

CADD: Utilizing Convolutional Neural Networks for Binary Classification of Coronary Artery Disease

Amogh Mahadar, Rishi Venkatesh, Shrey Jain, & Varun Koduri

Received August 20, 2024

Accepted November 17, 2024

Electronic access November 30, 2024

Coronary artery disease (CAD) is the most common form of cardiovascular disease with high mortality rates in the United States. While current methods for CAD diagnoses are effective, they may be subjected to misdiagnosis. Thanks to technological advancements, machine learning has shown potential to improve the medical industry and reverse such issues. To reduce the sheer number of misdiagnoses associated with CAD and offer an alternative to improve upon current methods, we utilized convolutional neural networks (CNNs) to determine if patients have CAD, based upon an MPV scan via binary classification by analyzing divots and constrictions in the input. We initiated an InceptionV3 pre-trained model constructed through Python 3.9 Keras and TensorFlow resources, named CADD, paired with custom layers to correctly and effectively diagnose CAD. We used a public dataset from Mendeley Datasets that consists of roughly 5,000 images of Mosaic Projection View (MPV) scans of the coronary arteries, essentially a compressed angiogram. Our training and testing data were split in a ratio of approximately 80:20. MPV scans were the primary input format into our model because of their interoperability. We applied numerous data augmentations to the training dataset to increase its complexity and fine-tuned the model with multiple iterations, resulting in a significant increase in accuracy. After these procedures, CADD achieved a final accuracy of 94.67% compared to our base model's 68.24% accuracy on the same dataset. These results highlight the potential that CNNs have in the medical industry and offer a promising advancement for CAD diagnosis.

Introduction

What is coronary artery disease?

Coronary artery disease (CAD) is the most common form of cardiovascular disease and affects about 20.5 million people in the United States alone¹. CAD holds the most deaths in the United States for both woman and men. CAD is a serious and potentially life-threatening condition that requires proper medical management. If unmanaged over an extended period of time, it can subsequently cause heart attacks and even death². CAD specifically occurs when the coronary arteries constrict due to blockage, preventing the heart from receiving necessary blood and oxygen³.

This blockage is generally attributed to the buildup of plaque in the inner lining of the arteries. The plaque has a composite build of fat, cholesterol, and calcium, which circulates throughout the bloodstream and can cause atherosclerosis. Moreover, the plaque can possibly break off and cause clotting in the arteries⁴. When plaque buildup persists for a long time (roughly 5 years), the plaque hardens. This can cause what is known as calcification, which is essentially the buildup of hardened calcium in the arteries⁵.

The Prevalence of CAD

Anyone can be diagnosed with CAD. Nonetheless, the likelihood of developing the disease increases after 45 years of age in men and after 55 in woman⁶. Family history, diabetes, high LDL cholesterol, low HDL cholesterol, high blood pressure, smoking, and obesity all contribute to the increased risk of acquiring CAD³. The globalization of the United States has inadvertently increased obesity rates by increasing the availability of highly processed foods, which contain large quantities of fat. Subsequently, certain cardiac and endocrine diseases, like CAD and diabetes, have also become more prevalent. In addition to lifestyle, genetics also plays a significant role in the development of CAD. For example, people with a family history of CAD are 1.5 times more likely to obtain it⁷, and it is estimated that CAD has a 40-60% rate of inheritance⁸. Over the years, the prevalence of CAD has increased drastically. Specifically, between 1910 and 2020, coronary heart disease was expected to grow by 120% for woman and 137% for men in developing countries⁹. Since CAD is the most common form of heart disease, a significant percentage of this growth can be attributed to it. Clinicians assess CAD severity by examining the number of obstructive coronary arteries with 50% or more narrowing, categorizing patients into groups including no significant stenosis, single-vessel disease (1-VD), two-vessel

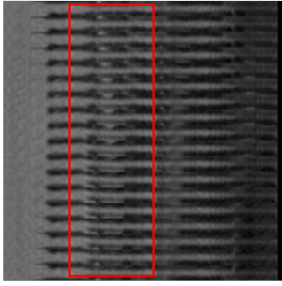


Fig. 1 Mosaic Projection View (MPV) of positive CAD patient¹⁴

disease (2-VD), three-vessel disease (3-VD), left main stem (LMS) disease, or cases with missing/no angiographic data¹⁰.

