International Journal of Science, Strategic Management and Technology
Volume 02 Issue 03 March-2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Automated Myocardial Infarction Classification from Cardiac MRI using Vision Transformer Architecture

[1] **Mrs.C.USHARANI**
Dept. of Artificial Intelligence and Data Science
Ramco Institute of Technology,
Rajapalayam.
cusha91@gmail.com

R. PRADEEPA [2]
Dept. of Artificial Intelligence and Data Science,
Ramco Institute of Technology,
Rajapalayam
rpradeepa2605@gmail.com

[3] **Dr.M.KALIAPPAN**
Dept. of Artificial Intelligence and Data Science
Ramco Institute of Technology,
Rajapalayam.
kalsrajan@yahoo.co.in

[4] **Dr.E.MARIAPPAN**
Dept. of Artificial Intelligence and Data Science,
Ramco Institute of Technology,
Rajapalayam
mapcse@gmail.com

https://doi.org/10.55041/ijsmt.v2i3.248

Cite this Article: PRADEEPA, R. & C.USHARANI, (2026). Deep Learning Framework for Myocardial Infarction Diagnosis from Cardiac MRI using Vision Transformers. International Journal of Science, Strategic Management and Technology, *02*(03). https://doi.org/10.55041/ijsmt.v2i3.248

ABSTRACT:

One of the main causes of death globally is a myocardial infarction, also referred to as a heart attack, which is a sudden reduction in the heart's blood flow. Survival rates are significantly increased by early identification of this illness. Because MRI can reveal the heart’s internal structure with high clarity, it can detect damaged heart muscle that standard tests may miss. However, manually analyzing large numbers of MRI images is time-consuming and heavily dependent on the observer’s expertise. In this work, an automated system for myocardial infarction detection from cardiac MRI images is proposed using a Vision Transformer (ViT) based deep learning model. The EMIDEC cardiac MRI dataset, which includes both healthy subjects and patients with myocardial infarction, is used in this study.To make key cardiac structures more visible, the MRI volumes are first transformed into two-dimensional slices and then preprocessed using noise reduction, contrast enhancement, scaling, and normalization. Further visual analysis with ROC curves and confidence-based assessment bolsters the system's efficacy and stability. The system has a high test accuracy of 97.63%, which shows that most of the MRI slices were correctly identified. This proposed system shows that Vision Transformer-based models can be effectively used for automatic myocardial infarction detection from cardiac MRI images, providing doctors with a useful tool for decision-support and facilitating prompt and precise diagnosis.

1. INTRODUCTION:

Cardiovascular diseases are and continue to be one of the leading causes of death globally and have caused millions to suffer over the years, thus becoming a great burden to all health facilities. However, the most critical and life-threatening heart disease, as discussed, refers to myocardial infarction, otherwise known as a heart attack, described as a condition when scarring of the heart occurs as a result of a blockage of blood flow to the heart muscles.This kind of condition, if not promptly diagnosed and managed, may cause serious detrimental effects, for example, cardiac failure. The early and correct diagnosis of this condition is critical to the prevention of myocardial infarction for the improved survival of patients.Medical imaging is an important aspect of diagnostic medicine used to detect various heart conditions. Among various medical imaging modalities, Cardiac Magnetic Resonance Imaging (MRI) is rated as one of the most reliable imaging modalities for cardiac imaging. Cardiac MRI offers a clear image of the heart's condition by depicting the heart muscles, blood circulation, and damaged areas of the heart through very high-resolution images. It is most appropriate for detecting myocardial infarction. During a regular clinical setting, a medical imaging radiologist undergoes analysis of the MRI images to detect heart conditions, determine the severity of damaged heart tissues, and recommend the most appropriate treatment plan. But the examination of the MRI images is time-consuming, demanding a lot of expertise from the medical imaging radiologists involved. Additionally, determination of the appropriate treatment is subjective to the radiologists involved, making decisions sometimes inconsistent.

