Medical Research Archives

European Society of Medicine

OPEN ACCESS

Published: September 30, 2023

Citation: Sabbah, M., et al., 2023. Atherosclerosis Detection using Autoencoders Applied to MR Images. Medical Research Archives, [online] 11(9). https://doi.org/10.18103/mra.v11i9.4350

DOI: https://doi.org/10.18103/mra.v11i9.4350

ISSN: 2375-1924

RESEARCH ARTICLE

Atherosclerosis Detection using Autoencoders Applied to MR Images

Maher Sabbah[1*], Yasmine Abou Adla[1,2], Sara El-Kadri[1], Jad Baalbaki[1], Mohamad El-Zein[1], Rached Zantout[1], and Mohamad Diab[1]

[1]Rafik Hariri University (RHU);
[2]American University of Beirut (AUB);

[*]sabbahmm@rhu.edu.lb

ABSTRACT

The early detection of atherosclerosis has been the interest of researchers in order to prevent diseases progression. And, most of cases diagnosed with atherosclerosis in late stages resulted of subjective ways of diagnosis in the healthcare sector.

Consequently, in order to help cardiologists diagnose atherosclerosis in early stages with an objective, fast and accurate way, we are proposing a machine learning model based on autoencoders that enables atherosclerosis detection from MRI images of murine subjects.

This novel way of automated system consists of applying various image processing techniques on the input MRI images. Then, after training, testing and validation it will be capable of classifying the images as atherosclerotic or not based on a specific threshold for the reconstruction error calculated from the autoencoders' output.

An autoencoder is a feed-forward neural network that has its input neurons equal to the output neuron. The classification works by comparing the reconstructed images to the original input images and evaluating loss between them. Since the autoencoder is trained on healthy images, reconstruction error of the healthy images would be low while that of atherosclerotic subjects would be higher. By setting a threshold for the loss, we can classify the images as healthy or atherosclerotic.

The pre-processing of these images was made using a Block Matching 3D (BM3D) filter to remove the noise in the images prior to application of a Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance the contrast. The next step included introducing the dataset into the autoencoder to start training on the healthy images, after increasing the images number with augmentation.

The results showed a reconstruction loss of 0.018 while using the stacked architecture of the autoencoder and 0.0366 when using the convolutional autoencoder architecture.

Keywords: Atherosclerosis, Autoencoder, MRI, Deep Learning, anomaly detection, pattern recognition.

I. Introduction

Cardiovascular diseases pose significant challenges in healthcare. These conditions can have a profound impact on individuals and their overall well-being. The primary focus in managing these diseases is to detect them at an early stage, allowing for timely intervention and prevention of associated complications. Atherosclerosis is a common and significant problem in cardiovascular disease, characterized by the buildup of plaque in the arterial walls[1]. The condition involves the gradual thickening and hardening of arteries, which can lead to restricted blood flow to organs and tissues throughout the body[1,2]. The plaque consists of fats, cholesterol, and other substances that accumulate in and on the artery walls[1]. Over time, the plaque can cause narrowing of the arteries, potentially leading to blockages and reduced blood flow[1,2]. One of the main challenges posed by atherosclerosis is the lack of noticeable symptoms in its early stages[2,3]. Mild atherosclerosis usually does not present any symptoms, making it difficult to detect without medical intervention.[1] Symptoms typically manifest when an artery becomes severely narrowed or completely blocked, preventing adequate blood supply to vital organs and tissues[1]. The specific symptoms experienced depend on the location of the affected arteries[1,2]. For example, atherosclerosis in the heart arteries may result in chest pain or pressure (angina), while blockages in the arteries leading to the brain can lead to numbness, weakness, speech difficulties, or temporary vision loss[1,2]. Atherosclerosis in the arteries of the arms and legs may cause symptoms such as leg pain when walking or decreased blood pressure in the affected limb[1]. Additionally, atherosclerosis in the arteries leading to the kidneys can contribute to high blood pressure or kidney failure.[1] The underlying causes of atherosclerosis are multifactorial, with various risk factors contributing to its development[1,2,3]. These risk factors include high blood pressure, high cholesterol, high triglycerides, smoking, diabetes, obesity, inflammation, and certain autoimmune diseases[1,2]. Damage to the inner layer of the artery, often caused by these risk factors, triggers the accumulation of substances and the formation of plaque[1,3]. Atherosclerosis is a progressive disease that can begin in childhood and worsen over time if left untreated[1]. Early diagnosis is crucial to prevent the progression of atherosclerosis and mitigate the risk of complications such as heart attacks, strokes, and other cardiovascular emergencies[1,2].

Monitoring atherosclerosis development is essential for many reasons, including early identification, cardiovascular event prediction, therapy effectiveness assessment, and personalized risk assessment. First, early detection of atherosclerosis facilitates the identification of people with subclinical disease, enabling rapid intervention and reduction of risk factors[4]. This is crucial for stopping disease development and lowering the risk of cardiovascular events. Second, monitoring atherosclerosis helps forecast the likelihood of future cardiovascular events such as heart attacks and strokes, by assessing the extent and severity of plaque buildup in the arteries[5]. Third, monitoring the development of atherosclerosis provides valuable information on the effectiveness of the treatment approaches and therapies. The effect of lifestyle modifications, medications, or other interventions can be assessed by

tracking changes in plaque characteristics[6]. This is particularly useful for people with existing atherosclerosis or those who are at high risk for cardiovascular illnesses. Finally, monitoring atherosclerosis enables customized management strategies based on unique risk profiles and treatment outcomes, which facilitates personalized medicine and risk assessment. Integrating molecular imaging techniques with clinical risk scores can improve risk assessment and guide personalized treatment decisions[7].

Moreover, invasive ultrasonic methods, like intravascular ultrasound (IVUS), provide detailed insights into atherosclerotic lesions by inserting an ultrasound catheter into arteries. IVUS yields high-resolution cross-sectional images, assessing plaque burden, composition, and distribution for interventional procedures like stent placement[8]. Additionally, virtual histology IVUS (VH-IVUS) classifies plaque components for high-risk plaque identification[9].

Non-invasive ultrasound techniques are pivotal in atherosclerosis monitoring. B-mode ultrasound visualizes plaque morphology and intima-media thickness (IMT)[10]. Doppler ultrasound assesses blood flow characteristics, while contrast-enhanced ultrasound (CEUS) enhances plaque vascularity visualization[11,12]. Shear wave elastography (SWE) measures plaque stiffness[13]. These non-invasive techniques are safe, cost-effective, and widely available, allowing for repetitive assessments over time and facilitating risk assessment, treatment monitoring, and disease progression tracking in atherosclerosis patients.

Coronary angiography is an invasive technique that visualizes coronary arteries and assesses atherosclerotic lesions. Using a contrast dye injected into coronary arteries and X-ray images, it identifies blockages or narrowed areas in vessel walls[14]. This provides detailed information about atherosclerotic plaques, aiding in coronary artery disease (CAD) diagnosis and management.

Quantitative Coronary Angiography (QCA) is a computer-assisted method that measures coronary artery stenosis. It involves analyzing angiographic images to determine arterial lumen diameter at specific reference points[15]. QCA accurately quantifies plaque burden, facilitating disease progression assessment and treatment outcome evaluation.

