İSTANBUL UNIVERSITY PRESS

Acta Infologica

Acta Infologica,
ACIN 2025, 9 (1): 34–54

https://doi.org/10.26650/acin.1474115
Submitted: 26.04.2024
Revision Requested: 04.11.2024
Last Revision Received: 23.11.2024
Accepted: 24.01.2025
Published Online 21.02.2025

Research Article

Open Access

Classification of Ventricular Septal Defect Disease Using Deep Learning

Kadir Barut [1], İhsan Pençe [2], Özlem Çetinkaya Bozkurt [3] & Melike Şişeci Çeşmeli [2]

[1] Burdur Mehmet Akif Ersoy University, Department of Management Information Systems, Burdur, Türkiye
[2] Burdur Mehmet Akif Ersoy University, Department of Software Engineering, Burdur, Türkiye
[3] Burdur Mehmet Akif Ersoy University, Department of Business Administration, Burdur, Türkiye

Abstract

Ventricular Septal Defect (VSD) disease is the most prevalent type of congenital heart disease. VSD is a hole between the left and right ventricles of the heart structure. VSD disease accounts for approximately one-fifth of all congenital heart disease types. Therefore, accurate disease diagnosis is paramount in determining the most appropriate treatment methods. This study aims to classify VSD disease using the deep learning algorithms VGG16, ResNET50, and Inceptionv3 on Computed Tomography (CT) images and compare the pre-trained algorithms used. One of the reasons why imaging methods such as echocardiography are generally used to detect congenital heart diseases is that there are almost no CT datasets related to this disease. The dataset used in this study is the ImageCHD dataset, which comprises 3D CT scans encompassing 16 distinct types of congenital heart defects. Hyperparameter optimization was performed using the grid search method to enhance the model performance, identifying the VGG16 model as the most effective. The model demonstrated a very high classification accuracy of 99.99% in the training dataset and 99.94% in the test dataset. Gradient-weighted Class Activation Mapping was employed to enhance model explainability, providing visualizations of the regions most critical for the classification, thereby enabling medical professionals to validate AI-driven predictions. An optimized model that successfully classifies VSD using 3D CT image data has been introduced to the literature for the first time. Therefore, this study assumes greater significance in the existing literature and sets a benchmark for future studies.

Keywords Computed tomography · deep learning · medical image classification · transfer learning · Ventricular Septal Defect disease

Citation: Barut, K., Pençe, İ., Bozkurt, Ö. Ç. & Çeşmeli, M. Ş. (2025). Classification of ventricular septal defect disease using deep learning. *Acta Infologica*, *9*(1), 34-54. https://doi.org/10.26650/acin.1474115

Corresponding author: İhsan Pençe ihsanpence@mehmetakif.edu.tr

Introduction

With the developing technology in the world, significant developments and innovations are also experienced in the health field. Mainly since the beginning of the 2010s, artificial intelligence technology has been aimed to be used in different fields, following the significant developments and successful results obtained in the field of artificial intelligence, and with the innovations brought by this technology, much more advanced systems have started to be used compared to the existing system. Health is one of these fields, but considering the importance of human health and treatment processes, it has become one of the leading fields. Especially regarding the correct diagnosis of the disease and applying the proper treatment to the patient, the correct detection of some diseases is a process that challenges healthcare professionals. When human health is taken into consideration, a wrong diagnosis or a wrong decision is of such severe importance that it can even lead to the death of a person.

Ventricular Septal Defect (VSD) disease is the most prevalent type of congenital heart disease. VSD is a hole between the left and right ventricles of the heart structure. VSD disease accounts for approximately one-fifth of all congenital heart disease types. Therefore, accurate disease diagnosis is paramount in determining the most appropriate treatment methods. This study aims to classify VSD disease using the deep learning algorithms VGG16, ResNET50, and Inceptionv3 on Computed Tomography (CT) images and compare the pre-trained algorithms used. One of the reasons why imaging methods such as echocardiography are generally used to detect congenital heart diseases is that there are almost no CT datasets related to this disease. The dataset used in this study is the ImageCHD dataset, which comprises 3D CT scans encompassing 16 distinct types of congenital heart defects. Hyperparameter optimization was performed using the grid search method to enhance the model performance, identifying the VGG16 model as the most effective. The model demonstrated an outstanding classification accuracy of 99.99% in the training dataset and 99.94% in the test dataset. Gradient-weighted Class Activation Mapping was employed to enhance model explainability, providing visualizations of the regions most critical for the classification, thereby enabling medical professionals to validate AI-driven predictions. An optimized model that successfully classifies VSD using 3D CT image data has been introduced to the literature for the first time. Therefore, this study assumes greater significance in the existing literature and sets a benchmark for future studies.

Congenital heart disease is a type of congenital heart disease seen in 8 out of every 1000 babies worldwide (Bernier et al., 2010). Heart diseases, one of the most common types of congenital diseases, can cause serious health problems and deaths (Hallıoğlu et al., 2018). Ventricular Septal Defect (VSD) and Atrial Septal Defect (ASD) are the two most common types of congenital heart defects and account for the majority of the most common congenital heart diseases (Geva et al., 2014). It has been a challenging and complex process for doctors to detect and accurately differentiate these diseases, which show severe differences in the heart structure. Specialized radiologists can detect the types of congenital heart diseases by observing the structural changes in some parts of the anatomical structure of the heart (Left Ventricle, Right Ventricle, Left Atrium, Right Atrium, Aorta). There are techniques generally used to diagnose heart diseases. The echocardiography (EC) imaging technique plays an important role. It is widely used, especially in some cases, for diagnosing whether the heart performs its routine functions properly or for diagnosing structural heart diseases (Tay Lik Wui et al., 2011). Thanks to EC, the work of the heart can be monitored live. Thanks to this technique, which transmits sound waves to the heart, the heart valves, vessels, and muscle movements can be monitored, and heart murmurs and structural status can be observed from different angles. It has been observed that some defects can be missed when only the listening method is used to diagnose congenital heart diseases, and it is emphasized that patients should also be evaluated using the EC imaging method (Yıldız et al., 2015). However, even though EC and similar imaging modalities are more commonly used in diagnosing heart diseases, Computed Tomography (CT) imaging in diagnosing congenital heart diseases has advantages over commonly used modalities such as EC. It can provide more successful results (Dillman & Hernandez, 2009).

Especially since 2010, the rapid rise of artificial intelligence applications worldwide has begun to impact the medical field. When the literature is examined, Sun et al. (2012) used heart sounds to classify VSD disease using a support vector machines (SVM) algorithm and achieved a 98.4% success rate in detecting the heart with VSD disease. Hassani et al. (2017) developed an artificial neural network (ANN) based method to classify heart sounds using discrete wavelet transform to detect the size of VSDs and obtained 96.6% accuracy for the classification of VSDs with small structural size and 93.3% accuracy for VSDs with large structural size. Aziz et al. (2020) used signal-processing techniques to classify ASD and VSD, two types of congenital heart diseases, and achieved an accuracy of 95.24%. They used the SVM algorithm to create the classification algorithm. Wang et al. (2020) obtained sensitivity scores of 96.0% and specificity scores of 96.7% using the automatic murmur recognition system (TAP-CRNN) model in their study using a dataset containing two classes of heart murmurs, normal and VSD. In the heart's second aortic and tricuspid regions, they observed 100% classification success of heart murmurs. They concluded that the automatic murmur recognition system has a promising potential for detecting VSD and other structural heart diseases. Sapitri et al. (2021) reported that they achieved 99.79% pixel accuracy and 97.82% overall accuracy by applying the semantic segmentation method with U-NET architecture on 2D ultrasound videos in their disease detection studies, including three disease classes: VSD, ASD, and AVSD, which is a combination of these two diseases . Yang et al. (2023) compared the classification success of VSD disease using several different models. They reported that the YOLOV5 model was the most successful in classifying VSD disease, with a success rate of 92.59%. In a study by Koundinya et al. (2023), they trained three different congenital heart diseases, including VSD disease, using the CNN model they created by converting heart sounds into spectrograms and achieving 97.12% classification success. Abbas et al. (2024) compared machine learning and deep learning techniques to detect heart diseases by analyzing noisy audio signals and found that the multilayer perceptron model performed the best with an accuracy of 95.65%. Tüzün and Özdemir (2023) demonstrated the efficacy of deep learning models in classifying brain tumors using MRI images, making a significant contribution to accelerating early diagnosis processes and minimizing human errors. This study particularly emphasizes that the EfficientNet-b0 model achieved the highest performance with an accuracy rate of 99.54%. Arslan and Ozdemir (2024) published the DEMxNET deep learning model, which demonstrates success in classifying the stages of Alzheimer's disease and possesses the potential to enhance the reliability of AI-based diagnostic tools in clinical settings through the explainability provided by the LIME method. Dörterler et al. (2024) presented an innovative approach to improve the accuracy of medical datasets by hybridizing meta-heuristic algorithms with K-Means.