Therefore, these problems have led researchers to incorporate artificial intelligence and deep learning techniques to improve medical image analysis. Recently, convolutional neural networks (CNNs) have been successfully applied to several medical image analysis problems, including image classification, segmentation, and disease detection. However, these models have mainly dealt with local information provided by an image, but not long-range information. This might be a limitation when dealing with medical images like MRI images of the heart, where long-range information is important for a better diagnosis. More recently, Vision Transformers, abbreviated as ViTs, have been proposed as a strong architecture compared to the traditional convolution-based CNNs for image-related tasks. Unlike the traditional CNNs, ViTs make use of a self-attention mechanism that allows the model to learn the relationship between different parts of the image. Thus, it is capable of learning both the local and global characteristics of an image. In practice, image-related tasks, especially in the domain of medicine, involve the analysis of patterns that appear across different parts of an image to detect diseases. Therefore, the Vision Transformers can be utilized for image-related tasks in the medical domain.

In the current work, we introduce an automated system that enables the detection of myocardial infarction by using images of the heart through MRI scans, and the algorithm employed in the current work is based on the Vision Transformer model. The primary aim of the current work is to provide an efficient and effective deep learning framework that assists in the classification of MRI images as normal or abnormal, which further assists the early diagnosis of myocardial infarction among affected patients. This has been achieved through the proper application of imaging techniques.

2. RELATED WORKS:

Medical imaging with the incorporation of AI has transformed the diagnosis and management of cardiovascular diseases. Amongst various imaging techniques, CMR is generally considered a very trustworthy method for the assessment of MI due to its provision of high-resolution details on cardiac structure, tissue characteristics, and blood flow. However, manual interpretation of CMR images is time-consuming and heavily reliant on clinical experience, which may result in inter-observer variations and delayed diagnosis. Many researchers have hence explored automated detection methods based on machine learning and deep learning for MI.Initially, there was a lot of research work done on traditional image processing techniques and handcrafted features that were computed from cardiac MRI images. These techniques relied on texture features and shape features along with intensity features. They also used machine learning techniques like SVM and RF classifiers. These techniques worked well for some cases but had limited potential as handcrafted features tend to perform badly under varying image quality and patient anatomy. Progress in the field of deep learning has enabled convolution neural networks to become the widely used techniques for image analysis, as machines have the capability to learn hard features from images directly [2].

VGGNet [6] and ResNet [5], which are based on CNNs, have been successfully applied in various medical imaging analysis areas, including disease classification, segmentation, and detection. For cardiac MRI data analysis, CNNs could be applied to identify myocardial infarction based on their learned spatial patterns of damage or scarring. Tao et al. [13] introduced their deep learning approach to diagnose MI from cardiac MRI images, confirming that CNN-based methods perform better compared to other machine learning methods. Bernard et al. [12] also applied deep neural networks to perform segmentation of structures in the heart, also emphasizing the effectiveness of deep learning in heart disease diagnosis.

There have also been several studies to improve the segmentation of the cardiac structures, as it will eventually lead to better MI detection. U-Net [4] was found to be one of the popular architectures in biomedical image segmentation due to their encoder-decoder structure with skip connections that maintain spatial information in the images. Further improvements to the U-Net were made by developing the Attention U-Net [10] and the nnU-Net [9]. It is seen that accurate segmentation of the images will help to extract the myocardium, which is necessary to detect the infarcted areas in the MRI images of the heart. Although CNNs succeed in their application, they have some limitations. CNNs traditionally concentrate on local features; however, they fail to capture long-range dependencies in an image. The latter drawback is very critical in cardiac MRI scans because infarct patterns may exist distantly in image regions. To eliminate the CNN limitations, a new model based on Transformers, which is initially designed for natural language processing tasks, is now applied in computer vision. The Vision Transformers (ViTs) process images as sequences of pixels and make use of self-attention to capture overall relationships between image regions [1]. Dosovitskiy et al. [1] showed that Vision Transformers can perform equally well or better than CNNs on large-scale image classification problems. Encouraged by this, some researchers began to explore Transformers on medical image processing tasks. However, a new model, TransUNet [7], combined CNNs with Transformers to better solve medical image segmentation problems, proving that Transformers are effective for representing global contexts of medical images. Another model, Swin-UNETR [8], achieved good performance on MRI-based medical image segmentation, especially for brain tumor and cardiac image analysis, with a hierarchical transformer structure.