The combination of coronary angiography with additional techniques enhances diagnostic capabilities. Intravascular ultrasound (IVUS) offers cross-sectional vessel wall images, assessing plaque burden, composition, and distribution[16]. Integrating IVUS with coronary angiography provides detailed morphology and vulnerability insights into atherosclerotic plaques. Optical coherence tomography (OCT) achieves high-resolution images using near-infrared light. This allows precise measurements of plaque thickness, fibrous cap thickness, and thrombus presence[17]. Combining OCT with coronary angiography yields valuable plaque characteristic insights, aiding risk stratification and therapeutic guidance.

Computed Tomography (CT) scan is a non-invasive imaging modality extensively used for atherosclerosis assessment. A specific application of CT, CT angiography (CTA), delivers detailed images of coronary arteries, enabling visualization and evaluation of atherosclerotic plaques[18]. During CTA, a contrast dye is injected into the bloodstream,

followed by X-ray images to construct a three-dimensional reconstruction of coronary arteries. CT scan offers rapid image acquisition, high spatial resolution, and the ability to assess both calcified and non-calcified plaque components.

Coronary CT angiography (CCTA) has emerged as a valuable non-invasive tool for detecting and characterizing coronary artery disease (CAD). CCTA identifies and quantifies atherosclerotic plaques, evaluates stenosis severity, and assesses plaque composition. By categorizing plaques based on characteristics like calcification, non-calcified plaque burden, and plaque remodeling, CCTA supports risk stratification and treatment guidance[19].

Advancements in CT technology, such as dual-source CT and third-generation scanners, have further improved image quality while reducing radiation exposure. Additionally, novel techniques like coronary artery calcium scoring (CACS) offer quantitative assessment of coronary artery calcification, a key marker of atherosclerotic burden and cardiovascular risk[20]. However, it's important to acknowledge limitations. Motion artifacts, blooming effects from calcifications, and challenges in detecting non-calcified plaques can affect image quality and interpretation. Moreover, CT scans involve ionizing radiation and contrast agents, raising potential risks and considerations for certain patient populations[21].

Magnetic Resonance Imaging (MRI) has emerged as a powerful non-invasive imaging technique for the detection, quantification, and monitoring of atherosclerosis. MRI offers several advantages over other imaging modalities, making it the most important tool in atherosclerosis assessment. Its unique capabilities provide comprehensive information on plaque morphology, composition, and functional parameters[22]. By employing various MRI sequences, such as T1-weighted, T2-weighted, and proton-density-weighted imaging, different plaque components, including lipid-rich necrotic cores, fibrous tissue, calcifications, and intraplaque hemorrhage, can be visualized[23]. This enables accurate characterization of plaques and identification of vulnerable lesions, which is crucial for risk stratification and guiding treatment decisions.

Moreover, MRI provides excellent soft tissue contrast and high spatial resolution, allowing for detailed imaging of the vessel wall and accurate measurement of plaque burden[24]. Quantitative parameters, such as plaque volume, plaque surface area, and the degree of stenosis, can be precisely evaluated using MRI techniques[25]. Additionally, advanced MRI techniques, such as diffusion-weighted imaging and dynamic contrast-enhanced imaging, offer insights into plaque microstructure, inflammation, and angiogenesis, enhancing the understanding of atherosclerosis pathophysiology. Furthermore, MRI enables functional assessment through techniques like arterial spin labeling and phase-contrast imaging, allowing evaluation of blood flow, wall shear stress, and endothelial function in affected vessels[26]. These functional parameters are crucial for assessing the hemodynamic consequences of atherosclerosis and its impact on cardiovascular health.

II. Background information

A. DIAGNOSIS OF ATHEROSCLEROSIS:

Numerous imaging techniques have been explored to detect plaque volume and

constituents for better clinical indicators of plaque vulnerability. While intravascular ultrasound can distinguish plaque constituents, it is invasive and not suitable for serial studies. B-mode ultrasound has been utilized to assess stenotic severity, plaque volume, and plaque constituents in the carotid artery, but its ability to measure plaque volume and visualize intraplaque features is limited by the plane of acquisition and incident angle of the ultrasound. Three-dimensional ultrasound can help measure plaque volume and detect ulcerations on the lumen surface but cannot characterize internal plaque composition. X-ray computed tomography has been used to identify and quantify coronary calcium, but it has not been tested for identifying non-calcified lesion components[27].

In contrast, MRI plays a critical role in detecting atherosclerosis, enabling the detection of the disease at an early stage, before symptoms appear, and allowing for early intervention to improve patient outcomes. Moreover, MRI provides a comprehensive evaluation of the artery wall and can assess plaque burden and identify plaque constituents. The primary objective of atherosclerosis MRI is to quantify plaque burden based on plaque volume. Achieving an adequate spatial resolution is crucial to visualizing lesion components and synchronizing MR imaging with electrocardiogram (ECG) and respiratory signals is preferred to reduce signal interference and obtain a clear image[27].

B. DIAGNOSING ATHEROSCLEROSIS USING ML TECHNIQUES:

In recent years, there has been a growing interest in the use of machine learning algorithms for the accurate and efficient diagnosis of cardiovascular diseases, including atherosclerosis. Machine learning techniques have the potential to increase the accuracy of disease diagnosis and avoid false positives and negatives. Recent studies have explored the use of machine learning techniques in the diagnosis of cardiovascular diseases, with a particular focus on atherosclerosis. In one such study, researchers extracted features from ECG and respiratory signals of 10 murine subjects to classify atherosclerosis using a random forest classifier. The features included average heart rate, R-R interval, R amplitude, isovolumetric contraction duration, inspiration and expiration peak, inspiratory and expiratory phases, and respiratory rate. The model achieved an accuracy of 98.18% on testing data[28].

Another recent study proposed the use of artificial neural networks (ANNs) for the classification of atherosclerosis in a dataset consisting of 270 subjects. The dataset included features such as age, gender, blood pressure, serum cholesterol, resting ECG, maximum heart rate, and others. After training the ANN on these features, the model achieved an accuracy of 96% in classifying atherosclerosis[29]. These studies suggest that machine learning techniques can significantly increase the accuracy of atherosclerosis diagnosis and classification, thereby improving patient outcomes. The use of non-invasive and automated methods, such as those enabled by machine learning, can provide valuable insights into disease diagnosis and management. Further research in this area can help refine these methods and improve their clinical utility.

III. Autoencoders: an introduction to unsupervised deep learning

Autoencoders (AEs) are a type of neural network that belong to the class of unsupervised deep learning models. Unlike supervised learning methods, which require labeled data to make predictions, unsupervised methods can automatically learn the underlying structure of data without any prior knowledge or labels. Autoencoders consist of two main components: an encoder network and a decoder network. As seen in Figure 1, the encoder network takes an input and maps it to a lower-dimensional latent representation, while the decoder network takes this latent representation and maps it back to the original input space. The encoder and decoder networks are trained jointly to minimize the difference between the input and output data, which is known as the reconstruction loss[30]. During training, the input data is fed through the encoder network, and the resulting lower-dimensional latent representation is fed through the decoder network. The output of the decoder network is then compared to the original input data to compute the reconstruction loss. The goal is to minimize the reconstruction loss, which encourages the autoencoder to learn a compressed representation of the input data that captures its essential features. Once the autoencoder is trained, it can be used for various downstream tasks by utilizing the encoder network to generate a lower-dimensional latent representation of new input data. One of the most common uses of autoencoders is for data compression. By compressing the input data into a lower dimensional latent space, using the encoder network, the amount of storage required to store the data is significantly reduced. The decoder network can then be used to reconstruct the original data from the compressed representation. Another application of AEs is feature extraction. By training the autoencoder on a large dataset, the encoder network can learn to extract useful features from the input data. These features can then be used as input to a supervised learning algorithm, such as a classifier. This can be particularly useful in situations where manual feature engineering is difficult or time-consuming. Autoencoders can also be used for anomaly detection, where the goal is to identify data points that deviate significantly from the norm. By training the autoencoder on a dataset of normal data, the model can learn to reconstruct normal data with low reconstruction error, while introducing high reconstruction error for anomalous data. Once the model is trained, it can be used to identify anomalous data points by computing their reconstruction error and comparing it to a threshold value. For example, in the context of medical imaging, the latent representation can be used for anomaly detection, disease classification, or image segmentation. One of the key advantages of autoencoders is their ability to learn meaningful representations of the input data in an unsupervised manner, without requiring labeled data. This makes them particularly useful in scenarios where labeled data is scarce or expensive to obtain. Additionally, autoencoders can be adapted to different types of data, including images, text, and time series data, making them a versatile tool for a wide range of applications.

