as input and then computes the votes for each target label. The prediction/result of the hard voting ensemble model is the label with the majority of votes after this computation (Peppes *et al*., 2021). For example, Predictions:

Classifier 1 predicts class A
Classifier 2 predicts class B
Classifier 3 predicts class B
Classifier 2 and Classifier 3 predict class B, so class B is the ensemble decision.

### *II. Soft-Voting*

Soft-voting is an ensemble technique in which the class that gets the most votes based on the probability score from each classifier is chosen. If all classifiers can assess the probability of classes using a function, then find the middle value of singular classifiers. The soft voting approach regularly performs better compared to the hard-voting approach.

Base models in soft voting should use the probability technique. Because it incorporates the predictions of several models, the voting classifier produces better overall results than other base models (Kumari *et al*., 2021). It is generally implemented when the built models predict probabilities for each included class, as well as when it is evident that a classifier produces superior output results than the other classifiers included (Peppes *et al*., 2021). Prediction is almost similar to the previous example, but use only class B as it is a binary classification problem:

Classifier 1 predicts class B with a probability of p1
Classifier 2 predicts class B with a probability of p2

Classifier 3 predicts class A with a probability p3 So, the ensemble model will predict class B with probability
p = (p1 + p2 + (100-p3))/3

### *Proposed Approach*

The outline of our work is shown in Fig. 11. It consists of the following steps: Acquisition of the ACDC dataset, data extraction, data pre-processing, data augmentation, choosing pre-trained models for transfer learning, feature extraction, and classification using the VGG16, InceptionV3, and MobileNetV2 models and ensemble using soft-voting and hard-voting methods. The converted ACDC dataset images were in the RGB format within the pixel range of [0,255]. In the pre-processing phase, all those images were rescaled into the range [0,1] as per the pre-trained model's requirements. In transfer learning, the classification layer of pre-trained models may not be helpful for the new classification task. Therefore, we replaced it with a fully connected layer at every model's top layer. The pre-trained network was kept frozen and thus only the weights of the top four layers and one classifier layer were modified during the training. In this proposed approach, a SoftMax layer was used on top of all the models as a classification layer. For training the first 10 epochs, an Adam optimizer with a learning rate of 0.0001 was used. We used a different optimizer called RMSprop from epoch 10 to 100 with a learning rate of 0.00001. Using a different optimizer suddenly boosted the model performance. We considered different parameters such as the number of trainable layers including original and extra added layers, epochs, learning rate, and optimizers for the fine-tuning process of all the models. Table 1 demonstrates the optimal hyper-parameters used in different experiments.

### *Over-Fitting Prevention Techniques*

The ACDC dataset contains scanned Magnetic Resonance Images (MRI) of hearts. Meanwhile, the ImageNet dataset does not contain similar image classes. Due to this dissimilarity between the source and the target dataset images, our model started overfitting during the training. We applied various techniques to overcome this issue. The most used technique to avoid over-fitting is applying data augmentation. Data Augmentation is used to increase dataset size by applying some geometrical techniques to the original dataset. It helps to create so many similar images and reduce overfitting (Pavanelli Vianna, 2018; Analytics, 2022). We generated 75000 augmented images from 5600 original images with the help of the Keras ImageDataGenerator method using different geometric techniques such as rotation, width shifting, zooming, horizontal flipping, and brightness changes.

The dropout algorithm can be used as a solution to over-fitting by which the model's performance is enhanced by altering neurons randomly on every iteration. A dropout map with the same neuron size is initialized randomly to mark the on or off state of the corresponding neuron in each iteration. During training iterations, the neurons with off status are removed from the network. The activation signal is disabled for forward propagation and the error signal for backward neuron propagation. All neurons are enabled during testing, but the activation signal is decreased to average turn-up during training (Garbin *et al*., 2020). The dropout rate of 0.5 gave optimum results in our experiments.

Weight regularization methods such as weight decay restrict the loss function as a neural network is trained to use small weights. A reliable model and less likely to overfit the training dataset could lead to smaller weights within a neural network, resulting in optimal results in predictions. A weight constraint is a trigger that controls the weight, size, or size of the weights and scales them all below a defined threshold (MLM, 2022). We applied Unit Norm weight constraints into the model to reduce over-fitting. It forced weights into a magnitude of 1.0. Then, we used white noise which is a special case of Gaussian noise with a standard deviation of 0.1 for the input samples to enhance the model's stability.

We also used the early stopping method (*Use Early Stopping to Halt the Training of Neural Networks at the Right Time*) (Venu, 2020) to solve for over-fitting. Early stopping is a kind of regularization in which a model is trained using an iterative method, like gradient descent. It suggests how many iterations can be performed before the overfitting starts. We applied the validation accuracy with the patience of six epochs as the monitoring measurement in this method. Therefore, it automatically stops the training if there is no improvement in validation accuracy in every six epochs.