Current CAD Diagnosing Methods

There are numerous well-developed cardiac imaging techniques that doctors use to examine the arteries for the presence of CAD. The most common include: electrocardiograms, a simple, non-invasive test that records the electrical signals of the heart using small sticky dots (electrodes) and wire leads placed on specific locations of the patient's chest, arms, and legs. These signals are then translated visually on a screen or paper for medical interpretation through an electrocardiograph machine¹¹; Computed Tomography (CT) scans, a technique which uses X-rays to create detailed cross-sectional images of the heart and blood vessels, with Coronary CT Angiography (CCTA) being especially useful for visualizing coronary arteries and detecting blockages or plaques¹²; angiograms, a method often considered the "gold standard" for CAD diagnoses as they provide direct visualization of blockages through an effective, but invasive, procedure using X-rays and often contrast dye to observe the coronary arteries¹³; and Mosaic Projection View (MPV) scans, a specialized post-processing technique that generates a standardized 2D projection of the examined artery lumen along a horizontal axis. MPV scans give the appearance of multiple horizontal lines, transforming the original three-dimensional vessel structure into a series of standardized views by projecting the artery lumen onto a 2D plane. This allows clinicians to examine the entire vessel circumference simultaneously for better detection of abnormalities. For further clarification regarding how the analyzed MPV scans look, reference fig.1 below.

Issues With Current CAD Diagnosing Methods While the current methods for CAD diagnoses are effective, they may be subjected to misdiagnosis and miss early developments of CAD. Misdiagnosis rates for heart failure range from 16% to

a substantial 68% depending on the situation¹⁵. Alongside the life risks, misdiagnosis have serious medical expenses for the individuals involved¹⁶. Some of the most common misdiagnosis include nonspecific chest pain, gastrointestinal issues, musculoskeletal pain, and arrhythmias, all of which impede urgent care for patients with CAD. Moreover, some potential image artifacts or factors that can result in misinterpretation of CAD and potentially lead to misdiagnosis include movement during scanning from breathing or heartbeats, bright spots from metal stents or calcium buildup that can make blockages look worse than they are, twisted blood vessels that are hard to see clearly, and poor image quality especially in larger patients or those who have trouble holding their breath. Additionally, electrocardiogram (ECG) tests, which are one of the most commonly used cardiac diagnostic tests, risk incorrect interpretation, further contributing to misdiagnosed cases. Some, but not all, of the limitations that can ultimately result in incorrect interpretation of ECGs include precordial electrode misplacement, electromagnetic interference (EMI) from smartphones and smart devices, external forces such as construction vibrations, and insufficient sampling collected from the ECG¹⁷. A substantial 50% of fatal heart attacks occur outside of the hospital setting without receiving proper treatment due to the fact that their complication was not effectively detected by diagnosing tests. Moreover, there is low adherence to guidelines for timely ECG testing due to barriers such as suboptimal patient flow and access to necessary diagnostic technologies. Ongoing uncertainty still remains about what the most optimal method is for diagnosing CAD¹⁸. The issues associated with current methods of CAD diagnosis underscore the necessity of developing a more effective and reliable way to identify the presence of a disease. Due to this, we believe that developing an alternative to detect CAD and more effectively interpret medical imaging by leveraging the capabilities of machine learning may allow for this to be possible.

Overview of Convolutional Neural Networks

A Convolutional neural network (CNN) is a branch of machine learning and is designed to adapt based on the information that it is trained upon. A CNN is typically composed of three main layers: convolution, pooling, and fully connected layers and is designed to automatically and adaptively learn special hierarchies of features in data through a backpropagation algorithm. Because of this, CNNs excel at modeling predictions and identifying patterns in data. In a CNN, the convolution and pooling layers are responsible for the feature extraction in data, while the fully connected layers print the extracted features into a final output. The convolution layer is one of the most key components of a CNN. This layer involves numerous mathematical operations which all operate with each other to print the final output. In the case of digital images, a CNN stores

each pixel of the image into a two-dimensional grid, such as an array of numbers. A small grid of parameters known as Kernel, which is an optimizable feature extractor, is applied to the image at every position. This makes CNNs incredibly advantageous in analyzing images¹⁹.