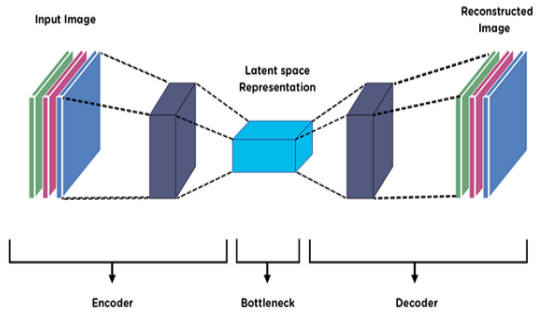

Figure 1: Architecture of Autoencoder

A. AUTOENCODERS FOR ANOMALY DETECTION IN MEDICAL IMAGING

Autoencoders have shown to be effective for anomaly detection, where the goal is to identify data points that deviate significantly from the norm. In the context of medical imaging, anomaly detection can be used to detect abnormalities or diseases in images that may not be apparent to the human eye. Autoencoders can be used for anomaly detection by training the model on a dataset of normal data, and then using it to identify data points that deviate significantly from this norm. During training, the autoencoder learns to reconstruct normal data with low reconstruction error, while introducing high reconstruction error for anomalous data. Once the model is trained, it can be used to identify anomalous data points by computing their reconstruction error and comparing it to a threshold value[31]. One of the advantages of using autoencoders for anomaly detection is that they can learn to capture the underlying distribution of the normal data, without explicitly modeling anomalies. This allows the model to generalize to new types of anomalies that were not present in the training data. Additionally, autoencoders can handle high-dimensional data, such as medical images, and can learn to extract meaningful features that are relevant for anomaly detection. To use AEs for anomaly detection in medical images, a dataset of normal images is first collected and used for training. During training, the autoencoder learns to reconstruct the normal images with low reconstruction error, while introducing high reconstruction error for anomalous images. Once the model is trained, it can be used to detect anomalous images by computing their reconstruction error and comparing it to a threshold value. Images with reconstruction error above the threshold value are considered anomalous and can be flagged for further inspection. In the context of atherosclerosis detection using MRI images, AEs can be used to identify anomalous arterial wall plaque that may not be visible to the human eye. By training the autoencoder on a dataset of healthy arteries, the model can learn to reconstruct healthy images with low reconstruction error, while introducing high reconstruction error for images with anomalous plaque. This allows the model to identify arteries that are likely to be affected by atherosclerosis and can help in early diagnosis and treatment of cardiovascular disease. Autoencoders can handle high-dimensional medical images, such as MRI images, and can learn to extract meaningful features that are relevant for anomaly detection. This is achieved by the encoder part of the autoencoder, which compresses the input image into a low-dimensional representation, and the decoder part, which reconstructs the image from the compressed representation. The compressed representation can be used as a feature vector to train a separate anomaly detection model, such as a support vector machine or a random

forest, for improved performance. Overall, AEs have shown to be a powerful tool for anomaly detection in medical imaging, particularly for early detection of diseases such as atherosclerosis. By leveraging the power of deep learning, AEs can identify subtle abnormalities in medical images that may go unnoticed by human experts, leading to better diagnosis and treatment of diseases.

B REVIEW OF AUTOENCODER ARCHITECTURES FOR UNSUPERVISED DEEP LEARNING

Autoencoder architecture plays a crucial role in the performance of autoencoder-based models for unsupervised deep learning. Over the years, several autoencoder architectures have been proposed in the literature, each with its strengths and weaknesses.

1) Standard Autoencoder: A standard autoencoder is a type of neural network that learns to reconstruct its input data. It is an unsupervised learning algorithm that can be used for feature extraction, data compression, and anomaly detection. The architecture of a standard autoencoder consists of an encoder network that compresses the input data into a low-dimensional latent space, and a decoder network that reconstructs the input data from the compressed representation[30]. The encoder network maps the input data to a lower dimensional latent space representation. The number of hidden layers and the size of the latent space can be adjusted to suit the needs of the application. The output of the encoder network is a compressed representation of the input data, often referred to as the bottleneck. The decoder network takes the bottleneck representation and maps it back to the original input space. The decoder network is designed to be symmetrical to the encoder network, so that it mirrors the operations of the encoder. The output of the decoder network is the reconstructed input data, which should be as close as possible to the original input data. During training, the autoencoder is optimized to minimize the difference between the input data and its reconstructed output. This is typically done by minimizing the mean squared error (MSE) between the input data and the reconstructed output. The weights and biases of the network are adjusted during training to minimize the MSE loss.

2) Convolutional Autoencoder: Convolutional autoencoders (CAE) are a specific type of autoencoder that is commonly used for image processing tasks. Unlike a standard autoencoder, which consists of fully connected layers, a convolutional autoencoder consists of convolutional layers in the encoder and decoder networks[32]. As seen in Figure 2, the encoder network of a CAE typically consists of multiple convolutional layers followed by a pooling layer. The convolutional layers apply a set of filters to the input image, which capture local patterns and features. The pooling layer then downsamples the feature maps to reduce their size, while preserving the most important information. The decoder network of a CAE consists of upconvolutional layers, also known as deconvolutional layers, and convolutional layers. The upconvolutional layers perform the opposite operation of the pooling layers, by upsampling the feature maps to increase their size. The convolutional layers then apply a set of filters to the upsampled feature maps, to reconstruct the original image. One advantage of using CAE for image processing tasks is that they can

learn to extract spatially meaningful features from the input image, which can be used for various downstream tasks such as image classification, segmentation, and object detection. Additionally, convolutional autoencoders can handle high-dimensional data, such as medical images, and can learn to extract features that are relevant for specific tasks. To train a CAE, the same principles as a standard autoencoder are applied. The model is trained on a dataset of input images, with the goal of minimizing the reconstruction error between the original and reconstructed images. Once the model is trained, it can be used for various tasks such as image compression, feature extraction, and anomaly detection.