There have also been applications for transformers in medical image classification problems. Surveys conducted by Azad et al. [15] and Shamshad et al. [16] showed that transformers perform better than traditional CNNs for a wide range of medical imaging problems, especially where the global context is critical. Gao et al. [17] demonstrated that the ViT family produces more stable classification results than traditional CNN architectures for medical image classification problems.In the field of cardiac MRI, hybrid models combining the power of CNNs and Transformers have been proposed to leverage the strengths of both architectures. In the paper by Wang et al. [18], the authors proposed a hybrid model combining the Transformer and CNN architectures for the task of medical image classification. In the specific case of myocardial infarction detection, both texture and spatial features may play a crucial role. This has been further accelerated with the availability of benchmark datasets. The dataset of the EMIDEC [21, 22] was made available; it consisted of labeled cardiac MRI from both healthy patients and those suffering from myocardial infarction. This dataset has then been used to develop and evaluate a series of deep learning approaches for the detection of MI. Each of those that utilized the dataset described herein demonstrated that these deep learning models are capable of identifying infarcted regions and characterizing patients as normal or abnormal with high accuracy. Overall, existing studies clearly indicate that deep learning, especially Transformer-based architectures, has great potential for the automated detection of myocardial infarction from cardiac MRI. However, many earlier works have focused primarily on CNNs or segmentation-based approaches. So far, the use of Vision Transformers for direct MI classification is quite new and presents significant scope for further improvement. This motivates the present work to further investigate the use of a fine-tuned Vision Transformer in classifying cardiac MRI images as normal or abnormal, with a view to providing a sound and efficient decision-support tool for clinicians.

Recently, transformer models have been applied to the area of medical image analysis. Vision Transformers have shown excellent performance in medical image analysis tasks like image classification and segmentation, often surpassing CNNs when a large amount of training data is available. Some recent works have applied ViTs to medical datasets and shown improved accuracy and generalization. These results encourage the application of Vision Transformers for myocardial infarction detection in this research.

3. DATASET AND PREPROCESSING:

For this research, the EMIDEC cardiac MRI dataset is employed to create and test the proposed system for the detection of myocardial infarction. The EMIDEC dataset is a collection of cardiac MRI scans obtained from healthy participants and patients with a myocardial infarction diagnosis. This dataset is very useful because it contains a wide range of cases, which makes it an ideal choice for training and testing deep learning models in a real-world environment. The dataset

International Journal of Science, Strategic Management and Technology
Volume 02 Issue 03 March-2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

contains comprehensive scans of the heart, which enables the training of the model to distinguish between healthy and infarcted heart tissue.

The cardiac MRI scans in the EMIDEC dataset are originally stored as three-dimensional data. The data is comprised of several slices that represent various sections of the heart. However, deep learning models like Vision Transformers are designed to process two-dimensional images. Therefore, the 3D data is first transformed into a series of 2D slices. This process enables the model to examine the minute details of the slices while also being able to identify differences in various parts of the heart.Following the extraction of slices, a number of preprocessing techniques are carried out to improve the quality of the images. The first preprocessing technique is noise removal. This is useful in eliminating random variations that may be present in the images during the MRI scan process. This technique improves the clarity of the images and ensures that the model does not learn irrelevant patterns from the images.

