International Journal of
Applied Data Science in Engineering and Health
https://ijadseh.com

Detection and Diagnosis of Congenital Heart Disease from Chest X-Rays with Deep Learning Models

Seyedhassan Sharifi[1*], Ali Donyadadi[2]

[1] *Pediatric Cardiology Subspecialist, Day General Hospital, Iran (Drhsharif52@gmail.com)*

[2] *Islamic Azad University, South Tehran Branch, Tehran, Iran*

Received date: December 11, 2024; Accepted date: January 1, 2025

Abstract

Congenital heart disease (CHD) is a leading cause of morbidity and mortality in children, requiring early and accurate diagnosis for effective management. In this study, we employed advanced deep learning models—EfficientNetV2, ResNeSt, and MobileNetV4—to classify CHD using chest X-ray (CXR) images. A dataset comprising 828 images, categorized into normal, atrial septal defect (ASD), ventricular septal defect (VSD), and patent ductus arteriosus (PDA), was utilized. The dataset was split into training, validation, and test sets in a stratified manner to ensure balanced evaluation. EfficientNetV2 achieved the best performance, with an accuracy of 92.5%, precision of 90.9%, sensitivity of 84.1%, and specificity of 88.6%, demonstrating its reliability for CHD diagnosis. ResNeSt closely followed, with an accuracy of 91.7% and precision of 90.4%, while MobileNetV4, though slightly less accurate at 88.6%, offered a lightweight alternative for resource-constrained environments.

Keywords*:* Congenital Heart Disease; Patent Ductus Arteriosus; Deep Learning.

Introduction

Congenital heart disease (CHD) is a collective term for structural and functional abnormalities in the heart and great vessels that are present from birth. It is among the most prevalent birth defects, affecting approximately 8 in every 1,000 live births globally, and represents a significant cause of infant morbidity and mortality [1,2]. CHD varies widely in its presentation, ranging from mild defects that may remain asymptomatic to severe malformations requiring immediate medical intervention. Three common forms of CHD include atrial septal defect (ASD), ventricular septal defect (VSD), and patent ductus arteriosus (PDA). ASD is characterized by an opening in the septum between the atria, which leads to increased pulmonary blood flow and may result in complications such as stroke, pulmonary hypertension, and heart failure. Similarly, VSD involves a defect in the interventricular septum, causing oxygen-rich blood to mix with oxygen-poor blood, leading to symptoms like fatigue, poor growth, and respiratory distress. PDA occurs when the ductus arteriosus, a vital fetal blood vessel, fails to close after birth, allowing abnormal circulation between the aorta and pulmonary artery, which can result in heart murmurs, pulmonary congestion, and eventual heart failure. The clinical management of CHD depends on its type and severity, ranging from pharmacological interventions like diuretics and ACE inhibitors to more invasive approaches such as catheterization and surgical correction. Early and accurate diagnosis is paramount to ensuring timely treatment and preventing irreversible complications, particularly in resource-constrained settings where advanced diagnostic tools may not be readily available [3,4].

Chest X-rays (CXR) remain an essential imaging modality in medical images analysis [5-8]. In pediatric cardiology, especially in regions where advanced diagnostic tools like echocardiography, MRI, or CT are unavailable. CXRs offer a non-invasive, cost-effective, and rapid means of evaluating cardiac and pulmonary anatomy, making them invaluable in the initial assessment of suspected CHD cases. For conditions like ASD, VSD, and PDA, CXRs can reveal key features such as cardiomegaly (enlarged heart silhouette), pulmonary vascular congestion, and abnormal positioning of cardiac or vascular structures [9, 10]. These findings provide critical insights that guide further diagnostic and therapeutic steps. Despite their utility, the accurate interpretation of CXRs requires specialized training, as subtle anomalies can be easily overlooked. This challenge is particularly pronounced in underdeveloped regions where skilled radiologists may be scarce. Consequently, there is an urgent need for automated tools capable of supporting clinicians by accurately detecting and highlighting potential abnormalities in CXR images, thereby bridging the gap in diagnostic expertise and improving patient care.