3) Variational Autoencoders: Variational autoencoders (VAEs) are a type of generative model that are capable of learning a low-dimensional representation of data, while simultaneously generating new data samples that resemble the original data. VAEs are a type of probabilistic autoencoder, where the encoder maps the input data to a probability distribution in the latent space, and the decoder maps samples from the latent space to a probability distribution over the input data.[33] Unlike standard autoencoders, VAEs are trained using a variational inference approach, where the objective is to minimize the difference between the true posterior distribution over the latent variables and a tractable approximate distribution. The approximate posterior distribution is typically modeled as a Gaussian distribution with mean and variance parameters, which are output by the encoder network. The latent variable samples are then generated by sampling from this Gaussian distribution. The objective function for a VAE includes two terms: the reconstruction loss, which measures the ability of the model to reconstruct the input data, and the Kullback-Leibler (KL) divergence between the approximate posterior distribution and the prior distribution over the latent variables. The KL divergence term encourages the approximate posterior distribution to match the prior distribution, which is typically modeled as a unit Gaussian distribution. By minimizing the KL divergence term, the VAE encourages the encoder network to learn a compact and disentangled representation of the input data in the latent space (Figure 3). One of the advantages of VAEs is that they can generate new data samples by sampling from the learned latent space. By sampling from the prior distribution over the latent variables, the decoder network can generate new samples that resemble the original data. This makes VAEs useful for tasks such as data generation, data augmentation, and transfer learning. In the context of medical imaging, VAEs have been used for a variety of tasks, such as image denoising, image reconstruction, and image synthesis. For example, VAEs can be used to generate synthetic medical images that can be used to augment training data for supervised learning algorithms. Additionally, VAEs can be used to generate new medical images that can be used for diagnostic purposes, such as identifying abnormalities or diseases. In summary, autoencoders are powerful unsupervised deep learning models that can be used for a variety of tasks, including data compression, feature extraction, anomaly detection, and image generation. They work by learning a compressed representation of the input data, which can be used for various downstream tasks. Over the years, various

architectures of autoencoders have been proposed, each with its own strengths and weaknesses. Some of the most popular architectures include the standard autoencoder, convolutional autoencoder, and variational autoencoder. It is important to carefully select the architecture of the autoencoder based on the task at hand and the nature of the data. In addition, hyperparameters such as the number of layers, neurons, and learning rate should also be optimized to ensure optimal performance. Overall, autoencoders have shown great promise in the field of unsupervised deep learning and are expected to play an increasingly important role in future research and applications.

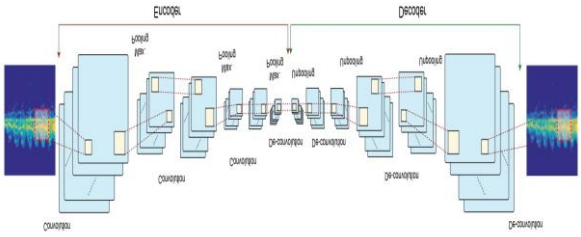

Figure 2: CAE Architecture[34].

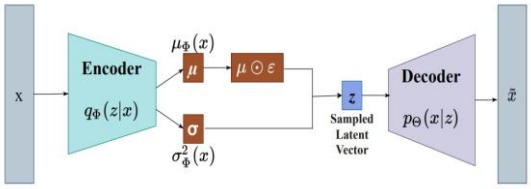

Figure 3: CAE Architecture[35].

IV. Transfer learning

A. OVERVIEW

Transfer learning is a technique in machine learning that enables a model to leverage knowledge gained from one task to improve its performance on another related task. In the context of deep learning, transfer learning involves reusing the learned weights of a pre-trained neural network as a starting point for training a new network for a different task[36]. One of the main advantages of transfer learning is that it enables us to train models with much smaller datasets. By using a pre-trained model, we can leverage the knowledge gained from a large dataset to improve the performance of a model trained on a smaller dataset. Additionally, transfer learning can reduce the time and computational resources required for training a new model from scratch. There are several different types of transfer learning techniques, including fine-tuning, feature extraction, and domain adaptation. Fine-tuning involves taking a pre-trained model and re-training it on a new dataset with a different set of output classes. Feature extraction, on the other hand, involves freezing the weights of the pre-trained model and using it as a fixed feature extractor for a new dataset. Finally, domain adaptation involves adapting a pre-trained model to a new domain with different characteristics, such as a different data distribution. Transfer learning has shown great success in many applications, including image classification, object detection, natural language processing, and speech recognition. In the context of medical applications, transfer learning has been used to improve the performance of deep learning models for a range of tasks, such as disease diagnosis, lesion segmentation, and medical image classification. One of the challenges of transfer learning in medical applications is the need to fine-tune the pre-trained models to the specific medical domain of interest, which may have different characteristics from the original domain the model was trained on. However, with appropriate domain adaptation techniques, it is possible to overcome these

challenges and achieve state of- the-art performance in medical applications. Overall, transfer learning is a powerful technique for improving the performance of deep learning models in various domains, including medical applications. It enables us to leverage the knowledge gained from large datasets to improve the performance of models trained on smaller datasets and reduces the time and computational resources required for training a new model from scratch. In this paper, we decided to use transfer learning on the visual geometry group (VGG), VGG-16 model, originally trained on the ImageNet dataset to classify images into 1000 categories. Using transfer learning on the VGG-16 model is important because it allows us to leverage the model's pre-trained weights and architecture, which have been shown to be effective for a wide range of image classification tasks. By fine-tuning the last few layers of the model on our own dataset, we can re-purpose the model to solve our own classification task with relatively few training examples. Compared to training a model from scratch, transfer learning can be significantly faster and require less training data, as the pre-trained model has already learned to extract useful features from images. This can be particularly useful in medical imaging, where annotated data can be scarce and time-consuming to obtain. Additionally, transfer learning can improve the performance of our model by providing a better initialization of the weights and biases of the neural network, which can lead to better convergence during training. However, it is important to note that transfer learning is not always guaranteed to improve the performance of a model. The choice of pre-trained model and the specific layers to fine-tune can have a significant impact on the performance of the final model. In some cases, it may be necessary to train a model from scratch to achieve optimal performance for a specific task.

B. VGG-16 *VGG16* is a convolutional neural network (CNN) architecture that was introduced by the Visual Geometry Group (VGG) at the University of Oxford in 2014. It is a deep learning model that consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. The VGG16 architecture achieved state-of-the-art results on the ImageNet dataset, which is a large-scale image recognition dataset containing millions of images.[37] As shown in Figure 4, the VGG16 architecture is characterized by its simplicity and uniformity. All convolutional layers have a fixed 3x3 filter size and a stride of 1, and all pooling layers have a 2x2 filter size and a stride of 2. This uniformity makes the architecture easy to implement and optimize. The first layer of the VGG16 model takes an input image of size 224x224x3 and applies a set of convolutional filters to generate a set of feature maps. The subsequent convolutional layers progressively reduce the spatial dimensions of the feature maps while increasing the number of channels. The pooling layers halve the spatial dimensions of the feature maps, further reducing the number of parameters in the model. The fully connected layers at the end of the VGG16 architecture perform the classification task. The output of the final fully connected layer is a probability distribution over the different classes in the ImageNet dataset.

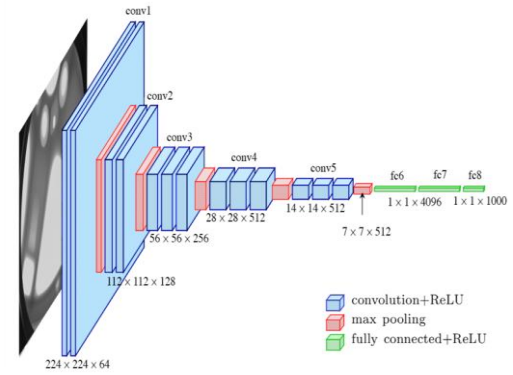

Figure 4: VGG-16 Architecture[38].