We trained the customized VGG16, MobileNetV2, and InceptionV3 models for 100 epochs using 5600 original images and 75000 augmented images.1200 images were used for validation and 280 images for testing. Keras libraries with Python 3.7.6 was used to build the proposed model since it provides access to load the pre-trained models on the ImageNet dataset directly and more convolutional layers can be easily added to the model. Python and TensorFlow were used on an Ubuntu machine with GTX 1080 GPU, $12^{th}$ generation i7 @ 4.600GHz CPU, and 64359 MiB RAM.

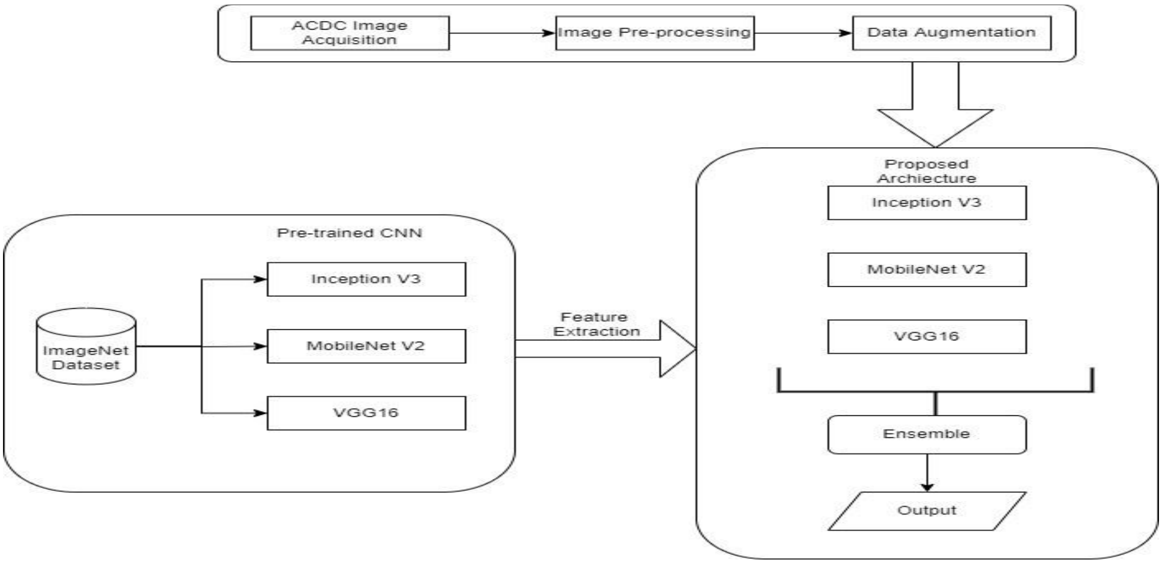



**Fig. 11:** Flow diagram of the proposed work

**Table 1:** Fine-tuning for each model

| Model | NL | NAL | NTL | Epochs | Learning rate | Optimizer |
|---|---|---|---|---|---|---|
| VGG16 | 18 | 5 | 13 | 1-100 | 0.00010 | Adam |
| | | | | 10-100 | 0.00001 | RMSprop |
| InceptionV3 | 310 | 5 | 64 | 1-100 | 0.00010 | Adam |
| | | | | 10-100 | 0.00001 | RMSprop |
| MobileNetV2 | 153 | 5 | 6 | 1-100 | 0.00010 | Adam |
| | | | | 10-100 | 0.00001 | RMSprop |

*NL: Number of original Layers, *NAL: Number of extra Added Layers, * NTL; the number of Trained Layers (original + customized)

## Results and Discussion

We evaluated the performance of the model using different accuracy measurements. The training and validation accuracies were measured to assure that the model has gained enough knowledge through the training and to ensure that the over-fitting is minimal. The testing accuracy was calculated after the completion of training by using 280 test images. For all these measurements we calculated the average training and validation accuracies of the last 20 epochs. Two different ensemble approaches such as soft voting and hard voting were applied to the results of individual models. After these ensembles, we achieved 97.94% test accuracy from hard voting and 98.08% from soft voting. The model has taken 16 hours to complete 100 epochs of training.

Accuracy is the most used metric to evaluate the performance of DL models. It is the proportion of the correct number of predictions to the total number of inputs used:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions\ made} \tag{1}$$

The training graphs for VGG16, MobileNetV2, and InceptionV3 models are given in Fig. 12, 13, and 14 respectively. It contains the accuracy and loss measurements that are used to evaluate the model's performance on both training and validation datasets. The graphs clearly show that there is no over-fitting during the model's training. Both the training and validation accuracies gradually increased for the epochs. In the initial stage of the training, there was some fluctuation in both the accuracies which became steady as the training progressed. It is also observed that the loss measurements gradually decreased until the last epoch. We can conclude from this analysis that the model achieved efficient results without any overfitting.

Table 2 gives the results from individual models and hard-voting ensembles. Concerning epochs, the accuracies kept on increasing in all the experiments. After 30 epochs of training, the MobileNetV2 model achieved the highest test accuracy 83.21% meanwhile InceptionV3 and VGG16 achieved test accuracies of 81.12% and 81.6% respectively. But the hard-voting ensemble accuracy reached 84.62% at this point which was higher than the test accuracy from MobileNetV2.