CNNs in Medical Imaging

CNNs show much promise in the healthcare industry due to their capabilities in efficiently analyzing and processing all forms of data. This includes the possibility of inputting medical imaging into a CNN and training it accordingly. So, we aimed to create a CNN model trained on positive and negative CAD medical imaging scans. The goal was to create a model that could effectively differentiate between the two and thus allow for an accurate means of disease detection via binary classification. Moreover, we intended to implement a unique procedure known as data augmentation to strengthen the data of the medical imaging input and thus directly increase the classification accuracy of our model. This would help mitigate the risk of misdiagnosis of CAD and subsequently improve upon current methods. Our final product was CADD, a CNN model that consistently analyzes different cases of CAD without having to sacrifice levels of confidence. Our implementation of deep augmentation and advanced CNN architecture aims to set CADD apart from other models and shed light on a method to further improve future CAD detection models. CADD achieved a validation accuracy of 94.67%, a significant improvement from its base model's 68.24% accuracy prior to fine-tuning procedures.

Scope of The Study

Although our model is relatively accurate in evaluating the presence of CAD, its scope in clinical applications is limited due to the fact that our model was exclusively trained to evaluate the presence of CAD in Mosaic Projection Views (MPV) of the coronary arteries. Because of this, our model lacks the ability to detect diseases from other diagnosing imaging scans outside of the MPV. Other methods of CAD scanning will not be able to be fed into our model. Future development of CNN models like CADD to encapsulate more dataset variations will be necessary. Additionally, we do not have sufficient funding to further our study and verify our model for worldwide use in the medical setting, and our model cannot yet be implemented in the medical setting as it does not have a front-end developed for it.

Methodology Overview

CNNs require high-quality data to be able to produce accurate results. To fulfill this, we located a Mendeley dataset consisting of both positive and negative CAD scans to train CADD upon. Once this data is trained, our model will be able to identify which

patterns correspond to positive and negative CAD. We utilized Mosaic Projection View's (MPV) to accurately detect CAD without working with intricate images such as the more common angiogram. Angiograms are unique scans that vary based on the arrangement of the arteries, this causes angiograms to be more complex when training a CNN model upon. However, because an MPV offers a standardized process of projecting a 2D view of the artery lumen along a horizontal axis, giving the appearance of lines, each scan will be relatively similar, providing an effective visual for the artery's path and ultimately making it easier for our model to evaluate any potential complications in the images. This data provided from MPV scans makes them perfect to be utilized in CNNs, enhancing their ability to accurately identify and assess CAD by learning from diverse perfusion patterns.

The identification of CAD and artery complications involves a comprehensive analysis of medical images. Radiologists and clinicians utilize specific methods such as visual search patterns, preferred zoom locations, and the identification of diagnostically relevant information within the images to accurately assess the presence of tightening, constriction, or other indicators associated with plaque, cholesterol, or calcium buildup. There are various variations within each case, thus methods for detection may differ. For example, plaque buildup and calcium deposits lead to a similar outcome, however, are diagnosed in completely different ways. Plaque buildup detection is typically more detailed than calcium buildup detection. Typically, more invasive procedures are used to detect plaque buildup to provide more detailed information regarding them, whereas scans such as Computer Tomography (CT) scans are typically used for calcium buildup detection. In our case, our developed CNN model will analyze these images by utilizing multiple layers and finding stating places (epochs) and increasingly recognize certain patterns. As the program finds hotspots within the images, the CNN will identify these hotspots within a given input and use them when concluding whether the patient is diseased or not. After finding and training CADD upon our primary dataset, we employed various fine-tuning techniques, most notably data augmentation, to increase the accuracy of CADD. Data augmentation aims to increase the complexity of the training dataset, thus increasing the versatility of our model. With augmented data, our model will be trained to evaluate even the most intricate patterns in an image.