Xu et al. (2021) published the world's first 3-dimensional (3D) ImageCHD dataset of CT images of congenital heart diseases. They achieved an overall classification success of 82% in the classification of 16 different congenital heart diseases, including VSD disease, in their congenital heart disease classification study with this dataset. The same study emphasizes that theirs is the first disease classification study developed using CT images on this subject. Their success rates are very suitable for improvement, as they emphasize the classification success they have achieved. When the literature is examined, it is stated that there are very few studies on the classification of congenital heart diseases and the importance of the ImageCHD dataset in terms of its contribution to new studies in this field. In their study, Dillman and Hernandez (2009) stated that although EC and Cardiac Angiography are the most common methods used in the diagnosis of congenital heart diseases, CT and Magnetic Resonance methods are also essential imaging techniques that can be used in the detection of this disease and that CT imaging method is more advantageous to be used in the diagnosis of congenital heart diseases when evaluated in terms of structural examination of the heart due to some technical limitations of EC and other imaging methods, especially when compared with CT scans. When the EC method is evaluated from this point of view, the fact that this dataset is the world's first dataset containing CT images of 16 different types of congenital heart diseases reveals the importance of the dataset. In particular, the medical images that make up this dataset are in 3D, providing the opportunity to examine the heart's structure in more detail and to examine the details of the physical structure of the heart significantly better in diagnosing congenital heart diseases.

VSD has become the most common type of congenital heart disease in newborns as a structural defect due to the formation of a hole between the right and left ventricle of the heart. While the EC method is widely known to be successful in diagnosing VSD disease by analyzing heart murmurs, the ImageCHD dataset used in this study plays a significant role in better structural understanding and detection of VSD disease with its 44 3D 512x512 resolution medical images in the 3D CT dataset. This can be better understood by examining other studies in which VSD disease has been detected using medical images.

Due to the wide variety of congenital heart diseases, the automatic diagnosis and classification of these diseases has been a field of study that has been felt lacking in the literature due to the scarcity of medical image data sets related to this subject and the difficulty of accessing these data. This study is one of the few studies using image datasets to detect and classify VSD disease on 3D medical images. This study categorizes 3D CT scans in the ImageCHD database into "VSD" and "Non-VSD". A grid search method was used to conduct hyperparameter optimization to ascertain the influence of the hyperparameters on the model efficacy. This approach facilitated the identification of the optimal configuration. The original images with a size of 512x512 were separated into 2D slices and resized to 128x128 after separation. The input layers of the VGG16, ResNET50, and InceptionV3 pre-trained deep learning models were modified to have 128x128x3 input layers and two classes in the output layers. In this study, Gradient-weighted Class Activation Mapping (Grad-CAM) was also employed to provide a better visual representation of the classified medical images and to highlight the regions that play a more significant role in the classification process. Grad-CAM improves the explainability of artificial intelligence models in healthcare, providing reliability and transparency. The novelty and contribution of the study are as follows:

- An optimized model that successfully classifies VSD using 3D CT image data has been introduced to the literature for the first time.
- The optimized VGG16 model achieved a high 99.94% classification accuracy on the test set by hyperparameter optimization and evaluating state-of-the-art pre-trained deep learning models.
- This research sets a benchmark for future studies by demonstrating how artificial intelligence models can support and improve the diagnostic accuracy in congenital heart disease, specifically in VSD.
- Integrating the Grad-CAM method enhances the interpretability of the results, allowing medical professionals to validate AI-driven predictions by visualizing the model's focus areas and pinpointing regions critical to diagnosis.

This article is organized as follows: In the materials and methods section, the ImageCHD dataset is explained, and the pre-processing and CNN architectures are given. The experimental studies of VGG16, ResNET50, and InceptionV3 are presented in Section 3. Finally, Section 4 presents the conclusions.

Materials And Methods

This study aimed to classify 3D CT scans obtained from the ImageCHD database into "VSDs" and "Non-VSDs." The original 512x512 images were sliced into 2D slices of 128x128 pixels during the image resizing. The deep learning models VGG16, ResNET50, and InceptionV3 were modified to accept 128 x 128 x 3 inputs and predict two output classes.

ImageCHD dataset

The data used in this study were obtained from the ImageCHD dataset. The ImageCHD dataset, which contains 110 3D CT images, is the first medical image dataset in the field to be published to detect congenital heart disease (Xu et al., 2021). This dataset contains 16 different congenital heart disease data, including 44 congenital heart diseases with VSDs and 110 congenital heart diseases with 512x512x3-dimensional medical images. In addition, some of the 3D CT scan images in the ImageCHD dataset contained more than one congenital heart disease. In this case, it is essential to note that any CT image in the dataset may contain more than one disease.

Table 1 shows the number of medical images of VSD and other diseases in the dataset. Some diseases can be a combination of two or more different diseases.

Table 1

Distribution of diseases in the ImageCHD dataset

Common Congenital Heart Diseases	**ASD**	**AVSD**	**VSD**	**TOF**	**PDA**	**TGA**	**CA**	**PuA**
	32	18	44	12	14	5	6	16
Rare Congenital Heart Diseases	**PAS**	**DORV**	**CAT**	**DAA**	**APVC**	**AAH**	**IAA**	**DSVC**
	3	8	4	5	6	3	3	8
Normal					6			

Because some diseases are more common than other types of diseases, the table is divided into "Common Diseases" and "Rare Diseases". In addition to 16 different disease types, the dataset includes six disease-free healthy heart CT scans.

Figure 1 shows the structural shapes of one of the 3D CT images in the ImageCHD dataset in different axes.

Figure 1

Illustration of the width (X), height (Y), and depth (Z) layers of a 3D image

Since the CT image is 3D, it is seen that training a 3D CNN model with much more detailed data is a more precise and accurate classification approach than a 2D classification model in terms of revealing structural defects in the classification of this disease. In addition, given that each 3D image is of the float64 type, each image has a higher file size and much more detail.

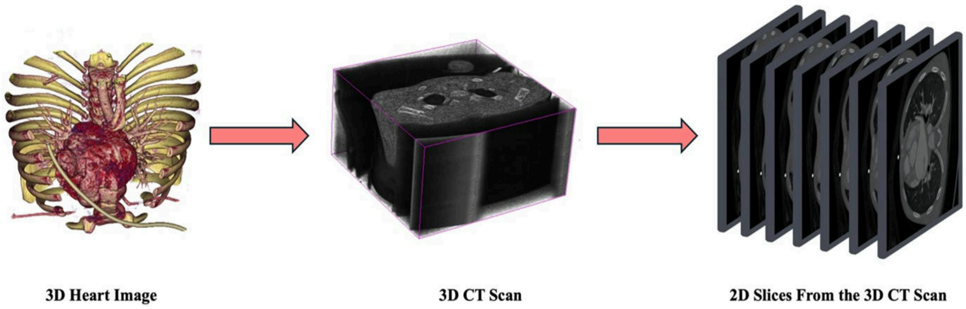
Transforming 3D images into 2D slices

Different methods exist for training a convolutional neural network (CNN) to classify 3D images. In this study, 3D medical images were reduced to 2 dimensions before training the CNN model. As a result, new 2-dimensional (2D) images are obtained from the segmented 3D image from the number of slices that make up the 3D image.

Figure 2 shows how one of the 3D tomography images in the dataset looks when converted to 2D.

Figure 2

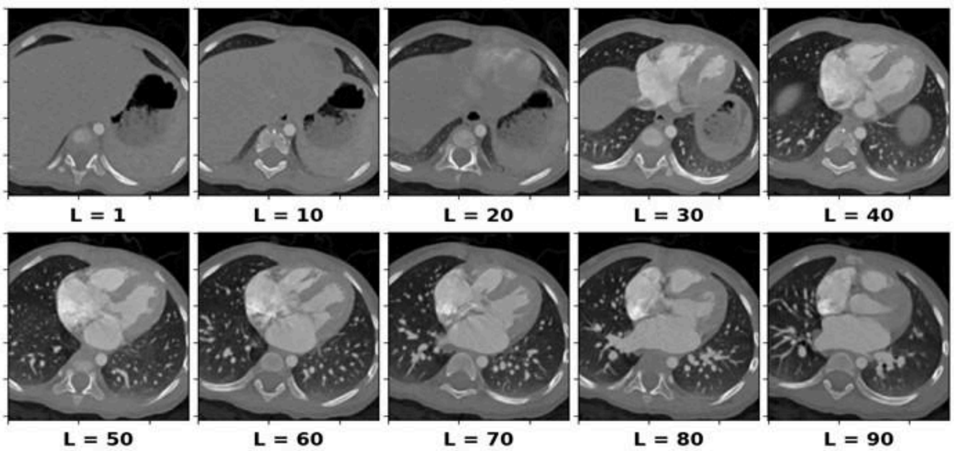
Representing a 3D image as 2D slices



This image, which has a size of (275x512x512), has become 275 (512x512) medical images after being reduced to 2 dimensions. Each of the 110 3D images was separated into two dimensions in this way, and then all the 2D images were resized to (128x128x3). In this way, it was suitable for pre-trained 2D CNN models for model training. Figure 3 shows the specific layers of a 3D image in the ImageCHD dataset in 2D.