The rapid advancement of artificial intelligence (AI) and machine learning (ML) has transformed the landscape of medical imaging, enabling automated, efficient, and highly accurate analysis of complex datasets. Among these, deep learning (DL), a subset of ML, has emerged as a powerful approach for analyzing medical images, owing to its ability to learn and extract intricate patterns from data. Convolutional neural networks (CNNs), a specialized type of DL architecture, have demonstrated remarkable success in tasks such as image classification [11-15], segmentation [16,17], and anomaly detection [18-20]. When applied to CHD diagnosis, DL models can process large volumes of CXR data to identify subtle abnormalities indicative of conditions like ASD, VSD, and PDA with a level of sensitivity and specificity that rivals or even surpasses human expertise. Moreover, these models can generate attention heat maps, which visually highlight regions of interest, such as the heart and pulmonary vasculature, enhancing the interpretability of their predictions and fostering trust among clinicians. By automating the detection and classification of CHD from CXRs, DL models hold the potential to revolutionize diagnostic workflows, particularly in underdeveloped regions where access to specialized care is limited. This study aims to harness the power of DL to develop an accessible and interpretable diagnostic system, ultimately improving early detection and treatment outcomes for children with CHD.

In this paper, we employ convolutional neural network (CNN)-based deep learning models to perform the classification of normal and three different abnormal hearts using chest X-ray images. The proposed approach focuses on leveraging the rich features extracted by CNNs to accurately distinguish between healthy and CHD-affected hearts, thereby laying the foundation for an efficient, scalable, and interpretable diagnostic tool that can assist clinicians worldwide.

Methods and Materials

Dataset

The dataset used in this study comprises 828 chest X-ray (CXR) images, categorized into four distinct groups to facilitate the classification of congenital heart disease (CHD) [21]. The first group includes normal images, representing individuals without any structural or functional heart abnormalities, which serve as a baseline for comparison. The second group contains X-rays of individuals diagnosed with atrial septal defect (ASD), characterized by features such as an enlarged heart silhouette and increased pulmonary blood flow. The third group consists of images exhibiting ventricular septal defect (VSD), where abnormalities include mixed oxygenated and deoxygenated blood flow leading to pulmonary congestion and cardiomegaly. The final group includes X-rays associated with patent ductus arteriosus (PDA), which show signs of abnormal blood flow between the aorta and pulmonary artery, often resulting in pulmonary vascular changes. This well-annotated dataset provides a robust foundation for training and evaluating the proposed deep-learning models in distinguishing normal and abnormal heart conditions.

Deep Learning Models

Deep learning is a powerful subset of artificial intelligence that relies on neural networks to extract and learn hierarchical features from data. At its foundation are artificial neural networks (ANNs), which consist of layers of interconnected nodes (neurons) that process input data through weights and activation functions. Deep neural networks (DNNs) extend this concept by incorporating multiple hidden layers, enabling the model to learn increasingly abstract and complex representations of the data. In particular, convolutional neural networks (CNNs), a specialized class of DNNs, have proven highly effective for image analysis tasks due to their ability to leverage spatial hierarchies within data. A CNN typically comprises convolutional layers, pooling layers, and fully connected layers. Convolutional layers apply trainable filters (kernels) that slide over the input image to detect local patterns, such as edges, textures, and shapes. These filters allow the network to preserve spatial relationships while identifying critical features. Pooling layers reduce the spatial dimensions of the data through operations like max-pooling or average-pooling, which enhance computational efficiency and make the model more robust to variations, such as scale or orientation. Finally, fully connected layers aggregate learned features to make final predictions.