Transfer learning on VGG16 is a popular technique in deep learning, especially for computer vision tasks. Transfer learning involves using a pre-trained model, such as VGG16, as a starting point for training on a new task or dataset. By using a pre-trained model, we can leverage the knowledge that the model has learned from a large dataset and transfer that knowledge to a smaller dataset. This can significantly reduce the amount of data required for training and improve the performance of the model. In our research, we decided to do transfer learning on VGG16 to classify medical images. The Image Net dataset, on which VGG16 was pre-trained, contains millions of natural images, which have different features than medical images. However, the lower layers of VGG16 learn generic features, such as edges and textures, which are also present in medical images. By freezing the lower layers of VGG16 and training only the top layers on our medical images, we can take advantage of the learned generic features while adapting the model to our specific domain. Compared to models trained from scratch, transfer learning on VGG16 can often achieve higher accuracy with less training data and training time. Additionally, VGG16 has been extensively tested and optimized, making it a reliable and robust architecture for many computer vision tasks.

V. Making ai transparent: explainable ai in medical diagnosis

Explainable AI (XAI) is a growing field of research that aims to develop machine learning models that are transparent and interpretable to humans. In the context of medical applications, XAI has the potential to improve patient outcomes and increase trust in AI-based decision-making. XAI can help clinicians understand how AI algorithms arrive at their decisions, identify potential biases, and improve the overall accuracy and reliability of AI models[39]. One approach to XAI in medical applications is to use visual explanations. Visual explanations provide an intuitive and interactive way to explain the decision-making process of AI models to clinicians and patients. One example of this is the use of saliency maps to highlight the most important features in medical images that contribute to a particular diagnosis. Another example is the use of attention maps to visualize the regions of interest in medical images that are most relevant for a particular task. Another approach to XAI is to use model-agnostic methods, which can be applied to any type of machine learning model. These methods aim to identify the most important features or inputs that contribute to the output of the model. This can be achieved through techniques such as feature importance ranking, partial dependence plots, and individual conditional expectation plots. These methods can help identify potential biases in the model and provide insights into how the model is making decisions. XAI in medical applications is still a

relatively new field, and there are many challenges to overcome. One of the challenges is the need for large and diverse datasets to train and validate AI models. Another challenge is the need for standardized evaluation metrics to assess the performance and interpretability of AI models. Despite these challenges, XAI has the potential to revolutionize the field of medical AI and improve patient outcomes by providing clinicians with more transparent and interpretable decision-making tools. A. Techniques for XAI in Medical Applications Explainable AI is an essential requirement for medical applications where decisions made by AI models must be transparent and explainable. There are various techniques used for XAI in medical applications, some of which are discussed below[40].

1). LIME: Local Interpretable Model-Agnostic Explanations (LIME) is a technique that can be used to explain the output of any machine learning model. LIME creates a local surrogate model around the prediction made by the original model and provides an explanation in terms of the features that the surrogate model finds most relevant to the prediction. In medical applications, LIME can be used to explain the output of deep learning models applied to medical images.

2) Grad-CAM: Gradient-weighted Class Activation Mapping (Grad-CAM) is another technique that can be used to explain the output of deep learning models. Grad- CAM generates a heatmap that highlights the regions of an image that contribute the most to the final prediction. In medical applications, Grad-CAM can be used to visualize the regions of medical images that the deep learning model has identified as being relevant to the diagnosis.

3) Decision Trees: Decision trees are a popular technique for building interpretable models. In medical applications, decision trees can be used to predict patient outcomes or diagnose diseases. The structure of decision trees can be visualized and provides insight into the reasoning behind the model's decisions.

4) Rule-based Systems: Rule-based systems are another technique that can be used for building interpretable models. Rule-based systems consist of a set of rules that are used to make predictions. In medical applications, rule-based systems can be used to diagnose diseases or predict patient outcomes. The rules used in the system can be easily understood by medical professionals and provide transparency into the reasoning behind the model's decisions.

5) Model Distillation: Model distillation is a technique that can be used to make complex models more interpretable. The technique involves training a smaller, simpler model to mimic the behavior of a larger, more complex model. The smaller model can then be used to make predictions and is more interpretable than the original complex model. The use of XAI techniques in medical applications can provide transparency into the decisions made by AI models, enabling medical professionals to make informed decisions.

VI. Methodology

In this section, we provide a detailed description of the methodology used in our study to develop and evaluate a deep learning model for the detection of atherosclerosis in murine subjects from gated

MRI images. We begin by describing the pre-processing steps, which involved the use of a Block Matching 3D (BM3D) filter and Contrast Limited Adaptive Histogram Equalization to enhance the contrast in the images. We also discussed how the dataset was augmented to increase the number of images and how the stacked autoencoder and convolutional autoencoder architectures were employed and compared to each other.

A. DATA ACQUISITION AND PRE-PROCESSING

The MRI images used in this study were collected using a high-performance 2-Tesla horizontal MR system with a 180 mT/m gradient set. Axial and coronal scout images of the heart, aortic root, and the carotid origin were obtained using a 2D gradient echo sequence. In addition, three high-resolution MRI sequences, GE (Gradient Echo), SE (Spin Echo), and FSE (Fast Spin Echo), were employed with specific acquisition parameters, TR/TE = 385/10 ms, 450/18 ms, and 2300/50 ms, respectively. To capture the axial T1-weighted images of the carotid origin, a 2D multislice SE sequence was used with specific parameters, including a TR of 290 ms, TE of 18 ms, matrix size of 256 × 256, slice thickness of 1 mm, and pixel size of 90 μm. The dataset available contains 215 murine subject images, including 188 atherosclerotic and 27 healthy images[41]. After data cleaning and selecting only the axial imaging plane we were left with 12 diseased images and 6 healthy.

To prepare the MRI images for training the autoencoder, a pre-processing step was implemented to enhance the quality of the images by removing noise and improving contrast. The images were first denoised using the Block Matching 3D filter, which groups similar blocks in the frequency domain and eliminates high-frequency noise by thresholding the transformed data[42]. The denoised blocks were then averaged to produce the final denoised image. To further improve the contrast levels of the denoised images, Contrast-Limited Adaptive Histogram Equalization (CLAHE) was applied. CLAHE divides the image into small tiles and calculates a local histogram before equalizing it, limiting the contrast based on a threshold.[43] The application of BM3D filter and CLAHE is shown in Figures 5 a and b respectively.

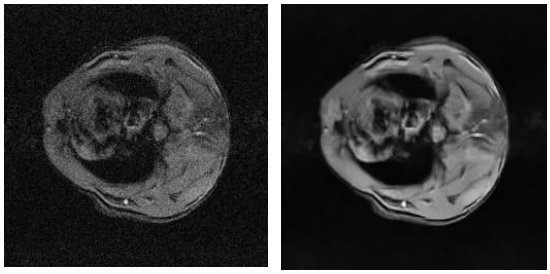

Figure 5.a: Before Processing
Figure 5.b: After Processing

B. DATA AUGMENTATION

Acquiring sufficient labeled data for real-world applications, particularly in the medical field, can be a daunting challenge for deep learning models. Such models require an extensive amount of training data to avoid overfitting, which is why data augmentation has become a critical aspect of training deep learning models with image data. Data augmentation can help increase the quantity and diversity of available training data, which is essential for achieving high accuracy levels in medical applications and other domains.

To augment our dataset, various image augmentation techniques were employed, using the ImageDataGenerator class in Keras to rotate, shift, rescale, zoom, and flip the

images. Specifically, we chose the axial point of view (POV) from the healthy and diseased images to train our model and applied all the aforementioned techniques to them. Initially, the dataset for this axial plane comprised only 6 healthy images and 12 diseased images. However, after applying image augmentation techniques, the number of healthy images increased to 1505 and the number of diseased images increased to 1212, enabling the training of a more robust and accurate deep learning model.