Figure 3

Some 2D slices of a 3D image



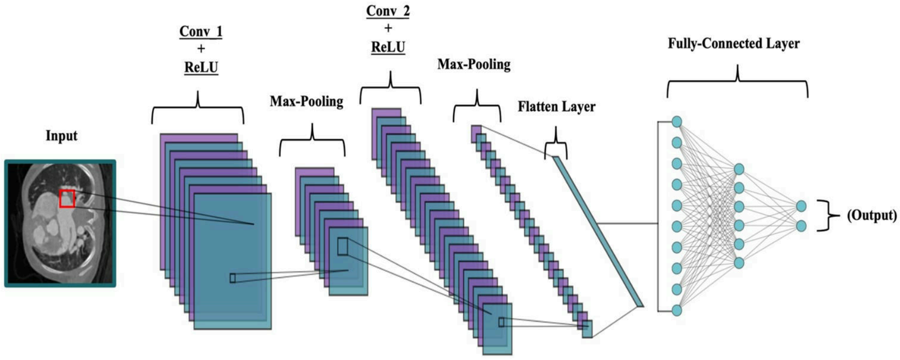
L denotes the number of layers on the 3D image in Figure 3. There are ten 2D CT images between each image. As the layers change, the structural differences in the image can be seen more clearly.

Convolutional neural networks

This neural network model, first introduced by Yann LeCun et al. in 1998, has seen a rapid increase in usage in recent years and has become the most popular neural network model used in image classification problems. It can give outstanding results, especially in classification problems. Today, CNN models are used in many fields, and different neural network models can be created in a specific field by developing the model according to the field needed. It is frequently used in image classification, natural language processing, style transfer, and text mining. Figure 4 shows the general structure of the CNN model.

Figure 4

The general structure of a CNN



The CNN structure, which is highly preferred in the medical field due to its successful results in disease classification in recent years, is an improved version of ANN. The CNN model, which works on image data, attempts to obtain the attributes of the object to be detected in the image by processing the images in layers consisting of some basic pre-processing before entering the ANN. Before these attributes enter the ANN, the image is reduced to a much smaller size than its original size, and the extracted attributes are reduced to a single dimension. This process is called flattening. After the flattening process, the data becomes one-dimensional and thus enters the ANN (Yamashita et al., 2018). The CNN model has become increasingly popular, especially after its highly successful results in high-impact image classification competitions like ImageNET (Russakovsky et al., 2014). Using the CNN model, binary or multiclass classification can be performed on the image.

Unlike the structure of the regular ANN model, the CNN model includes some essential layers, such as convolution and pooling. In this structure, pre-processing is performed to reduce the data size before the data enters the neural network and is trained. After obtaining the features that separate the attributes specific to these data instead of the entire data, they are included in the neural network.

ImageNet data set

When the ImageNet dataset was first published in 2010 under ImageNet LSVRC-2010, it contained 1.2 million high-resolution images in 1000 categories (Deng et al., 2009). As of 2024, it has become one of the most popular image datasets in the world, containing 14 million images. Many pre-trained deep learning models have been trained using this dataset. In addition, an image classification competition called "The ImageNet Large Scale Visual Recognition Challenge (ILSVRC)" has been organized annually since 2010. This competition determines the best algorithm for object detection using the ImageNet dataset (Russakovsky et al., 2014). InceptionV3, ResNET50, and VGG16 pre-trained deep learning models used in this study were also trained using the ImageNet dataset.

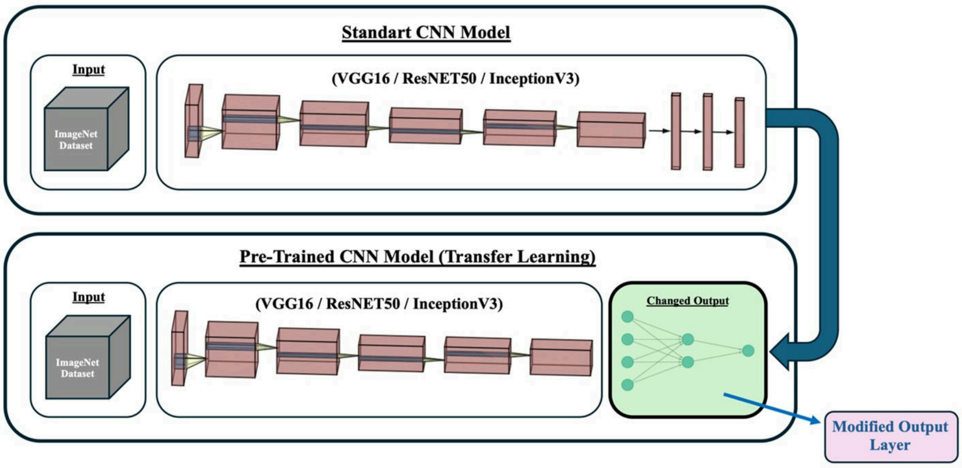
Transfer learning

Although CNN produces excellent results in image classification problems, it is a more challenging model for dealing with some problems. Some reasons are the lack of data sets for model training, the data size being too high, or the need for more costly hardware and time, as the CNN model consists of too many layers. Such problems that stand in the way of better model training and high classification success are significant challenges for scientists and researchers to overcome in their work. Transfer learning is often used to deal with these problems, and good results can be obtained. Transfer learning is the reuse of a previously trained artificial intelligence model by modifying it for different problems (Tan et al., 2018). Modifying and using some pre-trained neural networks with data that is too large to be obtained personally under normal conditions increase the success rate (Accuracy) obtained from the model. Some transfer learning models that gave successful results in most cases were used in this study. These models are called pre-trained deep learning models because they are trained with extensive data and have many layers in their structure.

Figure 5 shows a structural comparison between a typical CNN model and the modified version of the pre-trained deep learning models.

Figure 5

Comparison of the CNN and the pre-trained model



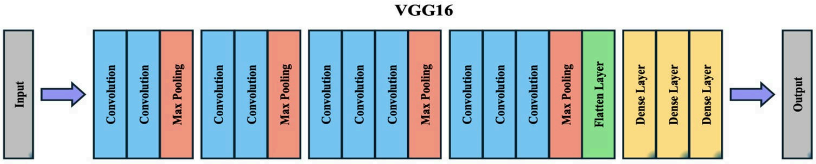
In this study, ResNET50, VGG16, and InceptionV3 pre-trained deep learning algorithms were utilized, and the original output layers were modified to have two neurons.

VGG16 model

In 2014, Karen Simonyan and Andrew Zisserman from the University of Oxford introduced the VGG16 and VGG19 models at the ILSVRC competition organized by ImageNet to investigate the effects of neural network depth on the success rate of large-scale data classification problems (Simonyan & Zisserman, 2015). The most important differences between these models and the previous deep CNN models are the smaller filter sizes and more convolutional layers. Figure 6 shows the structure of the VGG16 model. While developing these deeper CNN models, Krizhevsky et al. (2017) referenced their classification study using the ImageNet dataset. VGG models have been widely used in recent studies because of their high performance. This technique offers high performance primarily because it is already optimized for training with multidimensional data. Consequently, the model's feature map is highly developed (Kara, 2023).

Figure 6

Internal structure and layers of the VGG16 model



The structure of the VGG models consists of convolution, pooling, and fully connected layers. The input layer accepts 224x224 fixed-size inputs. The only pre-processing is to extract the average RGB (Red, Green, Blue) value from each pixel value. In the convolution layers, 3x3-dimensional filters are used. The five convolution layers have a 2x2 maximum pooling layer. The successive convolution and pooling layers are connected to fully connected layers of 1000 classes in the output layer (Simonyan & Zisserman, 2015).