The MobileNetV2 model also showed the highest training accuracy of 85.17% and validation accuracy of 84.34% during this period. The difference between the training and validation score of all three models was just around 1% only which ensured no overfitting during the model's training. The MobileNetV2 model again showed the highest training, validation, and test accuracies with 92.54%, 91.07%, and 90.21% respectively after 60 epochs of training. But the ensemble score of 91.07% was a little higher than the MobileNetV2 model's test accuracy. The VGG16 has achieved the highest test accuracy of 97.03% and validation accuracy of 96.92% among all the individual models and the least accuracy was 96.09% from MobileNetV2 in 100 epochs. The InceptionV3 model resulted in the highest training accuracy of 97.78%. The gap between training and validation score from all individual models were just less than 2% only which ensured all models trained well without any overfitting issue. An ensemble using hard voting of all these three models gave an accuracy of 97.94%.

The results from individual models and soft-voting ensembles are given in Table 3. The accuracies kept on increasing with epochs in all four experiments. The MobileNetV2 model had the best test accuracy of 83.88% after 30 training epochs, while InceptionV3 and VGG16 had test accuracies of 81.39% and 81.57%, respectively. However, the soft voting ensemble accuracy at that time was 84.79%, which was more than the MobileNetV2 test accuracy. The MobileNetV2 model also had the best training and validation accuracy, at 85.23% and 84.42%, respectively. All three models' differences between training and validation scores were less than 2%, showing no overfitting during model training. After 60 training epochs, the MobileNetV2 model once more demonstrated the best test accuracy of 90.21%. However, compared to the MobileNetV2 model, the ensemble score of 91.07% was somewhat higher. The VGG16 showed the highest training accuracy of 92.89% and validation accuracy of 92.81% at that time. The VGG16 has achieved the highest test accuracy of 97.17% and validation accuracy of 96.84% among all the individual models and the least accuracy was 95.62% from InceptionV3 in 100 epochs. The MobileNetV2 model resulted in the highest training accuracy of 98.12%. Soft-voting ensemble of all the three models gave an accuracy of 98.08% which was the highest so far. All individual models' gaps between training and validation scores were about 2% or less, ensuring that there were no overfitting issues.

The CVD classification accuracies of some of the existing works and our proposed work are given in Table 4. Available literature shows that the best classification accuracy achieved using the ACDC dataset is 96%. Our work achieved better test accuracy than other existing works in classifying cardiovascular disease from the ACDC dataset.