In the context of medical imaging, CNNs are particularly adept at recognizing subtle and complex patterns associated with pathological states. For example, in chest X-rays (CXR), CNNs can detect features such as cardiac enlargement, abnormal pulmonary vasculature, or structural anomalies indicative of congenital heart disease (CHD). These patterns may span various scales, from fine-grained textures to large-scale structural changes. By training CNNs on labeled datasets, the models can learn to differentiate between normal and abnormal conditions, leveraging feature extraction and classification capabilities. Deep learning models benefit significantly from large datasets, where diverse examples enable the network to generalize effectively to unseen data. During training, the model adjusts its weights iteratively through backpropagation and gradient descent, minimizing a loss function that quantifies the error between predictions and ground truth labels. The use of regularization techniques (e.g., dropout and weight decay) and data augmentation (e.g., rotations, flips, and intensity adjustments) further ensures robustness and reduces the risk of overfitting.

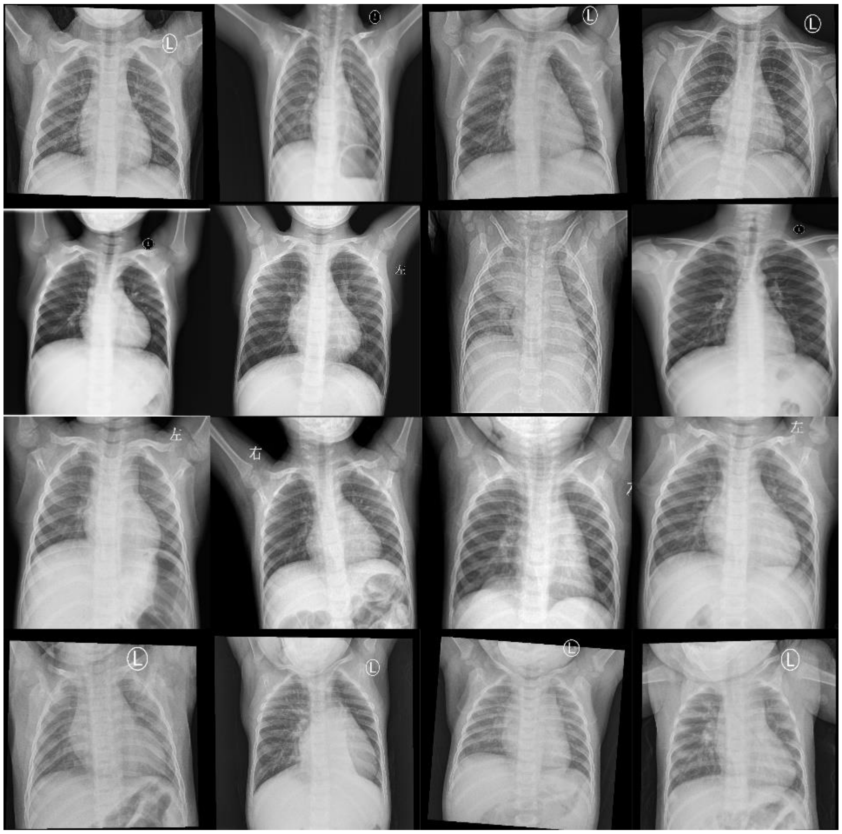

Figure 1 congenital heart diseases. first row: normal cases. second row: ASD. third row: PDA. third row: VSD [21]

By employing CNNs for medical image analysis, this study harnesses the network's ability to analyze and interpret complex patterns in CXRs. The model's hierarchical feature extraction pipeline enables it to identify both global structural abnormalities, such as cardiomegaly, and localized changes, like pulmonary congestion or septal defects. This capability forms the backbone of the proposed system for detecting and diagnosing congenital heart disease.

ResNeSt (Split-Attention Networks)

ResNeSt, short for ResNet Split-Attention Networks [22], is an enhancement of the classic ResNet architecture that introduces a novel split-attention mechanism to better capture multi-scale features in image data. The core innovation lies in the split-attention block, which divides the feature maps in each residual block into multiple splits. Each split is processed independently, with a dedicated attention mechanism applied to focus on the most relevant features within that split. The outputs from these splits are then aggregated to form the final feature representation for the block. This process enables the model to adaptively prioritize features at different spatial and channel scales, which is critical for tasks that require a detailed understanding of both global and local patterns.