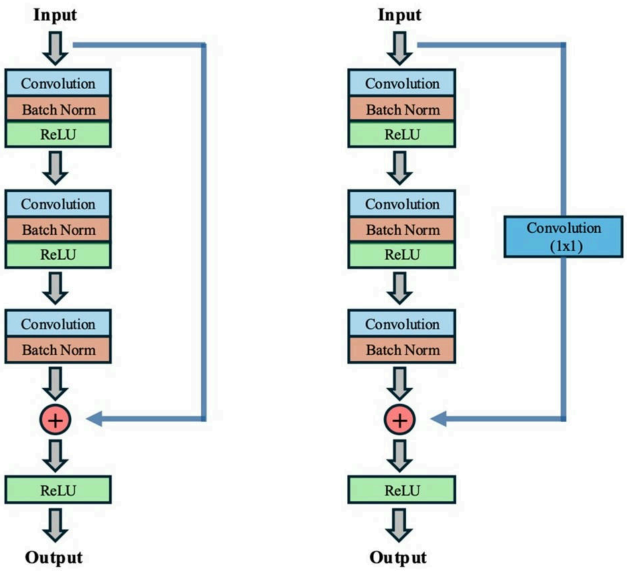
ResNET50 model

The ResNET50 model has less complexity than other popular deep learning models and simplifies the difficulty of training high-dimensional deep learning models with significantly more layers. Although it is much deeper than the VGG models, it has less complexity (He et al., 2016). The ResNET50 model trained with the ImageNet dataset won first place in classification in the ILSVRC competition held in 2015. In the Microsoft Common Object in Context (COCO) competition, they won first place in detection and segmentation (He et al., 2016).

The ResNET50 model accepts 224x224 color image data at the input layer. Instead of the standard convolution layer, a specially modified structure called "Residual Block" is used, and the model contains 16 Residual Blocks. Each of these so-called Residual Blocks contains two convolution layers. The convolution layers contain different filter sizes, such as 7x7 and 3x3, and there are connections between each Residual Block structure. Thanks to these connections, the gradient vanishing problem is minimized. The Residual Block structure of the ResNET50 Model is shown in Figure 7.

Figure 7

The residual block structure of the ResNET50 model



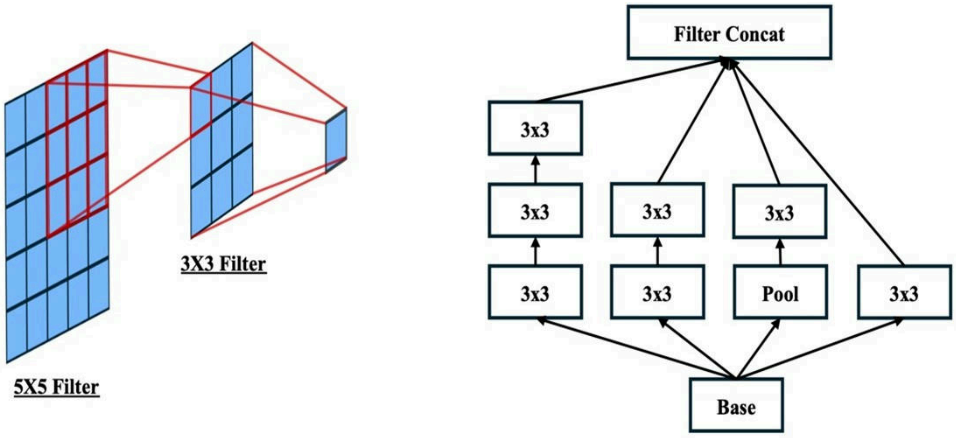
Since the ResNET50 model is trained with the ImageNet dataset, it consists of fully connected networks with 1000 classes in the output layer (He et al., 2016).

InceptionV3 model

Inception, also known as GoogLeNet, has been developed to outperform other large deep-learning models with respect to high computational cost and memory usage (Szegedy et al., 2016). InceptionV3 is a fast, powerful, and popular deep learning model with a high success rate in classification, even though it uses only 5 million parameters in training and is trained with a much smaller number of parameters compared to a popular deep learning algorithm such as VGG (Szegedy et al., 2016). The InceptionV3 model, such as the VGG16 and ResNET50 models, was one of the high-performing deep-learning models in the ILSVRC competition. Figure 8 shows the Asymmetric Convolution structure of the InceptionV3 model.

Figure 8

Asymmetric convolution structure of the InceptionV3 model



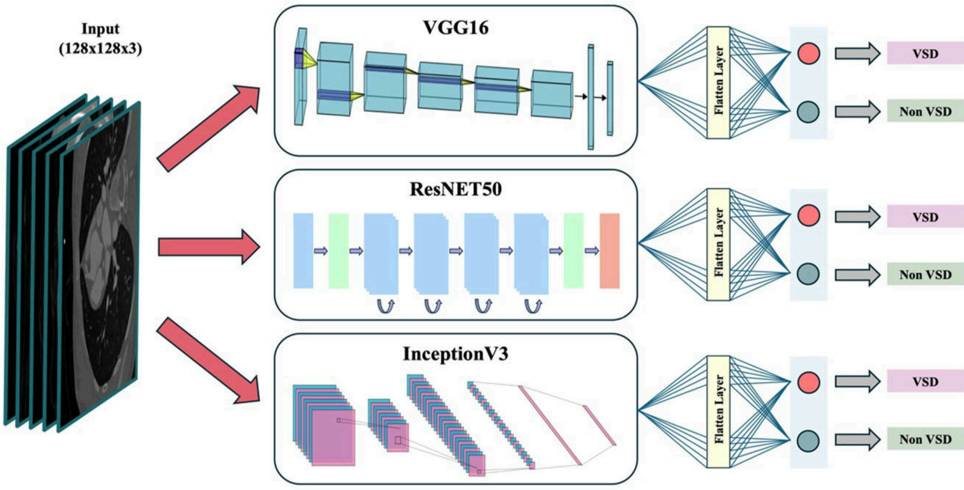
The factorized convolution process used in the model's structure is one of the most important reasons for the low hardware cost, speed, and high performance of the InceptionV3 model trained with few parameters. Filter sizes considered relatively large, such as 5x5 or 7x7, incur a disproportionate computational cost. For example, instead of a 5x5 filter, using two fully connected 3x3 filters is preferable. In the InceptionV3 model, 1x1 and 3x3 filters are generally preferred. In this way, the computational cost remains lower than that of large-size filters, and a significant performance increase is achieved.

Proposed method

For VSD disease classification, the ImageCHD dataset containing 3D CT images was converted into 2D CT images. VGG16, ResNET50, and InceptionV3 deep learning models were used in the training. These three models using the transfer learning method were modified, as shown in Figure 9, with input layers of 128x128x3 and output layers modified to be directly 2-class. The models were built with an input layer of 128x128x3 2D CT and an output layer of a categorical binary classifier as "with VSD" or "without VSD". To utilize the pre-trained weights, it was necessary to turn off the trainable feature of the layers before training the models. Furthermore, pre-trained with the ImageNet dataset, the default model weights were employed to avoid training all models from scratch.

Figure 9

Visualization of the classification method used



A 10-fold cross-validation technique and 25 training cycles (epoch) were preferred for the model training. The batch size was 16, and the activation function was "Softmax". For model optimization, the Adam optimization algorithm was preferred.

Cross-validation and performance metrics

The k-fold cross-validation method aims to evaluate the statistical success of an artificial intelligence model more objectively and to reveal the actual success rate more accurately. In this study, a 10-fold cross-validation was applied for model training.

The performance metrics used to evaluate the results of the VGG16, ResNET50, and Inceptionv3 models are analyzed in this section. Accuracy, Recall, Precision, F1-Score, Confusion Matrix, and Roc Curve performance metrics frequently used in image classification problems are used to analyze the trained models (Sevli, 2023).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{F1-Score} = \frac{2^{*}\ \text{Presicion}\ {}^{*}\ \text{Recall}}{\text{Presicion} + \text{Recall}} \tag{4}$$

Results

This study aimed to classify VSD disease using deep learning models. VGG16, ResNET50, and InceptionV3 pre-trained deep learning models were trained using 3D CT images from the ImageCHD dataset. The 3D CT scans were converted to 2D and resized to 128x128 before model training. The models were modified to have 2 class labels, "VSD" and "Not VSD", and the output layers were set as binary classifiers. To determine the effects of the hyperparameters on the model success, hyperparameter optimization was performed using the grid search method. In addition, to achieve statistically more accurate results, the "K-Fold Cross Validation" method was applied and set to work in 10 folds, and the entire data set was entered into the model for training as training and test data in pieces for each model.

Comparative analysis of the effects of the hyperparameters on the models

A comparative analysis of the effects of the hyperparameters on the models' classification metrics results was evaluated under varying batch sizes and learning rates over 10-fold cross-validation, as shown in Table 2. Nvidia P100 GPU and 16 GB VRAM were used as the computational resources.