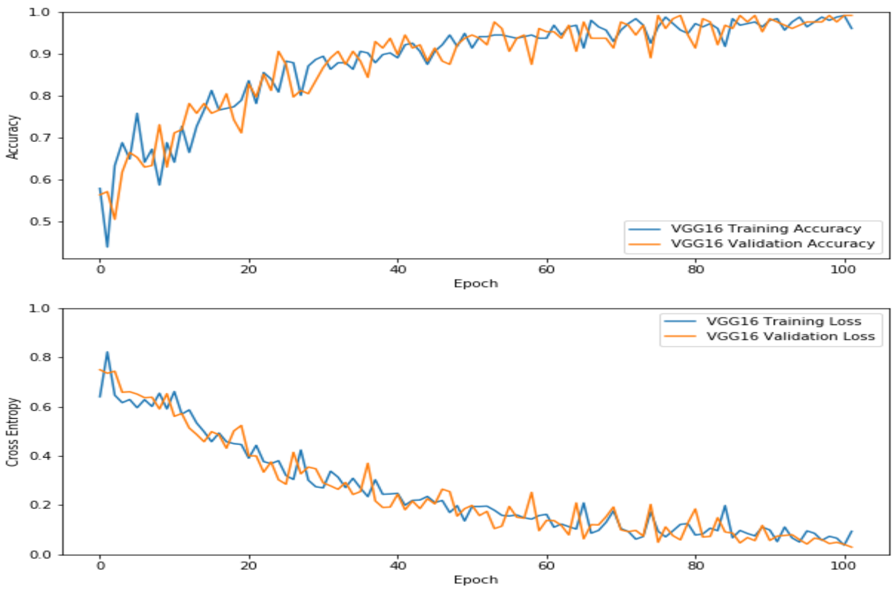



**Fig. 12:** Training graph of the VGG16 model

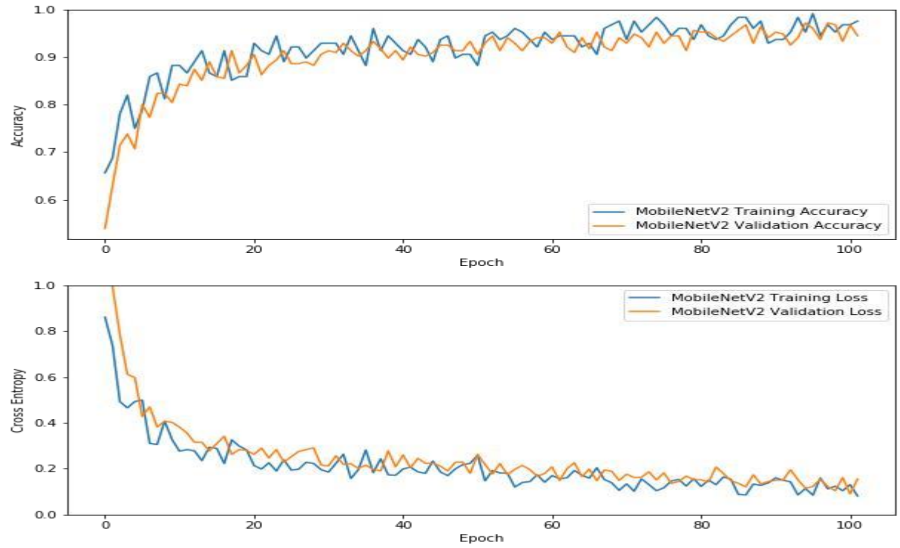



**Fig. 13:** Training graph of the MobileNetV2 model

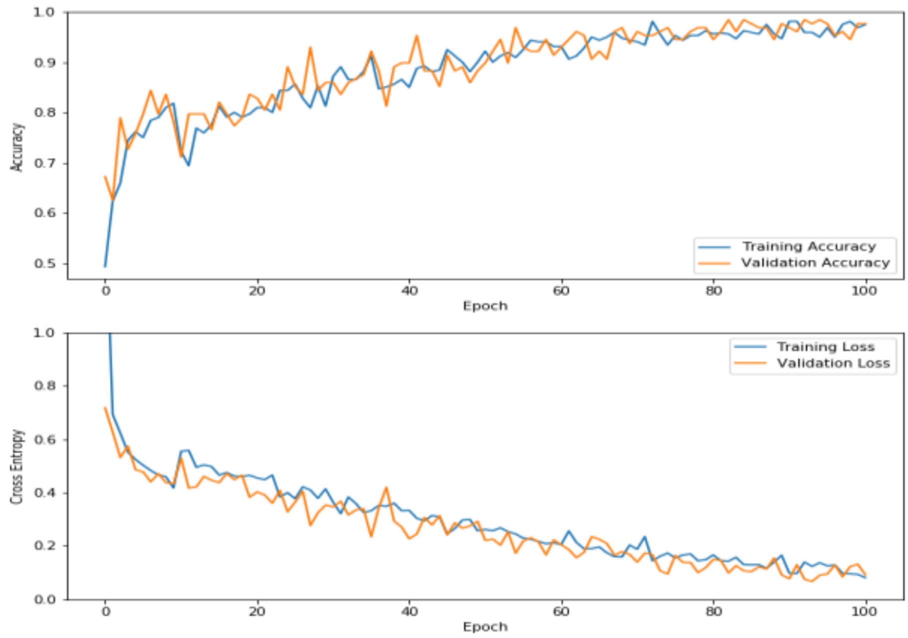



**Fig. 14:** Training graph of the InceptionV3 model

**Table 2:** Training result of ensemble with hard-voting

| Model | Epochs | Training accuracy (%) | Validation accuracy (%) | Test accuracy (%) |
|---|---|---|---|---|
| MobileNetV2 | 30 | 85.17 | 84.34 | 83.21 |
| | 60 | 92.54 | 91.07 | 90.21 |
| | 100 | 96.35 | 94.67 | 96.09 |
| InceptionV3 | 30 | 82.03 | 81.09 | 81.12 |
| | 60 | 91.33 | 90.19 | 89.32 |
| | 100 | 97.78 | 95.82 | 96.27 |
| VGG16 | 30 | 83.16 | 81.18 | 81.68 |
| | 60 | 92.01 | 91.56 | 90.16 |
| | 100 | 97.45 | 96.92 | 97.03 |
| Ensemble | 30 | | | 84.62 |
| | 60 | | | 91.07 |
| | 100 | | | 97.94 |

**Table 3:** Training result of the ensemble with soft-voting

| Model | Epochs | Training accuracy (%) | Validation accuracy (%) | Test accuracy (%) |
|---|---|---|---|---|
| MobileNetV2 | 30 | 85.23 | 84.42 | 83.88 |
| | 60 | 92.19 | 91.14 | 90.29 |
| | 100 | 98.12 | 96.56 | 96.83 |
| InceptionV3 | 30 | 82.11 | 81.27 | 81.39 |
| | 60 | 91.70 | 90.26 | 89.46 |
| | 100 | 97.91 | 95.76 | 95.62 |
| VGG16 | 30 | 83.14 | 81.12 | 81.57 |
| | 60 | 92.89 | 91.81 | 89.93 |
| | 100 | 97.89 | 96.84 | 97.17 |
| Ensemble | 30 | | | 84.79 |
| | 60 | | | 91.18 |
| | 100 | | | 98.08 |