The second preprocessing technique is contrast enhancement. This is useful in ensuring that the critical structures of the heart are visible. This technique is particularly helpful in ensuring that the differences between the normal and infarcted parts of the heart are visible, as these parts are likely to have different intensity patterns in the MRI images. Finally, the normalization step is carried out to ensure that the pixel values are within a uniform range. This helps to ensure that differences in brightness or intensity between various MRI images do not hamper the learning process. This step also helps to ensure that the training of the model is faster and more stable.The combination of the above steps helps to ensure that the quality of the MRI images is greatly improved and that the input to the Vision Transformer is clean and standardized. This helps to ensure that the model is able to learn features that are relevant to myocardial infarction.

Dataset	Number of Images	Percentage (%)
Training Set	702	64.9 %
Test Set	380	35.1 %
Total	**1082**	**100 %**

Fig 1: Dataset Distribution

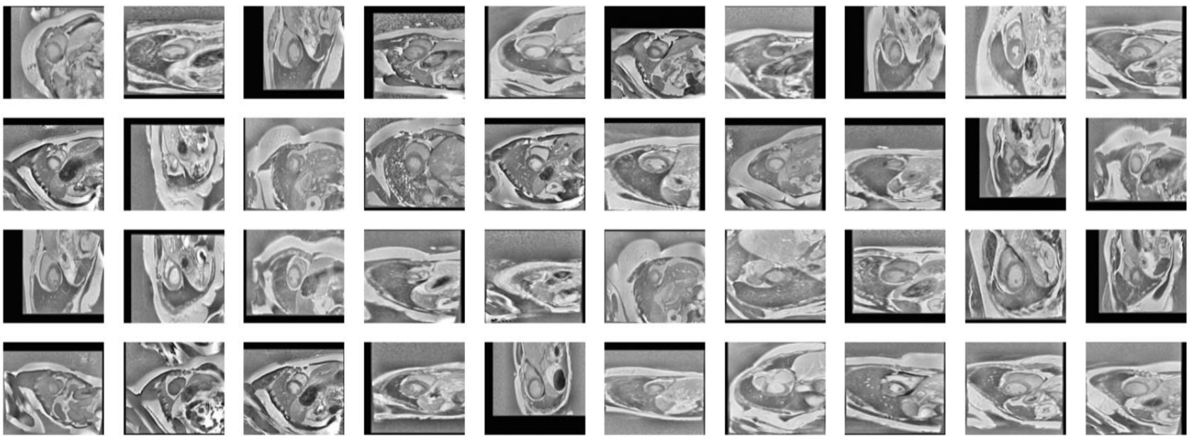

Fig.3. Sample Dataset

4. PROPOSED METHODOLOGY:

The proposed system for the detection of myocardial infarction is designed using a deep learning framework that is based on a Vision Transformer (ViT) architecture and is trained to predict two classes: normal and abnormal (myocardial infarction) based on the cardiac MRI slices. The overall approach for the proposed system includes data preparation, image processing, fine-tuning of the model, training, and evaluation.