Table 2

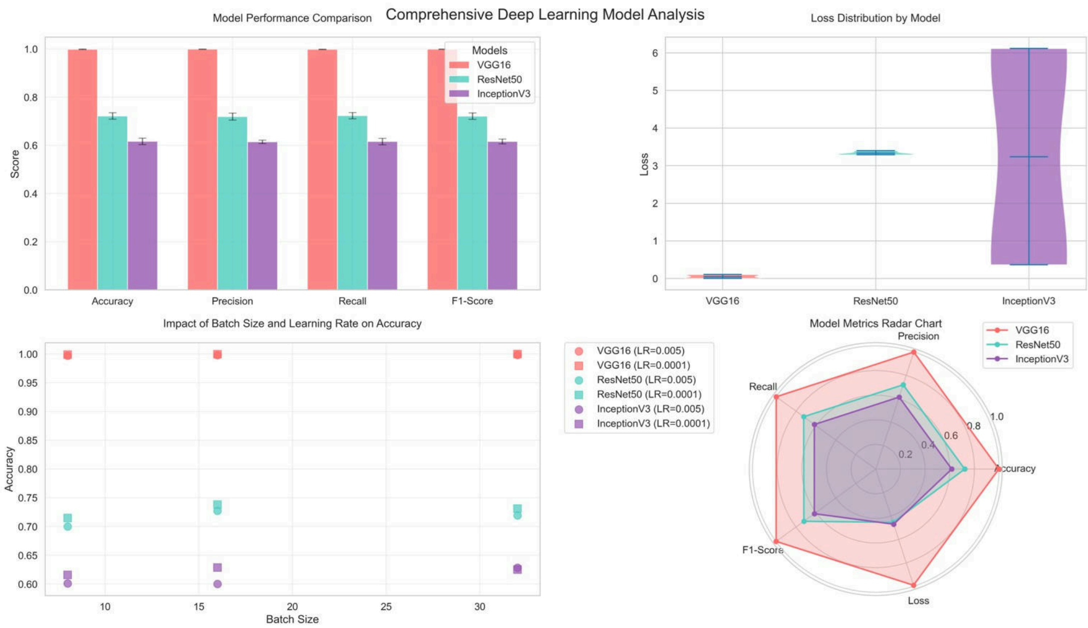
Demonstrating the effects of different hyperparameter values on the performance of the models through performance metrics

Model	Learning Rate	Batch Size	Train					Test				
			Accuracy	Loss	Recall	Precision	F1-Score	Accuracy	Loss	Recall	Precision	F1-Score
VGG16	0.005	8	0.9987	0.126	0.999	0.999	0.999	0.9967	0.123	0.997	0.997	0.997
	0.005	16	0.9981	0.204	0.998	0.998	0.998	0.9959	0.352	0.996	0.996	0.996
	0.005	32	0.9986	0.096	0.999	0.999	0.999	0.9973	0.104	0.997	0.997	0.997
	0.0001	8	0.9989	0.286	0.999	0.999	0.999	0.9977	0.342	0.998	0.998	0.998
	0.0001	16	0.9998	0.061	1.000	1.000	1.000	0.9993	0.110	0.999	0.999	0.999
	0.0001	**32**	**0.9999**	**0.004**	**1.000**	**1.000**	**1.000**	**0.9994**	**0.041**	**0.999**	**0.999**	**0.999**
ResNet50	0.005	8	0.7232	3392	0.723	0.754	0.720	0.7250	3305	0.725	0.755	0.723
	0.005	16	0.7272	4.225	0.727	0.753	0.720	0.7262	4.294	0.726	0.751	0.718
	0.005	32	0.6959	5.072	0.696	0.742	0.690	0.6956	5.132	0.696	0.744	0.690
	0.0001	8	0.6837	6.428	0.684	0.758	0.679	0.6829	6.267	0.683	0.756	0.678
	0.0001	16	0.6687	8.594	0.669	0.747	0.654	0.6705	8.755	0.671	0.747	0.656
	0.0001	32	0.7381	3.289	0.738	0.747	0.736	0.7346	3.124	0.735	0.745	0.733
InceptionV3	0.005	8	0.5627	6.117	0.563	0.550	0.512	0.5648	5.967	0.565	0.552	0.513
	0.005	16	0.6000	6.938	0.600	0.607	0.519	0.5972	7.168	0.597	0.573	0.516
	0.005	32	0.5786	6.448	0.579	0.569	0.495	0.5788	6.164	0.579	0.572	0.495
	0.0001	8	0.6248	0.900	0.625	0.612	0.545	0.6272	1.512	0.627	0.609	0.548
	0.0001	16	0.6260	0.527	0.626	0.614	0.558	0.6277	1.030	0.628	0.625	0.560
	0.0001	32	0.6286	0.368	0.629	0.613	0.561	0.6277	1. 217	0.628	0.610	0.559

Table 2 presents the classification metrics of these models, namely VGG16, ResNet50, and InceptionV3, that were trained using various combinations of batch size (8, 16, and 32) and learning rate (0.005, 0.0001). The variability in the VGG16 graphical network unfolds a clear advantage over other approaches in terms of illustrative performance and tenability, where the model performs a variety of fundamental classification techniques through vital metrics such as Accuracy, Loss, Recall, Precision, and F1-Score. These outcomes imply that VGG16 is the best and most stable model from the group for conducting the classification tasks that were tested. Specifically, the combination of a batch size of 32 and a learning rate of 0.0001 resulted in the highest classification performance, achieving an accuracy of 99.94%. Even though the ResNet50 model, being better than the InceptionV3 one, shows good classification performance, it has limitations with a loss value of 3.124, which means it cannot efficiently classify some samples. The network architecture is characterized by the integration of "residual blocks" specifically created to solve the vanishing gradient problem and give a structural privilege over the InceptionV3 algorithm in this case. However, the model's relatively high loss value indicates the need for further optimization or refinement in hyperparameter tuning to achieve greater reliability across diverse datasets. The InceptionV3 model demonstrated the lowest performance across all metrics in this study. With an accuracy of 62.77%, the model is significantly less effective than the VGG16 and ResNet50 models regarding the classification accuracy on this dataset. Even though it is an advanced model, the InceptionV3 model does not appear to be very successful in classification. One of the major drawbacks of InceptionV3 is that it uses multiple parallel convolutional layers to extract details at various feature scales, increasing the complexity. Although such an architecture is advantageous for particular datasets, it might introduce redundancy in all computing. Thus, the performance might be under productivity when other datasets with a simple or even less varied configuration are applied. In Figure 10, the classification success of the VGG16, ResNet50, and InceptionV3 models in all metrics is very detailed as comparative graphs.

Figure 10

Comparative visualizations presenting the results achieved by the VGG16, ResNet50, and InceptionV3 models

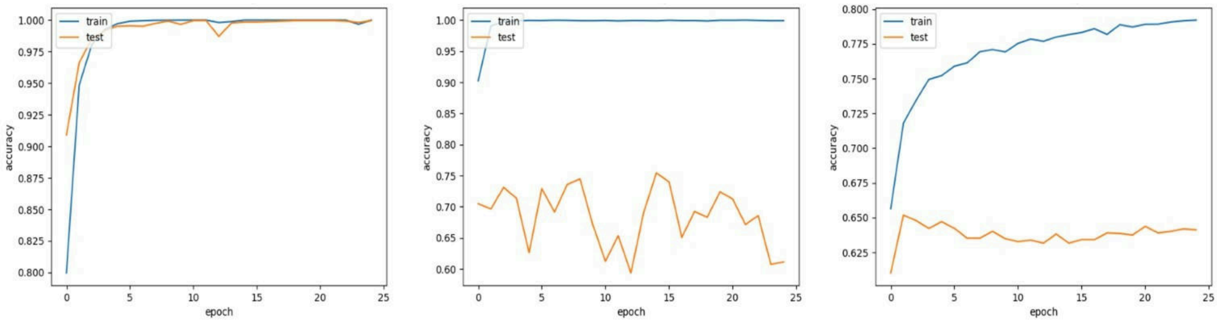


In Figure 10, the inconsistent results obtained from the InceptionV3 model are especially noticeable in the loss values graph. It is observed that the VGG16 and ResNet50 models obtain more consistent loss values. When the accuracy graph is examined, the accuracy results obtained from the models at different learning rate values (0.005, 0.0001) are seen. It is seen that the VGG16 model is again the most consistent and successful model with the results it obtains. The results of the ResNet50 model show that similar results were obtained at two different learning values, but the model trained with the 0.0001 learning rate performed slightly better. When the results of the InceptionV3 model are examined, it is more clearly seen that the model trained with the learning rate value of 0.0001 obtains more successful results than the model trained with the value of 0.005. When the radar chart is examined, the models are compared according to the metric results obtained. It can be observed very clearly that the VGG16 model is the most successful model that gives the best results by far. In addition, the superiority of VGG16 over the ResNe50 and InceptionV3 models, especially in the Loss value, is very clearly visible in the radar chart. It is seen that the Loss values of the other two models are close to each other. The charts reveal the superiority of the VGG16 model in classification success over the other models.

Figure 11 shows the accuracy curve graphs for the first fold of the VGG16, ResNET50, and InceptionV3 models with the specified learning rate of 0.0001 and batch size of 32.