**Table 4:** Classification results on the ACDC dataset

| Work | Model | Test accuracy |
|---|---|---|
| Ammar *et al*. (2021) | U-Net | 92% |
| Isensee *et al*. (2021) | U-Net | 92% |
| Khened *et al*. (2017) | CNN | 96% |
| Bernard *et al*. (2018) | Random forest | 96% |
| Our Model 1 | soft-voting | 97.94% |
| Our Model 2 | hard-voting | 98.08% |

## Conclusion

This study proposes and implements hard-voting and soft-voting ensemble techniques for cardiovascular disease detection from MRI images. It used VGG16, InceptionV3, and MobileNetV2 deep learning models that are already trained using the ImageNet dataset. The main contribution of this study is improved classification accuracy from both the ensemble voting techniques. Another contribution is the use of a limited dataset since transfer learning is adopted. The obtained results prove the effectiveness of our proposed method by giving an average accuracy of 97.94% from hard voting and 98.08% from soft voting from the ACDC dataset. We can conclude that this approach can be used to assist medical practitioners in diagnosing cardiovascular diseases. In the future, we plan to deploy this technique on real-world applications to monitor, detect and classify cardiovascular diseases automatically.

## Acknowledgment

This research was carried out at CHRIST (Deemed to be University), Bengaluru, India and we would like to extend our gratitude to the institution for providing us with the necessary computing resources and guidance.

## Authors Contributions

**Sibu Cyriac:** Problem identification, design, acquisition of data, result analysis, manuscript writing, and review.

**Sivakumar R.:** Problem identification, design, result analysis, manuscript review.

**Nidhin Raju:** Problem identification, design, acquisition of data, experiments, manuscript writing.

## Ethics

We would like to inform that the information stated here is genuine to the best of our knowledge. There are no ethical issues involved in this manuscript.

## References

Abdeltawab, H., Khalifa, F., Taher, F., Alghamdi, N. S., Ghazal, M., Beache, G., Mohamed, T., Keynton, R., & El-Baz, A. (2020). A deep learning-based approach for automatic segmentation and quantification of the left ventricle from cardiac cine MR images. *Computerized Medical Imaging and Graphics*, 81, 101717. https://doi.org/10.1016/j.compmedimag.2020.101717

AbuKhousa, E., & Campbell, P. (2012). Predictive data mining to support clinical decisions: An overview of heart disease prediction systems. *2012 International Conference on Innovations in Information Technology, IIT 2012*, 267–272. https://doi.org/10.1109/innovations.2012.6207745

Alqahtani, A., Alsubai, S., Sha, M., Vilcekova, L., & Javed, T. (2022). Cardiovascular Disease Detection using Ensemble Learning. *Computational Intelligence and Neuroscience*, 2022, 1–9. https://doi.org/10.1155/2022/5267498

Ammar, A., Bouattane, O., & Youssfi, M. (2021). Automatic cardiac cine MRI segmentation and heart disease classification. *Computerized Medical Imaging and Graphics*, 88, 101864. https://doi.org/10.1016/J.compmedimag.2021.101864

Asfi-Ar, M., Asif, R., Nishat, M. M., Faisal, F., & Hasan Udoy, M. (2021). Performance Evaluation and Comparative Analysis of Different Machine Learning Algorithms in Predicting Cardiovascular Disease Spectroscopy on Biological Tissue View project Analysis and Design of Optimized Controller for Power Converters by Nonlinear Optimization Techniques View project. https://www.researchgate.net/publication/3517061 15

Analytics, V. (2022). Guide to Prevent Overfitting in Neural Networks-Analytics Vidhya.

Baumgartner, C. F., Koch, L. M., Pollefeys, M., & Konukoglu, E. (2017). An Exploration of 2D and 3D Deep Learning Techniques for Cardiac MR Image Segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), 10663 LNCS, 111–119. https://arxiv.org/abs/1709.04496v2

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P. A., Cetin, I., Lekadir, K., Camara, O., Gonzalez Ballester, M. A., Sanroma, G., Napel, S., Petersen, S., Tziritas, G., Grinias, E., Khened, M., Kollerathu, V. A., Krishnamurthi, G., Rohe, M. M., … Jodoin, P. M. (2018). Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? *IEEE Transactions on Medical Imaging*, 37(11), 2514–2525. https://doi.org/10.1109/TMI.2018.2837502

Bizopoulos, P., & Koutsouris, D. (2019). Deep Learning in Cardiology. *IEEE Reviews in Biomedical Engineering*, 12, 168–193. https://doi.org/10.1109/RBME.2018.2885714

Brewer, L. P. C., Svatikova, A., & Mulvagh, S. L. (2015). The Challenges of Prevention, Diagnosis and Treatment of Ischemic Heart Disease in Women. *Cardiovascular Drugs and Therapy*, 29(4), 355. https://doi.org/10.1007/S10557-015-6607-4