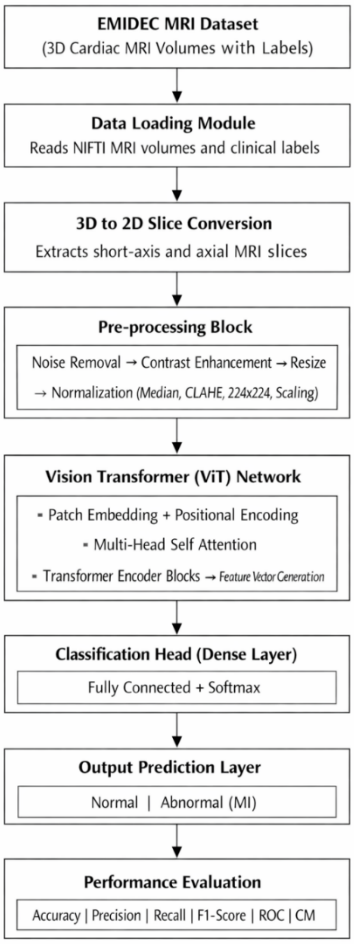

Fig 2: Proposed System Architecture

4.1 DATA PREPARATION:

The EMIDEC cardiac MRI dataset is employed as the main source of data for this research. This dataset comprises cardiac MRI images of healthy people and patients diagnosed with myocardial infarction. The images are delivered as three-dimensional volumes, with every volume comprising several cross-sectional slices of the heart.As deep learning models like Vision Transformers process two-dimensional images, every 3D MRI volume is initially transformed into a set of 2D slices. These slices deliver detailed information about the heart at various spatial points and contain enough information to detect myocardial damage. The slices are labeled based on the clinical diagnosis of the patient from whom the slices were derived, ensuring that every image is properly labeled as either normal or abnormal.To ensure consistency and prevent data imbalance, the dataset is carefully arranged to ensure that both normal and abnormal examples are adequately represented. This is an important step in helping the model learn relevant distinctions between healthy and infarcted heart tissues.

4.2. IMAGE PREPROCESSING:

Raw images obtained from the MRI may consist of noise, intensity variations, and artifacts due to patient movement during the scanning process. If these flaws are directly used as input to a deep learning model, they could adversely impact the performance of the model. To overcome this problem, a number of preprocessing steps are carried out to improve the quality of the images.The first step involves the use of noise reduction methods to eliminate random variations in the images. This process helps to make the boundaries of tissues and critical areas more distinct. The next step involves contrast enhancement to enhance the visibility of areas in the myocardium, especially those that are infarcted, which have intensity variations compared to normal areas.

Then, all images are resized to a fixed resolution corresponding to the input size required by the Vision Transformer. This helps ensure that the Vision Transformer model gets equal inputs, thereby stabilizing the learning process. Finally, the pixel values are normalized to ensure that the image values are scaled to a fixed range. This helps eliminate large differences in image brightness, thereby ensuring that the training process converges properly.By applying the above preprocessing techniques, the MRI slices are made clearer and more uniform, which helps the Vision Transformer learn features from them.

4.3. VISION TRANSFORMER MODEL:

A Vision Transformer (ViT) is employed as the backbone network for the new system. Unlike conventional convolutional neural networks, which rely on local filters to extract features, Vision Transformers treat an image as a sequence of small patches and examine the relationships among these patches using self-attention mechanisms. This enables the model to learn both local texture patterns and global structural patterns over the entire heart.The input MRI images are first split into fixed-size patches. The patches are flattened and projected into a high-dimensional embedding space. Positional embeddings are added to preserve the spatial information of the original image. This is important because the spatial configuration of cardiac structures is a critical factor in determining myocardial infarction.

The embeddings from the patch embeddings layer are then fed into multiple transformer encoder layers. Each transformer encoder layer consists of a multi-head self-attention block and a feed-forward neural network. The self-attention mechanism allows the model to compare every patch to every other patch, which is very useful in understanding how different parts of the heart relate to each other. This is especially useful in the identification of infarction, which may be present in multiple parts of the heart.

4.4 MODEL FINE-TUNING AND TRAINING:

The Vision Transformer is started with weights that have been pretrained on a large-scale image dataset. This is because the Vision Transformer is capable of using the knowledge that was acquired during the pretraining phase to perform well on the new task. This is referred to as transfer learning.During training, the preprocessed MRI images and their corresponding labels (normal or abnormal) are fed into the model. The result of the transformer is then passed through a classification head that predicts the probabilities of the two classes. A loss function, typically the cross-entropy loss function, is used to calculate the difference between the predicted labels and the actual labels. The model parameters are then updated using backpropagation to optimize the loss. The training procedure is tracked for overfitting by means of a validation set. Hyperparameters like learning rate, batch size, and number of epochs are adjusted to obtain the maximum possible performance.