Architecturally, ResNeSt retains the core components of ResNet, including skip connections and stacked residual blocks, ensuring compatibility with existing frameworks. However, it extends these components by incorporating the split-attention blocks. Each block includes three main steps: splitting the input feature maps, applying attention to each split, and combining the attended features.

This structure improves the model's representational capacity while maintaining computational efficiency. The scalability of ResNeSt allows it to perform well on both small-scale and large-scale datasets, making it ideal for high-precision classification, object detection, and medical imaging tasks like detecting abnormalities in chest X-rays. Its multi-scale feature extraction capability is especially valuable in applications where patterns of interest vary in size, shape, and position.

MobileNetV4

MobileNetV4 [23] represents the latest iteration in the MobileNet family, specifically optimized for use on mobile and edge devices. Its architecture builds upon the success of its predecessors (MobileNetV2 and V3) by improving both feature extraction efficiency and computational performance. The cornerstone of MobileNetV4 is the inverted bottleneck block, a design that uses an initial expansion layer to increase feature dimensionality, followed by a depthwise convolution to reduce spatial complexity, and finally a compression layer to project features back into a smaller space. This structure balances computational efficiency with the ability to capture meaningful patterns.

One key architectural improvement in MobileNetV4 is the introduction of multi-query attention modules, which enhance the model's ability to selectively focus on critical features. Unlike conventional attention mechanisms, these modules operate efficiently on low-power devices without significantly increasing model complexity. MobileNetV4 also standardizes the design of its inverted bottlenecks across layers, simplifying the model while improving consistency in feature representation. Additionally, MobileNetV4 includes optimizations for fused operations that combine convolutional and activation layers, reducing the number of operations and improving inference speed. The architecture of MobileNetV4 is particularly suited for scenarios with limited computational resources, such as mobile applications, embedded systems, and edge computing. It is widely used in real-time image classification, face recognition, and object detection tasks, where low latency and energy efficiency are critical requirements. The model's lightweight design and scalability make it a practical choice for deploying deep learning on devices with hardware constraints.

EfficientNetV2

EfficientNetV2 [24] builds on the original EfficientNet series, which introduced a systematic approach to scaling model size, depth, and input resolution. The key advancement in EfficientNetV2 lies in its ability to train faster and achieve higher accuracy through architectural optimizations and enhanced training techniques. One of the core architectural changes is the replacement of traditional MBConv blocks with Fused-MBConv blocks, which combine depthwise convolutions and pointwise convolutions into a single fused operation. This reduces computational overhead while preserving the ability to learn spatial and channel-wise features effectively. EfficientNetV2 also incorporates progressive learning, where the model is trained on progressively larger input sizes and more complex augmentations. This approach allows the model to learn robust features incrementally, improving generalization on large-scale datasets. Another key enhancement is the use of stochastic depth and RandAugment, which introduce random variations during training to reduce overfitting and increase robustness. The architecture further optimizes parameter usage by carefully balancing the number of filters, block depths, and input resolutions through a compound scaling formula.

EfficientNetV2 achieves state-of-the-art performance on benchmarks such as ImageNet while maintaining lower computational requirements compared to previous architectures. Its ability to handle large datasets efficiently makes it a popular choice for applications like medical imaging, where high-resolution data and accurate feature extraction are critical. For example, in tasks like disease diagnosis or organ segmentation, EfficientNetV2 can analyze large volumes of medical images with a high degree of precision and speed.

Experiments

Training and Configuration

The training process for the deep learning models is centered around optimizing the network's parameters to minimize the error between predictions and ground truth labels. This is achieved through backpropagation and the use of an optimizer, specifically the Adam optimizer, which combines momentum and adaptive learning rates for efficient convergence. The loss function employed is categorical cross-entropy, as it is well-suited for multi-class classification tasks. To improve generalization and prevent overfitting, regularization techniques such as dropout, which randomly deactivates neurons during training, and weight decay (L2 regularization), which penalizes large weights, are applied. Additionally, data augmentation is extensively utilized to artificially expand the dataset by applying transformations such as rotation, flipping, scaling, and brightness adjustments. The training process also incorporates early stopping, monitoring the validation loss to halt training once the model's performance ceases to improve, further mitigating overfitting risks.