Dinh, A., Miertschin, S., Young, A., & Mohanty, S. D. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC Medical Informatics and Decision Making*, 19(1), 1–15. https://doi.org/10.1186/S12911-019-0918-5/FIGURES/8

Channabasavaraju, B. D., & Vinayakamurthy, U. (2020). An Analysis of Heart Disease for Diabetic Patients Using Recursive Feature Elimination with Random Forest. *Journal of Computer Science*. https://doi.org/10.3844/jcssp.2020.105.116

Elsayad, A., & Fakhr, M. (2015). Diagnosis of Cardiovascular Diseases with Bayesian Classifiers. *Journal of Computer Science*, 11(2), 274–282. https://doi.org/10.3844/JCSSP.2015.274.282

Farag, A., & Fakhreldin, M. (2012). Heart Localization from Magnetic Resonance Images Sequence. *Journal of Computer Science*, 8(4), 499–505. https://doi.org/10.3844/JCSSP.2012.499.505

Friedman, J. H. (2017). The Elements of Statistical Learning: Data Mining, Inference and Prediction. http://dlib.hust.edu.vn/handle/HUST/19661

Garbin, C., Zhu, · Xingquan, Marques, · Oge, & Zhu, X. (2020). Dropout vs. batch normalization: An empirical study of their impact to deep learning. *Multimedia Tools and Applications*, 79, 12777- 12815. https://doi.org/10.1007/s11042-019-08453-9

Han, S., Mao, H., & Dally, W. J. (2015). Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding. 4th International Conference on Learning Representations, ICLR 2016-Conference Track Proceedings. https://arxiv.org/abs/1510.00149v5

Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., Sun, R., & Garciá-Magarinõ, I. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Information Systems, 2018. https://doi.org/10.1155/2018/3860146

Hayes, S. M., Dupuis, M., & Murray, S. (2008). Issues and challenges in the assessment, diagnosis and treatment of cardiovascular risk factors: Assessing the needs of cardiologists. *BMC Medical Education*, 8, 30. https://doi.org/10.1186/1472-6920-8-30

Holler, J. G., Bech, C. N., Henriksen, D. P., Mikkelsen, S., Pedersen, C., & Lassen, A. T. (2015). Nontraumatic hypotension and shock in the emergency department and the prehospital setting, prevalence, etiology and mortality: A systematic review. *PLoS ONE*, 10(3). https://doi.org/10.1371/JOURNAL.PONE.0119331

Isensee, F., Jaeger, P. F., Full, P. M., Wolf, I., Engelhardt, S., & Maier-Hein, K. H. (last visited June 2021). Automatic Cardiac Disease Assessment on cine-MRI via Time-Series Segmentation and Domain Specific Features.

Jang, Y., Hong, Y., Ha, S., Kim, S., & Chang, H.-J. (2017). Automatic Segmentation of LV and RV in Cardiac MRI. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), 10663 LNCS, 161–169. https://doi.org/10.1007/978-3-319- 75541-0_17

Janik, A., Dodd, J., Ifrim, G., Sankaran, K., & Curran, K. M. (2021). Interpretability of a deep learning model in the application of cardiac MRI segmentation with an ACDC challenge dataset. 11596, 861–872. https://doi.org/10.1117/12.2582227

Khaleel Faieq, A., & Mijwil, M. M. (2022). Prediction of of heart diseases utilising support vector machine and artificial neural network. *Indonesian Journal of Electrical Engineering and Computer Science*, 26(1), 374–380. https://doi.org/10.11591/IJEECS.V26.I1.PP374-380

Khened, M., Alex, V., & Krishnamurthi, G. (2017). Densely Connected Fully Convolutional Network for Short-Axis Cardiac Cine MR Image Segmentation and Heart Diagnosis Using Random Forest. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), 10663 LNCS, 140–151. https://doi.org/10.1007/978-3-319- 75541-0_15

Krittanawong, C., Rogers, A. J., Johnson, K. W., Wang, Z., Turakhia, M. P., Halperin, J. L., & Narayan, S. M. (2020). Integration of novel monitoring devices with machine learning technology for scalable cardiovascular management. *Nature Reviews Cardiology*, 18(2), 75–91. https://doi.org/10.1038/s41569-020-00445-9

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. https://doi.org/10.1145/3065386

Kumari, S., Kumar, D., & Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. *International Journal of Cognitive Computing in Engineering*, 2, 40–46. https://doi.org/10.1016/J.IJCCE.2021.01.001

MLM. (2022). How to Reduce Overfitting Using Weight Constraints in Keras.