Results:

Performance Evaluation

The performance of our model was evaluated using MacOS with an M2 processor, an 8-core graphics card, and 8GB of RAM. To construct our model, we utilized the Python 3.9 Keras and TensorFlow libraries. After the process of analyzing the input images from the validation dataset, the model prints a percentage

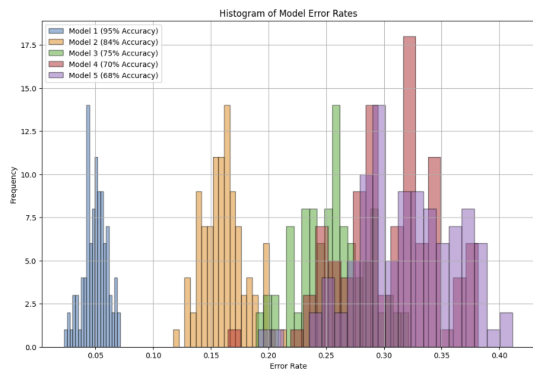


Fig. 2 Histogram of Model Error Rates

outcome based on how accurately it can classify the images. We chose Support Vector Machine (SVM) as a baseline comparison model due to its effectiveness in binary classification tasks and strong performance with high-dimensional data. The SVM model was chosen as a baseline comparison due to its structure in decision making. It works by plotting data points from the training dataset and input data onto a hyperplane. The number of hyperplanes is based upon the defined count of features. The structure of an SVM makes them highly effective for binary classification, as it simply separates data into two distinct classes using the hyperplanes²⁰. These capabilities in simplicity and strength in handling binary tasks are what we envisioned CADD to replicate, therefore it was imperative to utilize this model for comparison. This model achieved a maximum accuracy of 89.73%, an F1-score of 71.5, a precision of roughly 93.6, a sensitivity (or recall) of roughly 57.9, and a specificity of 98.8²¹. Additionally, a machine learning model specifically trained on a significant load of MPVs, like CADD, was also compared to allow for a direct, comprehensive, comparison of CAD detection performance solely through MPVs. Unlike CADD, this model was based off of a deep neural network (DNN) algorithm and achieved a maximum accuracy of roughly 0.93 and 0.91²². We utilized this model for further comparison in order to clearly assess CADD's efficiency in disease detection through MPVs with other current-made models trained on the same medical imaging technique. In doing so, we can evaluate if CADD is superior accuracy wise than current models trained on the same medical imaging type (coronary artery disease detection through MPVs). Moreover, this model differs from the previously evaluated model, SVM, since it is also trained on MPVs like CADD. Fig. 2 and Table 1 below compare the final accuracies of the different data augmentations performed on CADD. Note that each augmentation builds upon the previous one.

As shown by the two visuals, performing more data augmentations to the training dataset directly increased CADD's final classification accuracy. Performing various data

Table 1 CADD Performance Analysis for Primary Dataset

Model	Augmentation	Input type	Performance %
1	Base Model (ADAM Optimizer)	MPV	Accuracy: 68%
2	Horizontal and Vertical Shifts	MPV	Accuracy: 70%
3	Rotate by 40°	MPV	Accuracy: 75%
4	Rescale by a Factor of 1/255	MPV	Accuracy: 84%
5	Zooming and Flipping	MPV	Accuracy: 95%

augmentations improves CNN detection accuracy by artificially expanding the training dataset with modified versions of the original images. This creates variations that help the model learn to recognize the same features under different conditions (like different orientations, scales, or positions). These variations make the model more robust and generalizable as it learns to identify key features regardless of small transformations that might occur in real-world data. This process ultimately improves the model's ability to classify new, unseen images accurately, making it more implementable for clinical practice.