The dataset consists of 828 chest X-ray (CXR) images, divided into four classes: normal, atrial septal defect (ASD), ventricular septal defect (VSD), and patent ductus arteriosus (PDA). To ensure robust training and evaluation, the dataset is split into three subsets: 70% for training, 20% for validation, and 10% for testing. This stratified split maintains the proportional representation of each class across all subsets, ensuring balanced learning and evaluation. The training set is used to update the model weights, while the validation set monitors performance and aids in hyperparameter tuning. Finally, the test set provides an unbiased assessment of the model's generalization capabilities.

In terms of configuration, all input images are resized to a consistent dimension suitable for the chosen convolutional neural network (CNN) architecture. The batch size, learning rate, and number of epochs are carefully tuned to balance computational efficiency and performance. Training is performed using GPUs to accelerate computation, and checkpoints are saved periodically to preserve the best-performing model. This systematic approach ensures that the deep learning models are trained effectively to classify normal and abnormal heart conditions with high accuracy.

Results

The classification performance of the three deep learning models—ResNeSt, MobileNetV4, and EfficientNetV2—was evaluated using accuracy, precision, sensitivity, and specificity, providing a comprehensive assessment of their capabilities in diagnosing congenital heart disease. Accuracy measured the proportion of correctly classified cases, both normal and abnormal, out of the total number of samples, providing an overall assessment of model performance. Precision reflected the proportion of true positive predictions among all positive predictions, highlighting the model's ability to correctly identify abnormal cases without generating false positives.

Table1. Classification performance of deep learning models for congenital heart disease.

Model	Accuracy	Precision	Sensitivity	Specificity
ResNeSt	91.7%	90.4%	84.2%	85.1%
MobileNetV4	88.6%	85.1%	79.9%	80.4%
EfficientNetV2	92.5%	90.9%	84.1%	88.6%

Sensitivity, or recall, assessed the proportion of actual positive cases (abnormal hearts) that were correctly identified by the model, indicating its effectiveness in minimizing false negatives. Specificity evaluated the proportion of true negative cases among all actual negative cases, emphasizing the model's capability to accurately identify normal cases without false positives. Together, these metrics offered a comprehensive evaluation of the models' strengths and weaknesses, enabling a detailed comparison of the performance of EfficientNetV2, ResNeSt, and MobileNetV4 in diagnosing congenital heart disease from chest X-ray images.

EfficientNetV2 demonstrated the best overall performance with an accuracy of 92.5%, closely followed by ResNeSt at 91.7%, while MobileNetV4 lagged behind at 88.6%. This trend highlights the robustness of EfficientNetV2 and ResNeSt in correctly classifying both normal and abnormal heart conditions. In terms of precision, which measures the reliability of identifying abnormal cases, EfficientNetV2 achieved the highest score at 90.9%, slightly surpassing ResNeSt's 90.4%, with MobileNetV4 trailing at 85.1%, suggesting a higher propensity for false positives in the latter.
Sensitivity, an indicator of the models' ability to detect true abnormal cases, showed comparable performance between ResNeSt and EfficientNetV2, with values of 84.2% and 84.1%, respectively. MobileNetV4 exhibited the lowest sensitivity at 79.9%, indicating a higher likelihood of missing some abnormal cases. Specificity, which reflects the models' accuracy in identifying true normal cases, was highest for EfficientNetV2 at 88.6%, followed by ResNeSt at 85.1%, while MobileNetV4 again recorded the lowest performance at 80.4%. Overall, EfficientNetV2 emerged as the most balanced and reliable model across all metrics, followed closely by ResNeSt, while MobileNetV4 displayed relatively lower performance, particularly in terms of sensitivity and specificity. These results emphasize the superior capability of EfficientNetV2 and ResNeSt in diagnosing congenital heart disease from chest X-rays.