Martin-Isla, C., Campello, V. M., Izquierdo, C., RaisiEstabragh, Z., Baeßler, B., Petersen, S. E., & Lekadir, K. (2020). Image-Based Cardiac Diagnosis with Machine Learning: A Review. *Frontiers in Cardiovascular Medicine*, 0, 1. https://doi.org/10.3389/FCVM.2020.00001

Meda, S., & Bhogapathi, R. (2022). An Integrated Machine Learning Model for Heart Disease Classification and Categorization. *Journal of Computer Sciences.* https://doi.org/10.3844/jcssp.2022.257.265

Mijwil, M., & Shukur, B. S. (2022). A Scoping Review of Machine Learning Techniques and Their Utilisation in Predicting Heart Diseases. *Ibn AL-Haitham Journal for Pure and Applied Sciences*, 35(3), 175–189. https://doi.org/10.30526/35.3.2813

Mohan, S., Thirumalai, C., & Rababah, A. (2020). Classification of Heart Disease Using Cluster Based DT Learning. *Journal of Computer Science*. https://doi.org/10.3844/jcssp.2020.50.55

Muhammad, Y., Tahir, M., Hayat, M., & Chong, K. T. (2020). Early and accurate detection and diagnosis of heart disease using intelligent computational model. *Scientific Reports*, 10(1), 1–17. https://doi.org/10.1038/s41598-020-76635-9

Murugesan, B., Sarveswaran, K., Raghavan, V. S., Shankaranarayana, S. M., Ram, K., & Sivaprakasam, M. (2020). A Context Based Deep Learning Approach for Unbalanced Medical Image Segmentation. *Proceedings-International Symposiumon Biomedical Imaging*, 1949–1953. https://doi.org/10.1109/ISBI45749.2020.9098597

Muthuvel, M., Marimuthu, M., Abinaya, M., Hariesh, K. S., & Madhankumar, K. (2018). A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach Internet of Things View project Biometrics View project A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach. *Article in International Journal of Computer Applications*, 181(18), 975-8887. https://doi.org/10.5120/ijca2018917863

Marsland, S. (2011). *Machine learning: an algorithmic perspective*. Chapman and Hall/CRC.

Nahar, J., Imam, T., Tickle, K. S., & Chen, Y. P. P. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. *Expert Systems with Applications*, 40(4), 1086–1093. https://doi.org/10.1016/J.ESWA.2012.08.028

Neary, A., & Pinson, S. (2015). Cardiovascular examination in emergency departments. *Emergency Nurse*, 23(1), 32–37. https://doi.org/10.7748/EN.23.1.32.E1348

Nicholas Zakov, Z. (last visited May 2021). Book Reviews. www.ccjm.org

Nolf, E., Voet, T., Jacobs, F., Dierckx, R. A., & Lemahieu, I. (2003). (X)MedCon * An Open-Source Medical Image Conversion Toolkit. http://members.aol.com/royalef/gif89a.txt

Opitz, D., & Maclin, R. (1999). Popular Ensemble Methods: An Empirical Study. *Journal of Artificial Intelligence Research*, 11, 169–198. https://doi.org/10.1613/JAIR.614

Ordonez, C., Omiecinski, E., De Braal, L., Santana, C. A., Ezquerra, N., Taboada, J. A., Cooke, D., Krawczynska, E., & Garcia, E. V. (2001). Mining constrained association rules to predict heart disease. Proceedings-IEEE International Conference on Data Mining, *ICDM*, 433–440. https://doi.org/10.1109/ICDM.2001.989549

Organization, W. H. (2007). Prevention of cardiovascular disease: Guidelines for assessment and management of total cardiovascular risk. In Prevention of cardiovascular disease: Guidelines for assessment and management of total cardiovascular risk.

Paranthaman, M., Yaathash, B., Santhosh, S., & Sanjairam, M. (2022). Cardiovascular Disease Prediction using Deep Learning. *2022 6th International Conference on Trends in Electronics and Informatics*, *ICOEI 2022-Proceedings*, 1399– 1404. https://doi.org/10.1109/ICOEI53556.2022.9777135

Pasha, S. N., Ramesh, D., Mohmmad, S., Harshavardhan, A., & Shabana. (2020). Cardiovascular disease prediction using deep learning techniques. *IOP Conference Series: Materials Science and Engineering*, 981(2), 022006. https://doi.org/10.1088/1757-899X/981/2/022006

Patel, R., & Chaware, A. (2020). Transfer learning with fine-tuned MobileNetV2 for diabetic retinopathy. 2020 International Conference for Emerging Technology, *INCET 2020*. https://doi.org/10.1109/INCET49848.2020.9154014

Patravali, J., Jain, S., & Chilamkurthy, S. (2017). 2D-3D Fully Convolutional Neural Networks for Cardiac MR Segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), 10663 LNCS, 130–139. https://arxiv.org/abs/1707.09813v1

Pavanelli Vianna, V. (2018). Study and development of a Computer-Aided Diagnosis system for classification of chest x-ray images using convolutional neural networks pre-trained for ImageNet and data augmentation. https://pytorch.org/docs/master/torchvision/

Peppes, N., Daskalakis, E., Alexakis, T., Adamopoulou, E., & Demestichas, K. (2021). Performance of Machine Learning-Based Multi-Model Voting Ensemble Methods for Network Threat Detection in Agriculture 4.0. Sensors, 21(22), 7475. https://doi.org/10.3390/S21227475