4.5 PERFORMANCE EVALUATION:

After the completion of the training phase, the model is then tested on a test dataset that was not used in the training phase. The output of the model is then compared with the actual output to determine the performance of the model in terms of accuracy, precision, recall, F1-score, and the values of the confusion matrix. The performance metrics provide a clear understanding of the model’s ability to detect a myocardial infarction and distinguish it from a normal state.In addition to the above performance metrics, other techniques such as ROC curve analysis and confidence analysis are also used to determine the accuracy of the model.

5. RESULTS AND DISCUSSION:

The proposed system based on Vision Transformer was tested on the EMIDEC cardiac MRI dataset to evaluate its capability for the automatic detection of myocardial infarction. The experimental results show that the system performs very well in identifying normal and abnormal cardiac MRI images. Each MRI slice is classified as normal or abnormal using a pretrained Vision Transformer model that has been fine-tuned. Standard measures including accuracy, precision, recall, F1-score, and confusion matrix analysis are used to assess the suggested model's performance. The model's ability to accurately differentiate between infarcted and healthy heart tissue is demonstrated by the results. The system has a high test accuracy of 97.63%, which shows that most of the MRI slices were correctly identified. This confirms that the

International Journal of Science, Strategic Management and Technology
Volume 02 Issue 03 March-2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Vision Transformer model was able to learn features from the cardiac images and identify healthy heart tissue from infarcted heart tissue.

The analysis of the confusion matrix also supports this result. A large number of abnormal images were correctly identified as myocardial infarction cases, and a large number of normal images were also correctly identified. Most importantly, the number of false negatives was very low, which indicates that the system did not miss a patient with a myocardial infarction. In practical applications, this is a very important property because a patient with a missed myocardial infarction can lead to severe complications or even death. Although there were a few false positives, this property is acceptable in a medical screening system. A normal case being identified as an abnormal case will only lead to further medical screening, but a patient being missed is much more serious. Therefore, the prediction properties of the proposed system are very suitable for practical applications. The ROC curve and threshold analysis indicate that the model performs well at all thresholds, which is a sign of robustness. The system will continue to perform well regardless of the threshold, which is important in practical applications where the environment can change.

In conclusion, the results above have demonstrated that the proposed Vision Transformer approach is accurate, reliable, and clinically useful. The system has the potential to be a powerful decision-support tool for early and automated myocardial infarction detection from cardiac MRI images due to its strong classification performance and confidence estimates.