Future Works

The promising results achieved by the deep learning models in this study lay the foundation for further advancements in the automated diagnosis of congenital heart disease (CHD). Future work can explore several directions to enhance the robustness, interpretability, and clinical applicability of these models. One key area of improvement is expanding the dataset to include a larger and more diverse population, encompassing various age groups, ethnicities, and imaging conditions. A more extensive dataset would improve the generalizability of the models and ensure reliable performance across diverse clinical scenarios.
Another potential direction involves integrating multi-modal data into the diagnostic pipeline. Combining chest X-ray (CXR) images with other clinical data, such as echocardiography results, patient demographics, and laboratory findings, could provide a more holistic understanding of the patient’s condition and improve diagnostic accuracy. Additionally, exploring advanced architectures, such as hybrid models combining convolutional neural networks (CNNs) with transformers, may further enhance feature extraction and classification capabilities.

Improving model interpretability is another crucial aspect of future work. Generating more detailed and clinically relevant attention heat maps or saliency maps could help clinicians better understand the rationale behind the model's predictions, fostering trust and facilitating adoption in real-world settings. Furthermore, deploying these models in resource-limited environments will require optimization for low-power devices, such as mobile or edge computing platforms, to ensure accessibility and scalability.

Finally, clinical validation is an essential step to transition these models from research to practice. Conducting prospective studies in real-world healthcare settings will provide insights into the models' performance in aiding clinicians and improving patient outcomes. These efforts will pave the way for developing comprehensive, AI-driven diagnostic systems that can address the global burden of CHD, especially in underdeveloped regions where access to specialized care is limited. Through these advancements, deep learning-based tools can significantly enhance the early detection and management of CHD, contributing to improved healthcare equity and outcomes worldwide.

Conclusion

In this study, we evaluated the performance of three advanced deep learning models—EfficientNetV2, ResNeSt, and MobileNetV4—for classifying congenital heart disease (CHD) using chest X-ray images. The results demonstrated that EfficientNetV2 achieved the highest overall accuracy, precision, and specificity, establishing it as the most effective model for CHD diagnosis in this context. ResNeSt closely followed, showcasing robust performance in accurately classifying normal and abnormal heart conditions. MobileNetV4, while slightly less accurate, proved to be a viable option for environments with limited computational resources due to its lightweight architecture. Despite the promising results, the study highlighted areas for future work, such as expanding the dataset to improve model generalization, integrating multi-modal data for more comprehensive diagnostics, and enhancing model interpretability through attention heat maps.

Conflict of Interest

The authors imply no conflict of interest.

References

[1] Hoffman JI, Kaplan S. The incidence of congenital heart disease. Journal of the American college of cardiology. 2002 Jun 19;39(12):1890-900.

[2] Hoffman JI, Kaplan S, Liberthson RR. Prevalence of congenital heart disease. American heart journal. 2004 Mar 1;147(3):425-39.

[3] Sun R, Liu M, Lu L, Zheng Y, Zhang P. Congenital heart disease: causes, diagnosis, symptoms, and treatments. Cell biochemistry and biophysics. 2015 Jul;72:857-60.

[4] Leusveld EM, Kauling RM, Geenen LW, Roos-Hesselink JW. Heart failure in congenital heart disease: management options and clinical challenges. Expert review of cardiovascular therapy. 2020 Aug 2;18(8):503-16.

[5] Afrazeh F. Advances in Imaging Analysis for Understanding Brain Disorders. International Journal of Scientific and Applied Research (IJSAR), eISSN: 2583-0279. 2024 Nov 24;4(8):1-8.

[6] Huda W, Abrahams RB. X-ray-based medical imaging and resolution. American Journal of Roentgenology. 2015 Apr;204(4):W393-7.