Figure 11

Accuracy metric results for the first fold of the optimized VGG16, ResNET50, and InceptionV3 models

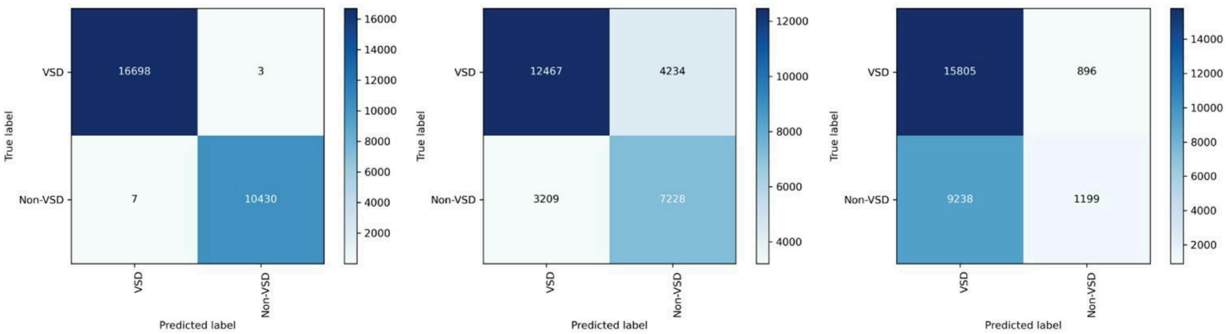


a) Accuracy Curve for VGG16 b) Accuracy Curve for ResNet50 c) Accuracy Curve for InceptionV3

The relationship between the actual and predicted data of the optimized models is illustrated in Figure 12, which depicts the confusion matrices. By looking at this table showing the TP, TN, FP, and FN values, inferences can be made about the performance of the models. When the results are compared with accuracy graphs and confusion matrix tables, inferences can be made more easily about the success of the models.

Figure 12

The confusion matrices of the optimized VGG16, ResNET50, and InceptionV3 models

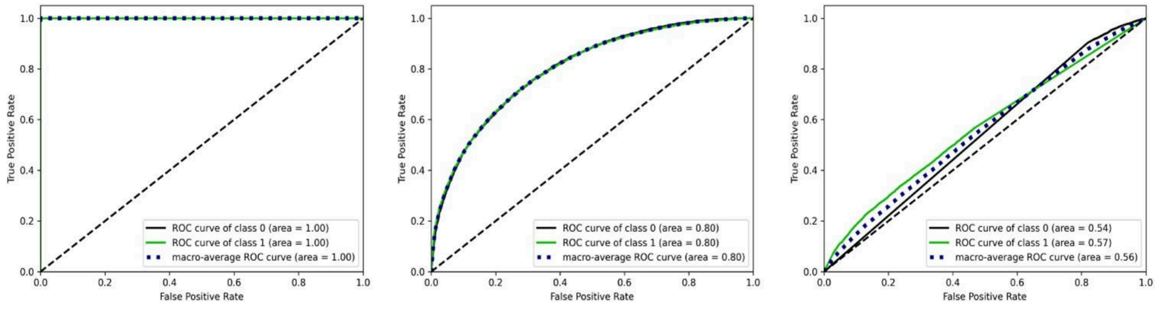


a) VGG16 Confusion Matrix b) ResNet50 Confusion Matrix c) InceptionV3 Confusion Matrix

When the confusion matrices are examined, the classification success of the VGG16 model attracts attention. The performance superiority of the VGG16 model, which stands out with its high accuracy rate and classification success, is seen compared to the ResNET50 and InceptionV3 models. The ResNET50 and InceptionV3 models achieve average classification success. When looking at the accuracy graphs, it is seen that the training and test curves are consistent and have a high success rate in the VGG16 model. At the same time, it is noteworthy that the training and test curves are inconsistent and have low success rates in the ResNET50 and InceptionV3 models. Figure 13 shows the Roc Curve graphs of the VGG16, ResNET50, and InceptionV3 models.

Figure 13

Roc curve of the optimized VGG16, ResNET50, and InceptionV3 models



a) VGG16 ROC Curve b) ResNet50 ROC Curve c) InceptionV3 ROC Curve

When the Roc Curve graphs, another statistical metric, are examined, the success of the VGG16 model stands out compared to the ResNET50 and InceptionV3 models. The results of the Roc Curve graphs obtained from the VGG16, ResNET50, and InceptionV3 models support the results of all other performance metrics. When the graph of the VGG16 model is examined, it has achieved a very high success. When the graphics of the ResNET50 model are examined, it is seen that it gives better results than the InceptionV3 model, but it is still not good enough and has a slightly above-average performance. When the Roc Curve chart of the InceptionV3 model is examined, it is seen to give bad results. According to these results, the InceptionV3 model can be interpreted as the model with the most unreliable results. It can be clearly stated that the most reliable and successful model is the VGG16. As a result, all models' performance metrics and performances were analyzed statistically. According to the results obtained, the performance metrics confirmed each other.

Comparative analysis of the effects of hyperparameters on the training times of the models

A comparative analysis of the effects of the hyperparameters on the models' training times under varying batch sizes and learning rates over 10-fold cross-validation is presented in Table 3. Nvidia P100 GPU and 16 GB VRAM were used as the computational resources.

Table 3

Comparative analysis of training durations for the VGG16, ResNet50, and InceptionV3 models evaluated under varying batch sizes and learning rates over a 10-fold cross-validation

Model	Number of Parameters	Learning Rate	Batch Size	Total Training Time (minutes)	Average Training Time per Fold (minutes)
VGG16	138 million	0.005	8	130.51	13.05
		0.005	16	116.17	11.61
		0.005	32	91.14	9.11
		0.0001	8	130.31	13.03
		0.0001	16	114.92	11.49
		0.0001	32	93.73	9.37
ResNet50	25.6 million	0.005	8	207.28	20.72
		0.005	16	193.72	19.37
		0.005	32	127.79	12.77
		0.0001	8	205.88	20.58

Model	Number of Parameters	Learning Rate	Batch Size	Total Training Time (minutes)	Average Training Time per Fold (minutes)
		0.0001	16	209.03	20.90
		0.0001	32	130.50	13.05
InceptionV3	23.8 million	0.005	8	254.98	25.49
		0.005	16	212.24	21.22
		0.005	32	133.80	13.38
		0.0001	8	253.97	25.39
		0.0001	16	210.16	21.01
		0.0001	32	137.92	13.79

Table 3 compares the VGG16, ResNet50, and InceptionV3 models regarding training times under different learning rates and batch size values. Each model's architectural features and number of parameters significantly affected the training times. Although the VGG16 model, with 138 million parameters, is significantly larger than the ResNet50 and InceptionV3 models, its training time is shorter than the other two. This shorter training time indicates that the architectural complexity of the models has a more significant impact on the training time than the number of parameters they contain. The results indicate that the training time significantly decreased with larger batch sizes, while the learning rate (0.005 and 0.0001) had minimal impact on the training duration. Although the InceptionV3 model has a lighter architecture than the VGG16 and ResNet50 models regarding the number of parameters, its training time is longer than that of the other models. The results demonstrate that with larger batch sizes, the training duration significantly decreased while the impact of the learning rates (0.005 and 0.0001) on the training time was minimal. The visualization hyperparameter analysis for the VGG16, ResNET50, and InceptionV3 models is shown in Figure 14.

Figure 14

Visualization hyperparameter analysis for VGG16, ResNET50, and InceptionV3 models

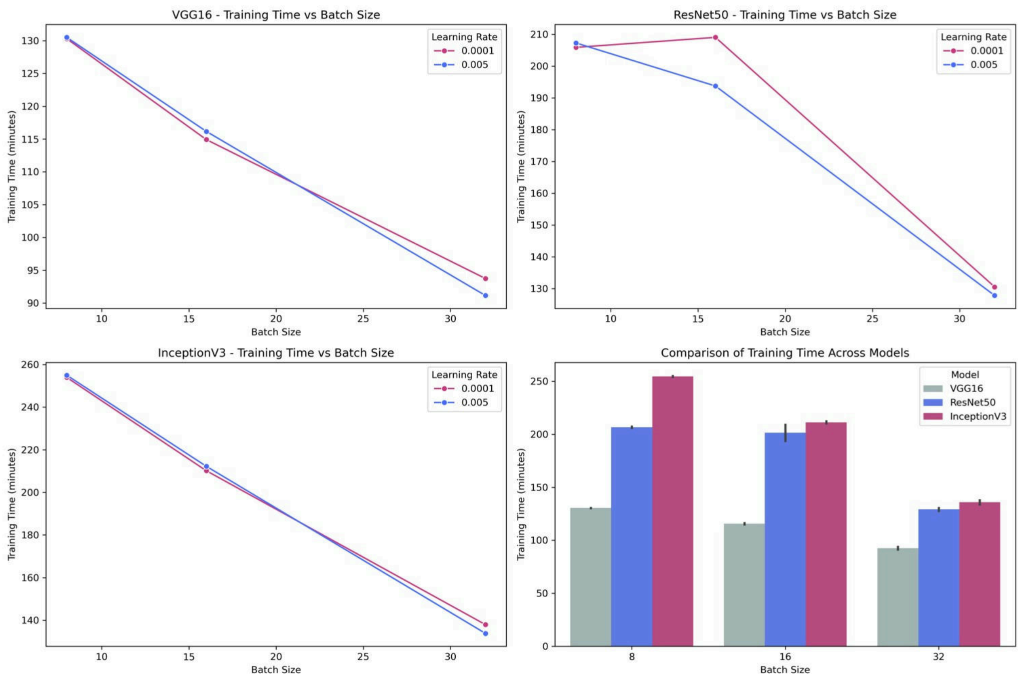


Figure 14 shows a comparative analysis of the results obtained by the VGG16, ResNet50, and InceptionV3 models when different hyperparameters (learning rate and batch size) were used. VGG16 is the model that exhibits the most efficient training time when testing different hyperparameter configurations. The lowest training time indicates that, among the models tested, VGG16 demonstrates the best training efficiency. It is the optimal choice for reducing the training time in this experimental context. These results demonstrate the necessity of grasping the influence of various deep learning models and hyperparameter settings on the training time. This model selection process should be treated as one of the most crucial factors in the decision-making process.