Statistical Significance and Clinical Relevance

After performing rigorous testing and data augmentations on our model, it achieved a final accuracy of 95% in classification of CAD through our training dataset. However, regardless of impressive accuracy, it is crucial to note that this number alone does not encapsulate the full scope of quality in such healthcare models. Coding systems, such as those involving AI automation, must also be subjected under continuous monitoring for completeness, relevance, and alignment with clinical and payer policies²³. Furthermore, specific use cases, such as value-based care or clinical registries, may require different quality metrics beyond simply raw accuracy to meet healthcare system requirements. This comprehensive approach to approving healthcare devices for official practice ensures that models are suitable not just for accuracy targets but also for their intended purpose in real-world healthcare settings.

Performance Metrics

To further assess the classification ability of CADD and provide a better visual, more accurate, assessment of our model's performance, we utilized 5 more means of statistical analysis: The confusion matrix, a F1 Curve, a Precision-Confidence curve, a Recall confidence curve, and a Precision-Recall curve. In these graphs, the terms precision, recall, are frequently seen. Precise definitions of both will be provided below.

Precision

Precision refers to the number of true positives divided by total number of positive predictions²⁴. In our case, this would mean the amount of MPV scans that our model correctly identifies as having CAD out of all the scans that actually have it.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Fig. 3 Precision and Recall formulas²⁵.

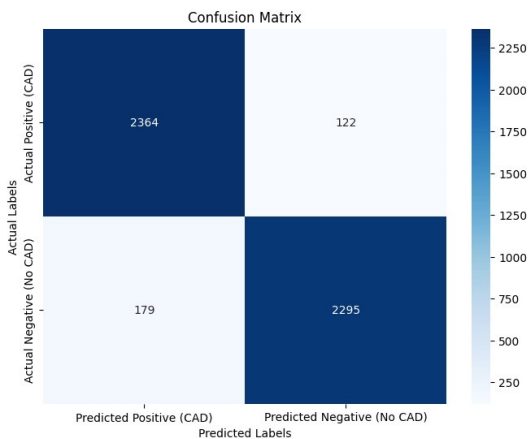


Fig. 4 Confusion matrix for CADD

Recall

Recall measures how correctly our model identifies these true positives²⁴. In our case, this means for all the MPV scans that actually have CAD, recall shows how many scans we correctly identified. Fig. 3 clarifies how precision and recall is calculated through formulas.

Confusion Matrix

In machine learning, a confusion matrix is used to effectively determine the performance of a model's classification by comparing predicted values alongside actual values²⁶. The main goal of a confusion matrix is to provide a clear visual summary of the performance of a model. Utilizing a confusion matrix allows us to assess exactly how many images the model is correctly classifying and calculate our model's specificity. We created a confusion matrix to evaluate CADD's performance in accurately determining if CAD is present or not in our input MPV scans. Reference fig.4 below.

Given these numbers, we can calculate the specificity of our model. Specificity measures the proportion of actual negatives that were correctly identified as negatives. In our case, it measures the proportion of images that our model correctly identified as negative CAD. Calculating our model's

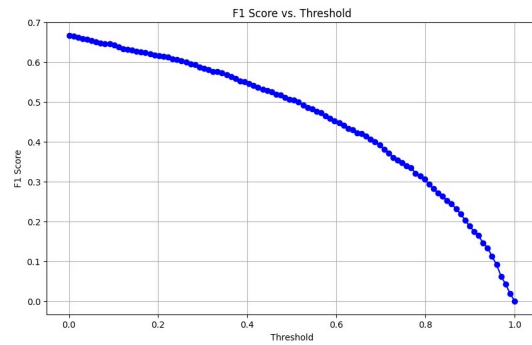


Fig. 5 F1 Curve Graph for CADD

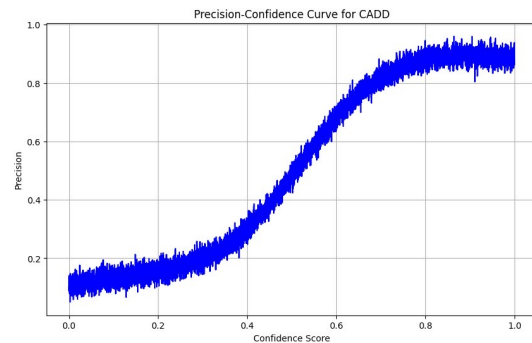


Fig. 6 Precision-Confidence curve graph for CADD

specificity provides an effective metric for us to clearly analyze its CAD classification performance. The formula for calculating specificity is $TN / (TN + FP)$, where TN = true negatives and FP = false positives. Given this formula, the calculated specificity of CADD is roughly 93.3%.