[7] Kasban H, El-Bendary MA, Salama DH. A comparative study of medical imaging techniques. International Journal of Information Science and Intelligent System. 2015 Apr;4(2):37-58.

[8] Afrazeh F, Ghasemi Y, Shomalzadeh M, Rostamian S. The Role of Imaging Data from Different Radiologic

Modalities During the Previous Global Pandemic. International Journal of Applied Data Science in Engineering and Health. 2024 Jul 27;1(1):9-17.
[9] Taylor AM, Thorne SA, Rubens MB, Jhooti P, Keegan J, Gatehouse PD, Wiesmann F, Grothues F, Somerville J, Pennell DJ. Coronary artery imaging in grown up congenital heart disease: complementary role of magnetic resonance and x-ray coronary angiography. Circulation. 2000 Apr 11;101(14):1670-8.
[10] Perloff JK. The Chest X-Ray in Congenital Heart Disease. Adult Congenital Heart Disease. 2009 Apr 17:208-20.
[11] Orouskhani M, Zhu C, Rostamian S, Zadeh FS, Shafiei M, Orouskhani Y. Alzheimer's disease detection from structural MRI using conditional deep triplet network. Neuroscience Informatics. 2022 Dec 1;2(4):100066.
[12] Minoo S, Ghasemi F. A Narrative Review: Dental Radiology with Deep Learning. International Research in Medical and Health Sciences. 2024 Oct 25;7(5):23-36.
[13] El Asnaoui K. Design ensemble deep learning model for pneumonia disease classification. International Journal of Multimedia Information Retrieval. 2021 Mar;10(1):55-68.
[14] Abbasi H. Transfer Learning and Advanced CNN Models for Detecting Brain Tumors using MRI. International Journal of Scientific and Applied Research (IJSAR), eISSN: 2583-0279. 2024 Dec 13;4(9):92-103.
[15] Norouzi F, Machado BL, Nematzadeh S. Schizophrenia Diagnosis and Prediction with Machine Learning Models. International Journal of Scientific and Applied Research (IJSAR), eISSN: 2583-0279. 2024 Dec 13;4(9):113-22.
[16] Mzoughi O, Yahiaoui I. Deep learning-based segmentation for disease identification. Ecological Informatics. 2023 Jul 1;75:102000.
[17] Livne M, Rieger J, Aydin OU, Taha AA, Akay EM, Kossen T, Sobesky J, Kelleher JD, Hildebrand K, Frey D, Madai VI. A U-Net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease. Frontiers in neuroscience. 2019 Feb 28;13:97.
[18] Abbasi H, Afrazeh F, Ghasemi Y, Ghasemi F. A Shallow Review of Artificial Intelligence Applications in Brain Disease: Stroke, Alzheimer's, and Aneurysm. International Journal of Applied Data Science in Engineering and Health. 2024 Oct 5;1(2):32-43.
[19] Pang G, Shen C, Cao L, Hengel AV. Deep learning for anomaly detection: A review. ACM computing surveys (CSUR). 2021 Mar 5;54(2):1-38.
[20] Mahmoudiandehkordi S, Yeganegi M. Transforming Gynecology with Artificial Intelligence: Advances in Clinical Practice. International Journal of Scientific and Applied Research (IJSAR), eISSN: 2583-0279. 2024 Dec 13;4(9):104-12.
[21] https://kaggle.com/competitions/congenital-heart-disease, 2023. Kaggle.
[22] Zhang H, Wu C, Zhang Z, Zhu Y, Lin H, Zhang Z, Sun Y, He T, Mueller J, Manmatha R, Li M. Resnest: Split-attention networks. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2022 (pp. 2736-2746).
[23] Howard A, Sandler M, Chu G, Chen LC, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V, Le QV. Searching for mobilenetv3. InProceedings of the IEEE/CVF international conference on computer vision 2019 (pp. 1314-1324).
[24] Tan M, Le Q. Efficientnetv2: Smaller models and faster training. InInternational conference on machine learning 2021 Jul 1 (pp. 10096-10106). PMLR.