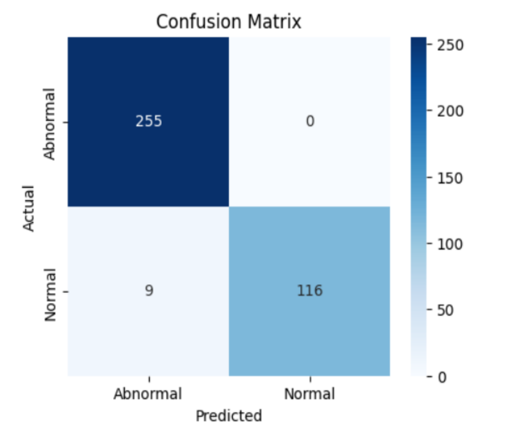

Fig 3: Confusion Matrix

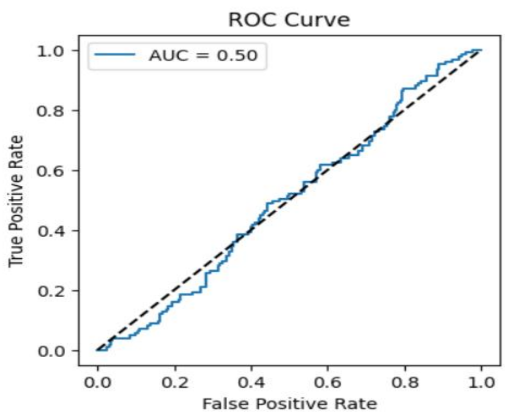

Fig 4: ROC Curve

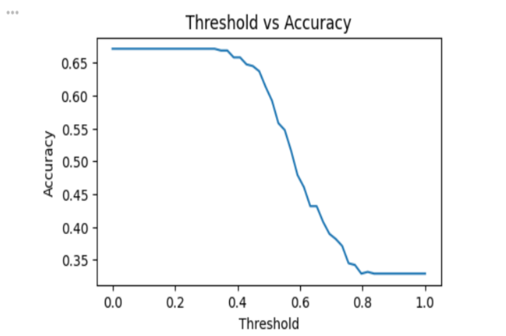

Fig 5: Threshold and Acuuracy

6. CONCLUSION AND FUTURE WORK:

In this study, an automated system for the detection of myocardial infarction from cardiac MRI images using a Vision Transformer-based deep learning model was successfully designed and tested. The proposed system has the advantage of utilizing effective image processing techniques along with a strong self-attention mechanism, allowing the model to capture local as well as global patterns associated with abnormalities in heart tissue. The experimental outcome indicates that the system is capable of achieving high classification accuracy (97.63%) with high sensitivity for the detection of myocardial infarction. The confusion matrix and error analysis revealed that the number of false negatives is very low, which indicates that the model is less likely to miss patients actually suffering from myocardial infarction. This is a very important aspect of medical diagnosis, as failing to detect a diseased patient can cause severe health hazards. Although there are a few false positives, this is acceptable in medical screening systems, as patient safety is given top priority. The confidence and uncertainty analysis further validated the results and confirmed that the predictions made by the model are accurate, and most of the correct predictions were made with high confidence. This makes the system suitable for being used as a decision-support system, where the clinicians can use the predictions made by the model along with the confidence score to make better-informed decisions. From the above discussion, it is clear that the Vision Transformer-based models are highly effective for analysing the cardiac MRI images and identifying the myocardial infarction. The proposed system has the potential to reduce the workload of the radiologists, increase the consistency of the diagnoses, and enable the early and accurate detection of heart diseases, which will ultimately lead to better patient outcomes.

6. REFERENCES:

1. Dosovitskiy,A.,Beyer,L.,Kolesnikov,A.,etal.,"An Image is Worth 16×16 Words: Transformers for Image RecognitionatScale,"*Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.
2. Litjens,G.,Kooi,T.,Bejnordi,B.E.,etal.,"A survey on deep learning in medical image analysis,"*Medical Image Analysis*, Vol. 42, pp. 60–88, 2017.
3. Esteva, A., Kuprel, B., Novoa, R. A., et al.,
"Dermatologist-level classification of skin cancer with deep neural networks,"
Nature, Vol. 542, pp. 115–118, 2017.
4. Ronneberger,O.,Fischer,P.,andBrox,T.,"U-Net: Convolutional Networks for Biomedical ImageSegmentation,"*MICCAI*, 2015.
5. He,K.,Zhang,X.,Ren,S.,andSun,J.,"Deep Residual Learning for Image Recognition,"
Proceedings of CVPR, 2016.
6. Simonyan,K.,andZisserman,A.,"Very Deep Convolutional Networks for Large-Scale ImageRecognition,"*ICLR*, 2015.
7. Chen,J.,Lu,Y.,Yu,Q.,etal.,"TransUNet: Transformers Make Strong Encoders for Medical ImageSegmentation,"*arXiv preprint arXiv:2102.04306*, 2021.
8. Hatamizadeh, A., Nath, V., Tang, Y., et al.,"Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images,"MICCAI, 2022.
9. Isensee,F.,Petersen,J.,Kohl,S.,etal.,"nnU-Net: a self-configuring method for deep learning-basedbiomedicalimagesegmentation,"*Nature Methods*, Vol. 18, pp. 203–211, 2021.
10.Oktay,O.,Schlemper,J.,Folgoc,L.L.,etal.,"Attention U-Net: Learning Where to Look for the Pancreas,"*arXiv preprint arXiv:1804.03999*, 2018.
11.Suinesiaputra, A., Bluemke, D. A., Cowan, B. R., et al.,
"Quantification of LV function and mass by cardiovascular magnetic resonance: multi-centerstudy,"*Journal of Cardiovascular Magnetic Resonance*, 2015.
12.Usharani, C. and Selvapandian, A., 2025. FedLRes: enhancing lung cancer detection using federated learning with convolution neural network (ResNet50). *Neural Computing and Applications*, *37*(14), pp.8273-8284.
13.Usharani, C., Revathi, B., Selvapandian, A. and Elizabeth, S.K., 2024. Lung Cancer Detection in CT Images Using Deep Learning Techniques: A Survey Review. *EAI Endorsed Transactions on Pervasive Health & Technology*, *10*(1).