C. PROPOSED MODEL

In this subsection, we outline the experimental protocol employed for the training process of both autoencoders,

1. Deep Learning Algorithm:

The dataset was split into three parts using an 80-10-10 ratio: the training set, representing 80% of the total images, and the testing and validation sets, each containing 10%. This approach helps to ensure that the model can generalize well to new data and not just memorize the training set. The training set consisted of only healthy images, while the testing and validation sets were balanced, with an equal number of healthy and diseased images.

The autoencoder took grayscale images with dimensions of 256 x 256 pixels as input, meaning that the input layer of the autoencoder had a shape of (256, 256, 1). The models were trained using a batch size of 128 for 50 epochs, with the Adam optimizer. The activation function used between the layers was ReLU, and sigmoid was used for the output layer. The models were evaluated using the Mean Squared Error (MSE) and Mean Absolute Error (MAE) as the loss criteria.

2. Stacked Autoencoder Model

The model was designed to be trained on healthy images and then tested on both healthy and diseased images. Images with a reconstruction error above a certain threshold were considered diseased. The number of layers and their sizes were selected using the Keras Hyper tuner to optimize the model's performance. The stacked autoencoder model used multiple fully connected layers, with the first layer being the input layer of shape (256,256,1). The model had several dense layers, with each dense layer having an initially large number of neurons that gradually decreased in size until the bottleneck was reached. The bottleneck layer had a neuron size of 32, and the output layer of the decoder had a size of (256x256), which was reshaped into (256,256,1) for image reconstruction. The use of a stacked autoencoder allowed for the model to learn more complex features and patterns in the data, leading to improved accuracy in identifying diseased images.

3. Convolutional Autoencoder Model

The convolutional autoencoder model consisted of an input layer, followed by a layer that added Gaussian noise to the images to make the reconstruction more robust. The model also had a convolutional layer using 16 filters and a kernel size of (3x3), followed by a max pooling layer. More convolutional and max pooling layers were added alternately as the encoder layers progressed, with the number of filters used in each convolutional layer decreasing until the bottleneck was reached. The decoder layer consisted of deconvolutional layers paired with upsampling layers that alternated in size until the size of the original input was reached.

4. Transfer Learning

The autoencoder is trained using a dataset comprising only normal MRI images. The objective is for the autoencoder to learn the underlying patterns and regularities present in the normal images, enabling it to identify anomalies during inference. The Adam optimizer with a learning rate of 0.001 is employed to optimize the model parameters. The mean squared error (MSE) loss function is used to measure the similarity between the input and reconstructed images. During training, the model is exposed to the training dataset in batches of 64 images. The training data is shuffled in each epoch to prevent the model from memorizing the order of the images and to enhance generalization. The training process is performed for a fixed number of epochs, typically 50.

To evaluate the performance of the trained autoencoder in detecting anomalies, a separate validation dataset is used. This dataset contains both normal and anomalous MRI images, with ground truth annotations indicating the presence of anomalies. The trained autoencoder reconstructs the input images from the validation dataset, and the similarity between the reconstructed and ground truth images is measured using the MSE loss. A threshold on the loss value is set to classify images as either normal or anomalous. Performance metrics such as precision, recall, and F1 score are computed to assess the accuracy of the proposed approach in detecting anomalies. Visual analysis of the reconstructed images is also performed to gain insights into the autoencoder's ability to capture anomalous regions.

All experiments are conducted on a suitable hardware platform with sufficient computational resources. The model implementation is based on widely-used deep learning frameworks such as Keras and TensorFlow. Hyperparameter tuning experiments are performed to optimize the model performance. Parameters such as learning rate, batch size, and layer configurations are systematically varied, and their impact on the detection accuracy is evaluated.

5. Grad Cam

In this study, we applied the Grad-CAM technique to visualize the activations in three different layers of the trained autoencoder. After training the autoencoder on a dataset comprising healthy MRI images, we implemented Grad-CAM to gain insights into the importance and localization of features at each layer of the encoder. Grad-CAM provides a heat map highlighting the regions in the input image that contribute significantly to the activations at a specific layer. We followed a step-by-step methodology, starting with a forward pass through the trained autoencoder to generate the encoded feature maps. Next, we selected three different layers within the encoder to analyze the feature activations, representing the transition from general to specific features. By performing a backward pass and computing the gradients, we weighted the gradients and feature maps to obtain the activation map. Finally, we overlaid the activation map on the input MRI image, generating a heat map that visually represented the areas contributing most significantly to the activations at each layer. This methodology allowed us to visualize and analyze the changes in feature activations across the selected layers, providing valuable insights into the hierarchical nature of the autoencoder's encoding process.

VII. Results

In this section, we present the results of training the CAE and SAE models on a single POV of healthy medical images. We evaluated the performance of the models using MSE and MAE loss functions. The reconstruction loss was used as a metric to evaluate the quality of the generated images by the autoencoders, with lower reconstruction loss indicating better autoencoder performance. Table 1 displays the results of training, validation, and testing loss for MSE and MAE across the two architectures.

Table 1: Convolutional and Stacked Autoencoder reconstruction loss across different loss functions.

		MSE	MAE
CAE	Training	0.0024	0.0272
	Validation	0.005	0.0461
	Testing	0.0049	0.0457
	Healthy	0.0032	0.0403
	Diseased	0.0068	0.052
SAE	Training	0.0076	0.0466
	Validation	0.0246	0.0932
	Testing	0.0243	0.089
	Healthy	0.0078	0.0478
	Diseased	0.04	0.1299

The results presented in Table 1 clearly show that the SAE and CAE architectures perform differently in terms of reconstruction, as measured by the MSE and MAE metrics during training, validation, and testing phases. It is also important to note the difference in reconstruction between unseen healthy and unseen diseased images. which shows that the model can detect the outliers within the dataset and allows us to classify images accordingly. The table also shows that the CAE provides better reconstruction across all loss functions.

From Figure 6, we can see the visual quality of the reconstructed images using the two models. Qualitatively, the images reconstructed by the CAE architecture appear to be more visually similar to the input images compared to those reconstructed by the SAE architecture. This suggests that the CAE architecture is better able to capture the important spatial features of the input images, which leads to better reconstruction quality.

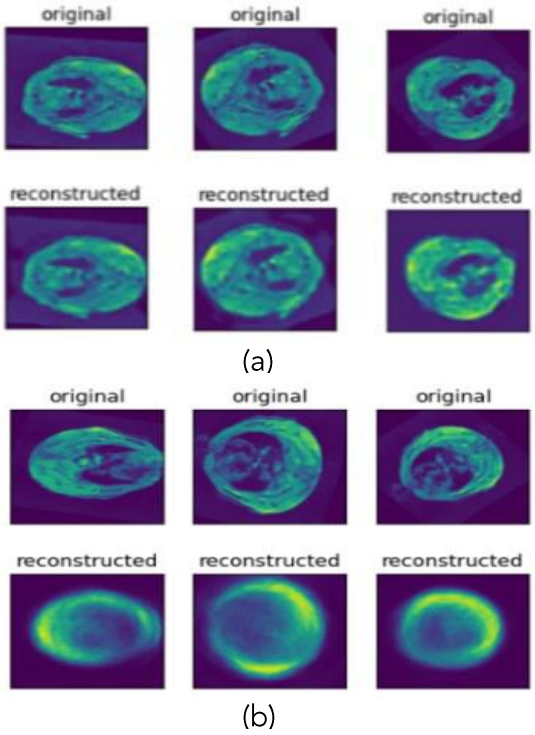

Figure 6 a and b: Original & Reconstructed Images using CAE (a) and SAE (b)

The performance of the proposed VGG16-based autoencoder is compared with other existing anomaly detection methods for MRI images. Evaluation metrics, including accuracy, precision, recall, and F1 score, are used for quantitative comparison. This quantitative analysis allowed a systematic comparison of the VGG16-based autoencoder with other established methods.