Gradient-weighted Class Activation Mapping Analysis

Grad-CAM is an eminent interpretable artificial intelligence (XAI) technique that seeks to elucidate the inner workings of artificial intelligence models, including their decision-making processes and outputs, for better human understanding. The approach thus allows the connection between the magnitude of AI programs and human thought to be bridled by providing visualizations of the rationale behind a model's predictions.

In this experiment, the Grad-CAM algorithm (Selvaraju et al., 2017), one of the famous XAI techniques, was implemented as a method of localization of regions in CT scans believed by the model to be affected by VSD disease. Grad-CAM creates heatmaps by distributing larger activation values to the pixels influencing the model's prediction. This heatmap leads to a color-coded diagram according to different diseased parts of the patient, henceforth providing doctors with a map indicating where the model is looking.

The Grad-CAM algorithm was employed to assess the output of a deep learning model trained for VSD disease diagnosis. By partially covering the CT images with the overlayed heatmaps, the regions of the disease were visually separated from the rest of the image. Therefore, the zones with more intensive red hues are where the model the presence of disease, and the blue zones are where no disease was found. This approach furnishes a comprehensible rationale for the model's thinking to healthcare experts, thus making it easily understood. The Grad-CAM output for a CT scan image with and without VSD Disease is presented in Figure 15.

Figure 15

View of CT scans with and without VSD Disease using Grad-CAM

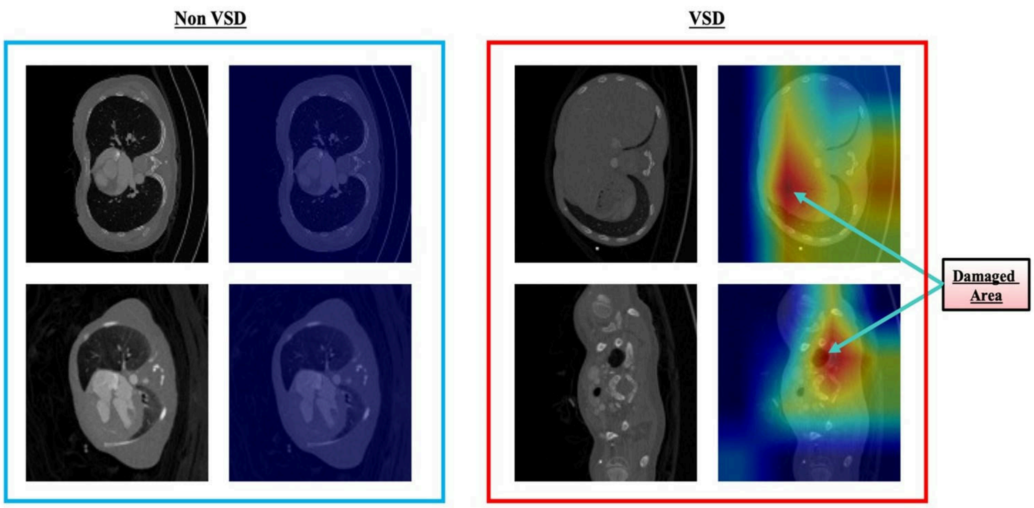
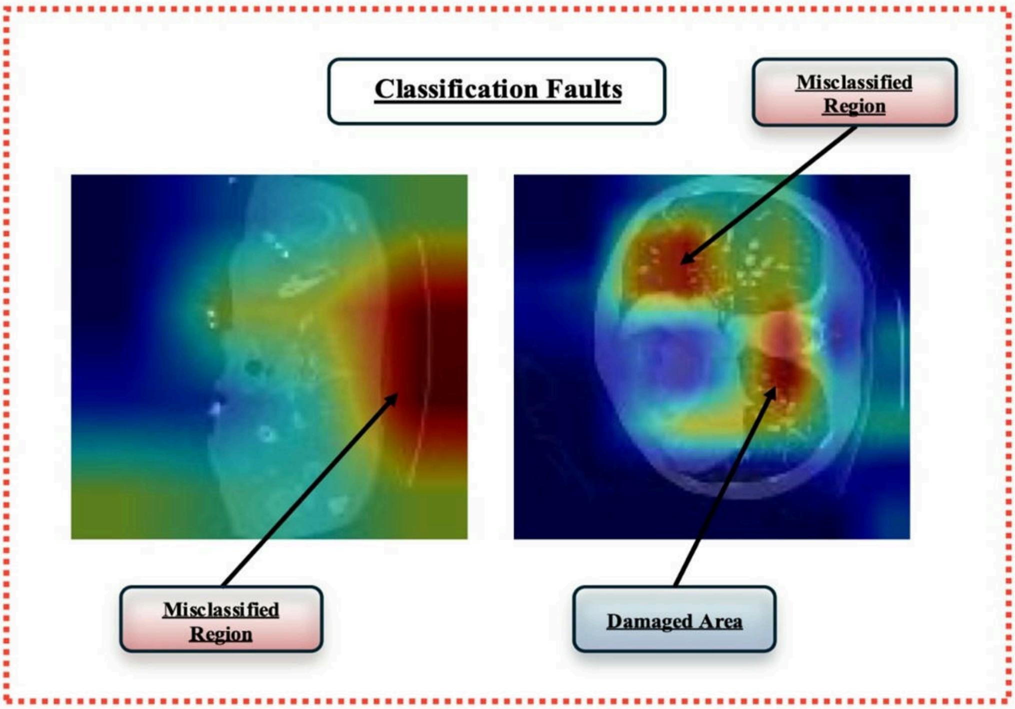


Figure 15 presents an example of the Grad-CAM output for a CT scan image. The red-highlighted areas represent regions where the model detects VSD disease, whereas the blue-highlighted areas correspond to regions where the disease is absent. This visual representation enables a straightforward and interpretable understanding of the model's predictions. Visualization of the classification faults can be seen in Figure 16.

Figure 16

Visualization of the classification faults



The Grad-CAM method, as shown in Figure 16, mistakenly marked a part of a CT scan image as a VSD case. This picture shows a problem that the deep learning model applied in this study has since it wrongly identified the normal area as the one with the disease. Therefore, the Grad-CAM heatmap was incorrectly focused on this image region indicated by the red spot, which could confuse the medics interpreting the results. Noise in the input data: Artifacts or inconsistencies in CT scans are the main reasons for wrongful feature extraction by the model. The misclassification in Figure 16 underscores the importance of not relying solely on AI-based models for clinical decision-making, especially in high-stakes fields like healthcare.

In addition to verifying the predicted results of the deep learning model, these visual outputs help clinicians interpret the data. Thus, it acts as a vital item. Using this method, medical professionals can act as supervisors and thus verify or disprove the model's reasoning. Thus, physicians will have additional confidence in their already great expertise in diagnosis. In conclusion, Grad-CAM is a useful tool for understanding artificial intelligence models. However, it must be applied with other validation methods and domain expertise to ensure reliable results.

Conclusion

The fact that VSD disease is the most common type of congenital heart disease reveals the importance of new treatment methods to be developed for this disease. To develop methods, it is of great importance to diagnose the disease correctly. Deep learning algorithms can achieve successful results in disease detection.

However, the fact that there are very few available data sets for VSD disease, especially the almost complete absence of medical image data sets, has caused the number of studies in this field to remain limited.

ImageCHD, the first 3D CT medical image dataset published on congenital heart diseases, was used in this study. Because of the research, when the literature is examined, this study is the first study conducted on the classification of VSD disease on CT image data. In order to ascertain the optimal model and parameters for classifying the VSD disease, hyperparameter optimization was conducted using the grid search method. By using optimized VGG16, InceptionV3, and ResNET-50 pre-trained deep learning models for classification, it was determined that the model with the highest degree of accuracy was the VGG16 model with a value of 99.99% in the training set and 99.94% in the test set. The high classification success achieved shows that CT image data are suitable for detecting or classifying this disease.