F1 Curve

We also utilized an F1 Curve vs Threshold graph to evaluate the performance of CADD. This graph displays the F1 score (harmonic mean of precision and recall)²⁷. An F1 graph encapsulates the precision and recall performance of a classifier model and allows us to visualize how the model performs across a range of possible thresholds²⁸. A higher F1 score indicates that our model is achieving high precision and high recall at that moment. Reference fig. 5 below for a clear visual of the F1 Score vs. Threshold graph for CADD.

Precision-Confidence Curve

A precision confidence curve measures how the precision of model classification changes depending on the confidence interval²⁹. This makes for another useful evaluation of CADD's classification performance. Fig. 6 below displays the precision-confidence curve for CADD.

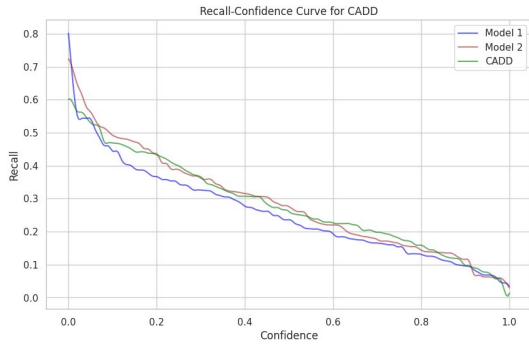


Fig. 7 Recall-Confidence curve graph for CADD

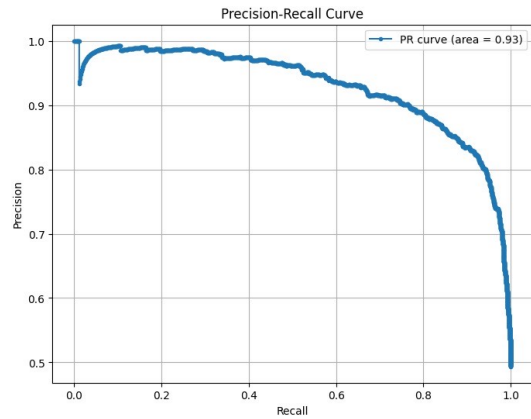


Fig. 8 Precision-Recall graph for CADD

Recall-Confidence Curve

Similar to a precision-confidence curve, a recall-confidence curve measures how the recall of the model classification changes depending on the confidence interval³⁰. Fig. 7 below displays the recall-confidence curve for CADD. Model 1 includes data augmentations up to rotating by 40 degrees, whereas model 2 includes all the augmentations up to rescaling by a factor of 1/255. CADD represents the final model with all the augmentation changes performed. As the confidence threshold increases, the recall value decreases, meaning the model identifies fewer true positive CAD cases. This decline in recall is particularly critical in medical diagnostics, where missing a CAD case could lead to delayed treatment and potential patient harm. However, our extensive data augmentation strategy helps mitigate this risk. By introducing controlled variations in the training data through multiple augmentation techniques (angle, size, and zoom adjustments), we expand the model’s ability to recognize CAD patterns across diverse patient populations. Each additional layer of augmentation improves the model’s efficiency by exposing it to a broader range of possible CAD presentations, effectively teaching it to maintain higher recall rates even at increased confidence thresholds. This is evidenced by CADD’s increasingly superior performance as we performed more data augmentations, exemplified by table 1 above.

Precision-Recall Curve

Lastly, a Precision-Recall graph was used to evaluate the performance of the binary classification of CADD. This graph plots precision against the recall at various thresholds. Utilizing a Precision-Recall graph allows us to further evaluate the classification performance of CADD across multiple thresholds³⁰. Fig. 8 below displays the Precision-Recall graph for CADD.