Piña, I. L., Apstein, C. S., Balady, G. J., Belardinelli, R., Chaitman, B. R., Duscha, B. D., Fletcher, B. J., Fleg, J. L., Myers, J. N., & Sullivan, M. J. (2003). Exercise and Heart Failure. *Circulation*, 107(8), 1210–1225. https://doi.org/10.1161/01.CIR.0000055013.92097.40

Qian, X., Li, Y., Zhang, X., Guo, H., He, J., Wang, X., Yan, Y., Ma, J., Ma, R., & Guo, S. (2022). A Cardiovascular Disease Prediction Model Based on Routine Physical Examination Indicators Using Machine Learning Methods: A Cohort Study. *Frontiers in Cardiovascular Medicine*, 0, 1535. https://doi.org/10.3389/FCVM.2022.854287

Radhakrishnan, P., Ramaiyan, K., Vinayagam, A., & Veerasamy, V. (2021). A stacking ensemble classification model for detection and classification of power quality disturbances in PV integrated power network. *Measurement*, 175, 109025. https://doi.org/10.1016/J.measurement.2021.109025

Rajalakshmi, S., & Madhav, K. V. (2019). A Collaborative Prediction of Presence of Arrhythmia in Human Heart with Electrocardiogram Data using Machine Learning Algorithms with Analytics. *Journal of Computer Science*, 15(2), 278–287. https://doi.org/10.3844/JCSSP.2019.278.287

Rezende, E., Ruppert, G., Carvalho, T., Theophilo, A., Ramos, F., & de Geus, P. (2018). Malicious Software Classification Using VGG16 Deep Neural Network's Bottleneck Features. *Advances in Intelligent Systems and Computing*, 738, 51–59. https://doi.org/10.1007/978-3-319-77028-4_9

Saeed, M., Van, T. A., Krug, R., Hetts, S. W., & Wilson, M. W. (2015). Cardiac MR imaging: Current status and future direction. *Cardiovascular Diagnosis and Therapy*, 5(4), 290. https://doi.org/10.3978/J.ISSN.2223- 3652.2015.06.07

Santhi, B., & Renuka, K. (2020). Study and Analysis of Prediction Model for Heart Disease Data Using Machine Learning Techniques. *Journal of Computer Science*. https://doi.org/10.3844/jcssp.2020.344.354

Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding Machine Learning: From Theory to Algorithms. *Understanding Machine Learning: From Theory to Algorithms*, 9781107057135, 1-397. https://doi.org/10.1017/CBO9781107298019

Shivamurthy, V. K. N., Tahari, A. K., Marcus, C., & Subramaniam, R. M. (2014). Brain FDG PET and the Diagnosis of Dementia, 204(1), W76–W85. https://doi.org/10.2214/AJR.13.12363

Shorewala, V. (2021). Early detection of coronary heart disease using ensemble techniques. *Informatics in Medicine Unlocked*, 26, 100655. https://doi.org/10.1016/J.IMU.2021.100655

Shu, T., Zhang, B., & Tang, Y. Y. (2017). Effective Heart Disease Detection Based on Quantitative Computerized Traditional Chinese Medicine Using Representation Based Classifiers. *Evidence-Based Complementary and Alternative Medicine*, 2017. https://doi.org/10.1155/2017/7483639

Tama, B. A., Im, S., & Lee, S. (2020). Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble. *BioMed Research International*, 2020. https://doi.org/10.1155/2020/9816142

UCSD. (2021). UC San Diego's Practical Guide to Clinical Medicine. https://meded.ucsd.edu/clinicalmed/heart.html

Understanding DICOM. How to read, write and organize… | by Alexander Weston, PhD | Towards Data Science. (last visited February 2022.

Use Early Stopping to Halt the Training of Neural Networks At the Right Time. (n.d.). https://machinelearningmastery.com/how-to-stop_training-deep-neural-networks-at-the-right-time_using-early-stopping/

Venu, S. K. (2020). An ensemble-based approach by fine_tuning the deep transfer learning models to classify pneumonia from chest X-ray images. *ICAART 2021- Proceedings of the 13th International Conference on Agents and Artificial Intelligence*, 2, 390–401. https://doi.org/10.5220/0010377403900401.

Wang, C., Chen, D., Hao, L., Liu, X., Zeng, Y., Chen, J., & Zhang, G. (2019). Pulmonary image classification based on inception-v3 transfer learning model. *IEEE Access*, 7, 146533–146541. https://doi.org/10.1109/ACCESS.2019.2946000

Xia, Y., Ravikumar, N., Greenwood, J. P., Neubauer, S., Petersen, S. E., & Frangi, A. F. (2021). Super_Resolution of Cardiac MR Cine Imaging using Conditional GANs and Unsupervised Transfer Learning. *Medical Image Analysis*, 71, 102037. https://doi.org/10.1016/J.MEDIA.2021.102037

Yang, X., Bian, C., Yu, L., Ni, D., & Heng. P. A. (2018). Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), 10663(January), 215– 223. http://link.springer.com/10.1007/978-3-319-75541-0