Moreover, the activation map generated by the Grad-CAM method, considering the weights in the deep learning model's last convolutional layer, showed that the model focused on the exact regions and made successful predictions. The Grad-CAM method allows for the visualization of the regions of an image that are the focus of the model's attention during the decision-making process. This visualization enables medical professionals to gain insight into and validate the decisions made by artificial intelligence systems. Additionally, the Grad-CAM model can be employed to analyze instances of inaccurate predictions, which can help identify the model's limitations and facilitate improvements in the development process.

This study will facilitate new research endeavors in this domain in the coming years. This is the first study to use CT images to classify VSD disease, and it has yielded the most optimal outcomes regarding general classification efficacy. In future work, classification studies will be conducted for the disease categories in the ImageCHD dataset.

Peer Review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- Ö.Ç.B.; Data Acquisition- K.B.; Data Analysis/Interpretation- İ.P., M.Ş.Ç.; Drafting Manuscript- K.B., İ.P.; Critical Revision of Manuscript- Ö.Ç.B., M.Ş.Ç.; Final Approval and Accountability- K.B., İ.P., Ö.Ç.B., M.Ş.Ç.; Supervision- İ.P.

Acknowledgments: The numerical calculations reported in this paper were fully/partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources).

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

Author Details

Kadir Barut

[1] Burdur Mehmet Akif Ersoy University, Department of Management Information Systems, Burdur, Türkiye

0000-0002-6288-5954

İhsan Pençe

[2] Burdur Mehmet Akif Ersoy University, Department of Software Engineering, Burdur, Türkiye

0000-0003-0734-3869 ihsanpence@mehmetakif.edu.tr

Özlem Çetinkaya Bozkurt

[3] Burdur Mehmet Akif Ersoy University, Department of Business Administration, Burdur, Türkiye

0000-0002-6218-2570

Melike Şişeci Çeşmeli
[2] Burdur Mehmet Akif Ersoy University, Department of Software Engineering, Burdur, Türkiye
0000-0001-9541-2590

References

Abbas, S., Ojo, S., Al Hejaili, A., Sampedro, G. A., Almadhor, A., Zaidi, M. M., & Kryvinska, N. (2024). Artificial intelligence framework for heart disease classification from audio signals. *Scientific Reports*, *14*(1), 3123. doi:10.1038/s41598-024-53778-7

Arslan, N. N., & Ozdemir, D. (2024). Analysis of CNN models in classifying Alzheimer's stages: comparison and explainability examination of the proposed separable convolution-based neural network and transfer learning models. *Signal, Image and Video Processing, 18*(S1), 447–461. DOI: 10.1007/s11760-024-03166-5

Aziz, S., Khan, M. U., Alhaisoni, M., Akram, T., & Altaf, M. (2020). Phonocardiogram signal processing for the automatic diagnosis of congenital heart disorders through the fusion of temporal and cepstral features. *Sensors*, *20*(13), 3790. doi:10.3390/s20133790

Bernier, P. L., Stefanescu, A., Samoukovic, G., & Tchervenkov, C. I. (2010). The challenge of congenital heart disease worldwide: epidemiologic and demographic facts. *Seminars in Thoracic and Cardiovascular Surgery: Pediatric Cardiac Surgery Annual*, *13*(1), 26–34. https://doi.org/10.1053/j.pcsu.2010.02.005

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255). IEEE. doi: 10.1109/CVPR.2009.5206848

Dillman, J. R., & Hernandez, R. J. (2009). Role of CT in the evaluation of congenital cardiovascular disease in children. *American Journal of Roentgenology*, *192*(5), 1219–1231. doi: 10.2214/AJR.09.2382

Dörterler, S., Dumlu, H., Özdemir, D., & Temurtaş, H. (2024). Hybridization of meta-heuristic algorithms with k-means for clustering analysis: Case of medical datasets. *Gazi Journal of Engineering Sciences, 10*(1), 1–11. https://doi.org/10.30855/gmbd.0705N01

Geva, T., Martins, J. D., & Wald, R. M. (2014). Atrial septal defects. *The Lancet*, *383*(9932), 1921–1932. https://doi.org/10.1016/S0140-6736(13)62145-5

Hallıoğlu, O., Karpuz, D., Giray, D., Demetgül, H., & Öztaş, A. (2018). Frequency of the congenital heart diseases according to the risk groups: Fetal echocardiographic screening. *The Journal of Gynecology-Obstetrics and Neonatology*, *15*(1), 1–4.

Hassani, K., Jafarian, K., & Doyle, D. J. (2017). Heart sounds feature used for the classification of ventricular septal defect size in children. In *IFMBE Proceedings* (Vol. 61, pp. 28–31). Springer Verlag. doi:10.1007/978-981-10-4220-1_6

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). IEEE. doi: 10.1109/CVPR.2016.90

Kara, İ. (2023). Query by image examination: Classification of digital image-based forensics using deep learning methods. *Acta Infologica*, 7(2), 348-359. doi:10.26650/acin.1282567

Koundinya, T. S., Sudhanva, S., Shashank, C. S., Chandrasekar, S., & Munavalli, J. R. (2023). Congenital heart disease detection using spectral analysis and CNN. In *2023, 4th International Conference for Emerging Technology, INCET 2023*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/INCET57972.2023.10170429

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

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. doi: 10.1109/5.726791

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*(3), 211–252. doi:10.1007/s11263-015-0816-y

Sapitri, A. I., Nurmaini, S., Sukemi, Rachmatullah, M. N., & Darmawahyuni, A. (2021). Segmentation atrioventricular septal defect by using convolutional neural networks based on U-NET architecture. *IAES International Journal of Artificial Intelligence*, *10*(3), 553–562. doi: 10.11591/ijai.v10.i3.pp553-562

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 618–626). IEEE. doi: 10.1109/ICCV.2017.74

Sevli, O. (2023). A deep learning-based classification study for diagnosing corneal ulcers on ocular staining images. *Acta Infologica*, 7(2), 281-292. https://doi.org/10.26650/acin.1173465

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations (ICLR 2015)*, 1–14. Retrieved from http://arxiv.org/abs/1409.1556

Sun, S., Jiang, Z., Wang, H., Fang, Y., & Tao, T. (2012). Heart sound feature parameter distribution and support vector machine-based classification boundary determination method for ventricular septal defect auscultation. *Journal of Computational Science and Technology*, *6*(3), 198–206. doi:10.1299/jcst.6.198

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2818–2826). IEEE. https://doi.org/10.1109/CVPR.2016.308

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. *The 27th International Conference on Artificial Neural Networks (ICANN 2018)*. Retrieved from http://arxiv.org/abs/1808.01974

Tay Lik Wui, E., Yip, J. W. L., & Li, W. (2011). Echocardiography. In *Diagnosis and Management of Adult Congenital Heart Disease* (pp. 28–43). Elsevier. https://doi.org/10.1016/B978-0-7020-3426-8.00005-8

Tüzün, B. N., & Özdemir, D. (2023). Classification of brain tumors using deep learning models. *Journal of Scientific Reports-A*, 054, 296–306. doi: 10.59313/jsr-a.1293119

Wang, J.-K., Chang, Y.-F., Tsai, K.-H., Wang, W.-C., Tsai, C.-Y., Cheng, C.-H., & Tsao, Y. (2020). Automatic recognition of ventricular septal defect murmurs using convolutional recurrent neural networks with temporal attentive pooling. *Scientific Reports*, *10*(1), 21797. doi:10.1038/s41598-020-77994-z

Xu, X., Wang, T., Zhuang, J., Yuan, H., Huang, M., Cen, J., ... Shi, Y. (2021). ImageCHD: A 3D computed tomography image dataset for classification of congenital heart disease. *Segmentation, Classification, and Registration of Multi-modality Medical Imaging Data*, 4, 77–87. Retrieved from http://arxiv.org/abs/2101.10799

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, *9*(4), 611–629. doi:10.1007/s13244-018-0639-9

Yang, Y., Wu, B., Wu, H., Xu, W., Lyu, G., Liu, P., & He, S. (2023). Classification of normal and abnormal fetal heart ultrasound images and identification of ventricular septal defects based on deep learning. *Journal of Perinatal Medicine*, *51*(8), 1052–1058. doi:10.1515/jpm-2023-0041

Yıldız, J., Çetin, İ. İ., Aktaş, D., Arı, M. E., Kocabaş, A., Ekici, F., & Şaylı, T. R. (2015). Is an echocardiographic evaluation necessary for all children with a cardiac murmur? *Turkish Journal of Pediatric Disease*. doi:10.12956/tjpd.2015.158