The results of this comparative evaluation revealed that the transfer learning-based autoencoder performed worse than the autoencoder trained from scratch. This judgment was reached based on a number of significant findings. First, The reconstruction loss was consistently higher for the transfer learning-based model, indicating a lower fidelity in capturing the essential information from the input images. This greater loss in reconstruction implied a lower degree of fidelity in preserving essential image elements and attributes. The visual assessments of the reconstructed images also exhibited decreased sharpness and detail compared to the autoencoder trained from scratch, as well as indicated by multiple performance metrics, including accuracy, precision, F1 score, and recall. The predicted images using a VGG16-based autoencoder are shown in *Figure 7*.

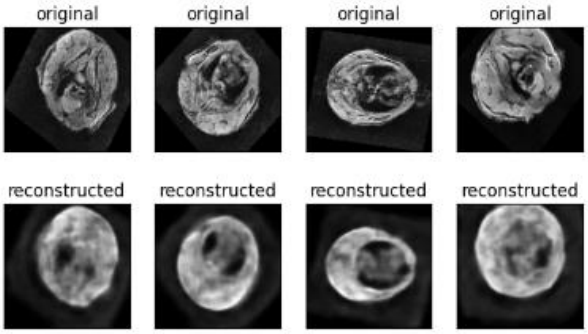

Figure 7: Original & Reconstructed Images using VGG16

In addition to evaluating the performance of the autoencoders based on their reconstruction quality, we also compared several practical aspects between the normal autoencoder and the VGG16-based autoencoder. Table 2 presents a comparison of the number of parameters, size of the code, inference time, and energy consumption for both models.

Table 2: Comparison between the performance and characteristics of convolutional and VGG16

	CAE	VGG16
Number of parameters	Total parameters (4,385) Trainable (4,385) Non trainable (0)	Total parameters (36,218,307) Trainable (21,503,619) Non trainable (14,714,688)
Size of the model	60KB	309,589KB
Inference time	0.2165ms	0.3024ms
Consumption of energy	0.00324 J	0.00453 J
Performance metrics	Accuracy: 0.9669 Precision: 0.971 Recall: 0.9571 F1 score: 0.964	Accuracy: 0.9133 Precision: 0.9375 Recall: 0.9022 F1 score: 0.8695

The normal autoencoder, trained from scratch, exhibited a lower number of parameters compared to the VGG16-based autoencoder. This indicates that the normal autoencoder had a more compact architecture, requiring fewer resources for storage and computational operations. In contrast, the VGG16-based autoencoder, which leveraged the pre-trained weights of the VGG16 model, had a higher number of parameters due to the inclusion of the pre-trained encoder.

Regarding the size of the model, which refers to the memory footprint required to store the autoencoder models, the normal autoencoder had a smaller size compared to the VGG16-based autoencoder. This is attributed to the reduced number of parameters in the normal autoencoder, leading to a more efficient utilization of memory resources.

In terms of inference time, which measures the time required for the autoencoder to process and reconstruct an input image, the normal autoencoder exhibited faster inference times

compared to the VGG16-based autoencoder. The simpler architecture of the normal autoencoder allowed for more efficient computations, resulting in quicker reconstructions.

Finally, considering the energy consumption which measure the total amount of power used by the laptop during its operation represented as thermal design power (TDP) multiplied by the total time of operation, the normal autoencoder demonstrated lower energy requirements compared to the VGG16-based autoencoder. The reduced number of parameters and faster inference times of the normal autoencoder contributed to its energy efficiency.

VIII. DISCUSSION

The obtained results reflect the effectiveness of the Convolutional and Stacked Autoencoder architectures in the context of image reconstruction. The different performance patterns between the two designs are highlighted, highlighting the potential benefits of adopting the CAE for medical picture reconstruction rather than the SAE. The fact that the models can distinguish between photos of healthy and diseased tissue based on the learned features suggests that this has implications for disease diagnosis and classification. The CAE frequently outperforms other reconstruction methods in terms of loss functions, which suggests that it can extract important information from the input images, making it an important tool for accurate image production and analysis.

The quantitative results are further supported by a visual comparison of the reconstructed photos, with the CAE-generated images showing greater fidelity to the original input images. This demonstrates how crucial it is for medical imaging applications for the CAE to accurately replicate small details and structures. The CAE's ability to maintain image quality further increases its potential for use in clinical settings, supporting medical professionals in making precise diagnoses and treatment plans.

The comparison of the VGG16-based autoencoder with existing MRI abnormality detection methods provided valuable insights into the effectiveness of learning transfer in this context. The use of rigorous evaluation scales enabled robust quantitative evaluation, highlighting trade-offs associated with the use of pre-trained models. The observed underperformance of the transfer learning-based autoencoder, as shown by increased reconstruction loss and reduced visual quality, emphasizes the difficulties transfer learning can present in some contexts. Pre-trained models like VGG16 can be useful in some situations but applying them to a given task may not always result in better results. The performance gap between transfer learning and starting from scratch stresses the importance of carefully weighing the task at hand, dataset properties, and the architecture's domain-specific adaptability.

The quantitative results are vividly supported by the visual evidence shown in Figure 7, which features both the original and reconstructed images produced by the VGG16-based autoencoder. The obvious disparities in image quality point out the VGG16-based autoencoder's limitations in accurately capturing the minute details necessary for anomaly detection in MRI scans.

In addition to the quality of the reconstruction, practical aspects such as the characteristics of the model were also taken into account. A comparison of a CAE and a VGG16-based autoencoder highlighted trade-offs between model complexity, memory footprint, inference time, and power consumption. The CAE showed advantages in terms of the number of parameters, model size, inference speed, and energy efficiency. These practical considerations underscore the importance of tailoring model architectures to the requirements of specific applications, as a CAE has proven useful in scenarios where resource efficiency is critical.

IX. Conclusion and perspectives

In conclusion, our research presents a deep learning model that can aid in the accurate and efficient detection of anomalies in MRI images. This can greatly benefit cardiologists and medical staff in identifying different types of anomalies, including atherosclerosis, with high accuracy using auto-encoders. The success of this model encourages its implementation in clinical practice. Moving forward, we plan to explore the use of Generative Adversarial Networks (GANs) to augment medical data, and also investigate other deep learning models to determine which produces the most accurate results. This will help improve the performance and reliability of our model for detecting anomalies in medical images.