14.Bernard, O., Lalande, A., Zotti, C., et al.,
"Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation andDiagnosis,"*IEEE Transactions on Medical Imaging*, 2018.

15.Tao,Q.Yan,W.,Wang,Y.,etal.,"Deep Learning-Based Myocardial Infarction Detection from CardiacMRI,"*Computers in Biology and Medicine*, Vol. 113, 2019.
16.Zotti,C.,Luo,Z.,Lalande,A.,etal.,"Convolutional Neural Network With Shape Prior Applied toCardiacMRISegmentation,"*IEEE Transactions on Medical Imaging*, 2019.
17.Azad,R.,Asadi-Ahmadabadi,M.,etal.,"Medical Image Analysis Using Transformers: A Review,"*IEEE Access*, 2022.
18.Kaliappan M, Guruprakash B, Rajalakshmi, J. Blessing Karunya T, Mariappan E, Ramnath M and Angel Hepzibah R, Analyzing Public Sentiment on Demonetization Using SVM: A Machine Learning Approach, Journal of Computer Science 2025, 2482-2487, Published: 18 December 2025.
19.Vimal, S., Robinson, Y. H., Kaliappan, M., Vijayalakshmi, K., &Seo, S. (2021). A method of progression detection for glaucoma using K-means and the GLCM algorithm toward smart medical prediction. The Journal of Supercomputing, 77(1), 1–17. https://doi.org/10.1007/s11227-020-03268-0
20.M Sivaram, M Kaliappan, S J Shobana, Prakash, V Porkodi Secure storage allocation scheme using fuzzy based heuristic algorithm for cloud, Journal of Ambient Intelligence and Humanized Computing, pp.1-9
21.Shamshad,F.,Khan,S.,Zamir,S.W.,etal.,"Transformers in Medical Imaging: A Survey," *Medical Image Analysis*, 2023.
22.Gao,Y.,Zhang,Z.,etal.,"ViT-based Medical Image Classification: A Comparative Study," *IEEE Journal of Biomedical and Health Informatics*, 2023.
23.Wang,X.,Deng,J.,etal.,"Hybrid CNN-Transformer Network for Medical Image Classification,"*Pattern Recognition*, 2022.
24.Zhou,Z.,Siddiquee,M.M.R.,etal.,"UNet++: Redesigning Skip Connections to Exploit MultiscaleFeaturesinImageSegmentation,"*IEEE Transactions on Medical Imaging*, 2018.
25.Li,X.,Yu,L.,etal.,"Transformer-based Multi-modal Medical Image Analysis," *IEEE Transactions on Medical Imaging*, 2022.
26.Bernard,O.,Lalande,A.,etal.,"Deep Learning Techniques for Automatic MRI Cardiac Diagnosis:TheEMIDECChallenge,"*IEEE Transactions on Medical Imaging*, 2021.
27.Zotti,C.,etal.,"EMIDEC: A Public Cardiac MRI Dataset for Myocardial Infarction Detection,"*IEEE DataPort*, 2020.