Conflict of Interest Statement:

None

Funding Statement:

None

Acknowledgement Statement:

None

References:

1. Mayo Clinic. Arteriosclerosis/atherosclerosis - Symptoms and causes. Retrieved from https://www.mayoclinic.org/diseases-conditions/arteriosclerosis-atherosclerosis/symptoms-causes/syc-20350569

2. Cleveland Clinic. Atherosclerosis: Risk Factors, Symptoms & Treatment. Retrieved from https://my.clevelandclinic.org/health/diseases/16753-atherosclerosis-arterial-disease

3. WebMD. Atherosclerosis. Retrieved from https://www.webmd.com/heart-disease/what-is-atherosclerosis

4. Erbel R, Lehmann N, Möhlenkamp S, et al. Subclinical coronary atherosclerosis predicts cardiovascular risk in different stages of hypertension. *Hypertension*. 2012; 59(1):44-53. doi:10.1161/hypertensionaha.111.180489

5. Bentzon JF, Otsuka F, Virmani R, Falk E. Mechanisms of plaque formation and rupture. *Circulation Research.* 2014; 114(12):1852-1866. doi:10.1161/CIRCRESAHA.114.302721

6. Rossi A, Merkus D, Klotz E, et al. Stress-induced changes of myocardial perfusion imaging in patients undergoing computed tomography- and catheter-based invasive coronary angiography. *European Heart Journal Cardiovascular Imaging.* 2013; 14(7):624-631. doi:10.1093/ehjci/jet315

7. Rudd JHF, Myers KS, Bansilal S, et al. Atherosclerosis inflammation imaging with 18F-FDG PET: Carotid, iliac, and femoral uptake reproducibility, quantification methods, and recommendations. *Journal of Nuclear Medicine.* 2008; 49(6):871-878. doi:10.2967/jnumed.107.050294

8. Mintz GS, Popma JJ, Pichard AD, Kent KM, Satler LF. Intravascular ultrasound: clinical, angiographic, and intravascular ultrasound parameters. American Heart Journal. 2001; 142(2):200-203.

9. Nair A, Kuban BD, Tuzcu EM, Schoenhagen P, Nissen SE. Coronary plaque classification with intravascular ultrasound radiofrequency data analysis. Circulation. 2002; 106(17):2200-2206.

10. Stein JH, Korcarz CE, Hurst RT, et al. Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the American Society of Echocardiography Carotid Intima-Media Thickness Task Force. Journal of the American Society of Echocardiography. 2008; 21(2):93-111.

11. Vernhet H. Carotid Doppler ultrasound investigation of the vessel wall. European Journal of Vascular and Endovascular Surgery. 2009; 37(3):268-275.

12. ten Kate GL, van den Oord SC, Sijbrands EJ, et al. Current status and future developments of contrast-enhanced ultrasound of carotid atherosclerosis. Journal of Vascular Surgery. 2012; 56(6):1651-1658.

13. Tutar O, Turhan H, Ertas F, et al. Assessment of plaque stiffness by shear wave elastography in patients with stable and unstable carotid artery disease. Medical Science Monitor. 2016; 22:2537-2544.

14. Falk E. Pathogenesis of atherosclerosis. Journal of the American College of Cardiology. 2013; 41(7):S7-S12.

15. White CW, Wright CB, Doty DB, et al. Does visual interpretation of the coronary arteriogram predict the physiologic importance of a coronary stenosis? New England Journal of Medicine. 1984; 310(13) :819-824.

16. Mintz GS, Popma JJ, Pichard AD, Kent KM, Satler LF. Intravascular ultrasound: clinical, angiographic, and intravascular ultrasound parameters. American Heart Journal. 2001; 142(2):200-203.

17. Tearney GJ, Regar E, Akasaka T, et al. Consensus standards for acquisition, measurement, and reporting of intravascular optical coherence tomography studies: a report from the International Working Group for Intravascular Optical Coherence Tomography Standardization and Validation. *Journal of the American College of Cardiology*. 2012; 59(12):1058-1072.

18. Goldstein JA, Gallagher MJ, O'Neill WW, et al. A randomized controlled trial of multi-slice coronary computed tomography for evaluation of acute chest pain. JACC: Cardiovascular Imaging. 2007; 50(7):581-586.

19. Maffei E, Seitun S, Martini C, Cademartiri F. Coronary CT angiography in the assessment of coronary artery disease: current and emerging applications. BioMed Research International. 2012; 2012.

20. Budoff MJ, Shaw LJ, Liu ST, et al. Long-term prognosis associated with coronary calcification: observations from a registry of 25,253 patients. Journal of the American College of Cardiology. 2006; 47(11):2229-2237.

21. Raff GL, Gallagher MJ, O'Neill WW, Goldstein JA, O'Neil BJ. Diagnostic accuracy of noninvasive coronary angiography using 64-slice spiral computed tomography. Journal of the American College of Cardiology. 2005; 46(3):552-557.

22. Fayad ZA, Mani V, Woodward M. Molecular imaging of atherosclerosis: clinical state-of-the-art. Heart. 2008; 94(7):949-959.

23. Saam T, Ferguson MS, Yarnykh VL, et al. Quantitative evaluation of carotid plaque composition by in vivo MRI. Arteriosclerosis, Thrombosis, and Vascular Biology. 2005; 25(1):234-239.

24. Botnar RM, Fayad ZA. Molecular imaging of atherosclerosis. In: Handbook of Experimental Pharmacology. Springer; 2008:455-487.

25. Saam T, Hatsukami TS, Takaya N, et al. The vulnerable, or high-risk, atherosclerotic plaque: noninvasive MR imaging for characterization and assessment. Radiology. 2007; 244(1):64-77.

26. Zhao XQ, Yuan C, Hatsukami TS, Phan BA. Imaging atherosclerosis with MRI: present and future. Journal of Magnetic Resonance Imaging. 2011; 34(6):1323-1336.

27. Yuan C, Kerwin WS, Yarnykh VL, et al. MRI of atherosclerosis in clinical trials. NMR in Biomedicine. 2006; 19(6):636-654.

28. Sabbah MM, Abou Adla YA, Kasab MW, Al-Ghourabi MI, Diab MO, Aloulou NJ. Murine atherosclerosis detection using machine learning under Magnetic Resonance Imaging. *2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. Published online 2021. doi:10.1109/iecbes48179.2021.9398816

29. Terrada O, Cherradi B, Raihani A, Bouattane O. Classification and prediction of atherosclerosis diseases using machine learning algorithms. *2019 5th International Conference on Optimization and Applications (ICOA)*. Published online 2019. doi:10.1109/icoa.2019.8727688

30. Bank D, Koenigstein N, Giryes R. Autoencoders. 2021.

31. Shvetsova N, Bakker B, Fedulova I, Schulz H, Dylov DV. Anomaly detection in medical imaging with deep perceptual autoencoders. IEEE Access. 2021; 9:118571-118583.

32. Suganuma M, Ozay M, Okatani T. Exploiting the potential of standard convolutional autoencoders for image restoration by evolutionary search. 2018.

33. Kingma DP, Welling M. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning.* 2019; 12(4):307-392.

34. Seyfiog˘lu MS, O¨ zbayog˘lu AM, Gu¨rbu¨z SZ. Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities. *IEEE Transactions on Aerospace and Electronic Systems.* 2018; 54(4):1709-1723.

35. Addo D, Zhou S, Jackson JK, et al. Evae-net: An ensemble variational autoencoder deep learning network for covid-19 classification based on chest x-ray images. *Diagnostics.* 2022; 12(11).

36. Zhuang F, Qi Z, Duan K, et al. A comprehensive survey on transfer learning. 2020.

37. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2015.

38. Ferguson M, ak R, Lee YT, Law K. Automatic localization of casting defects with convolutional neural networks. 2017; 12:1726–1735.

39. Gohel P, Singh P, Mohanty M. Explainable ai: current status and future directions. 2021.

40. Amann J, Blasimme A, Vayena E, et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC medical informatics and decision making. 2020; 20(1):1-9.

41. Alsaid H, Sabbah M, Bendahmane Z, et al. High-resolution contrast-enhanced MRI of atherosclerosis with digital cardiac and respiratory gating in mice.

42. Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3D transform-domain collaborative filtering.

43. Pizer SM, Amburn EP, Austin JD, et al. Adaptive histogram equalization and its variations.