Table 2 Comprehensive Performance Analysis of CADD vs Compared SVM and Angiogram Implementation

Method	Accuracy (%)	Precision (%)	Sensitivity (recall) (%)	F1-Score (%)	Specificity
CADD	94.68	~94 (max)	~60 (max)	68.2 (max)	93.3
SVM	89.73	93.6	57.9	71.5	98.8
CT Coronary Angiography	91 ³¹	N/A	90 ³¹	N/A	40 ³¹

Performance Comparison:

To comprehensively analyze CADDs performance with both existing AI models and traditional diagnostic methods, we created a detailed table. We chose angiogram implementation as our traditional diagnostic method as it is typically the most commonly used one. Table 2 below presents a comparison table with statistical analysis comparing CADD, our proposed model, to both existing AI models and traditional diagnostic methods.

Discussion

There has been an increasing demand for a lightweight, low-cost CAD detection method. Recent models often implement stand-alone CNNs paired with custom layers. However, our model’s implementation of data augmentation and advanced CNN architecture sets it apart from other models. Prior to performing data augmentations and numerous other tuning techniques to the model, the original validation accuracy was 69% and increased significantly after each following augmentation. This pattern shows that extensive fine-tuning of data significantly improves a CNN model’s accuracy. The final validation accuracy of the model is roughly 95%. Although this accuracy exceeded our initial expectations, it does not confirm the ability for our model to be successfully implemented for clinical practice, as many other factors other than raw accuracy will need to be evaluated to confirm its success. If our model meets the other requirements

needed to be put into meaningful use in clinical practice, then our objective to create an improved alternative for an accurate disease diagnoses method through the capabilities of machine learning will be fulfilled. Regardless, our developed model, CADD, shows improved performance, notably accuracy, over the other CAD detection models that were compared. This observation sheds light on a potentially improved method of disease detection than pre-existing machine learning models intended to perform the same function. Moreover, as shown by the performance comparison table with existing AI Models and traditional diagnostic methods, we can see that CADD demonstrates the highest accuracy, however metrics such as precision, sensitivity, F1-score, and specificity are relatively the same with the compared SVM. The sensitivity of the compared traditional diagnostic method appears to be significantly better than CADD and the compared SVM, however its specificity is lower than both models and its accuracy is better than the compared SVM but lower than CADD. The results collected by the confusion matrix show that CADD is correctly classifying a vast majority of the contents of the dataset (2364 positive and 2295 negative) as opposed to only approximately 300 misdiagnoses. Though this is impressive, it is important to note that this does not guarantee our model's effective performance in the clinical setting, as this only shows that our model performed well with the provided dataset. In the F1 curve graph, the F1 score's decrease across thresholds reflects a fundamental precision-recall trade-off in our model. At lower thresholds, the model achieves both good precision and recall, resulting in higher F1 scores. However, as the threshold increases, while the model gains precision by reducing false positives, it loses recall by missing true CAD cases. This trade-off significantly impacts the model's performance in the clinical setting: higher thresholds make the model more conservative, reducing unnecessary interventions but potentially missing cases requiring treatment. Moreover, given this graph, it can be seen that the maximum F1-score of roughly 68% occurs at a threshold that is close to 0.0. The precision-confidence curve shows that as the precision increases, our model's confidence subsequently increases as well. This pattern suggests that when our model is confident in its classifications, they are likely to be correct. In this graph, it can be seen that the precision is at its maximum of roughly 94% at a confidence level close to 1.0. The recall-confidence graph shows that the recall decreases as the confidence of our model increases. This indicates that our model is becoming more selective as its confidence increases, which means the model is identifying fewer true positives. In a medical setting, this reduction in recall could cause extreme problems. Missing a potential diagnosis could have serious clinical consequences to the patient and to the hospital. For the patient, a missed diagnosis of coronary artery disease could lead to delayed treatment, disease progression, increased risk of heart attack or stroke, and potentially life-threatening complications. For the

hospital, a missed diagnosis may result in medical malpractice lawsuits, significant financial liability, damage to institutional reputation, and potential loss of patient trust. Moreover, our augmented final version of CADD demonstrates a greater decrease in recall than its other compared models as the confidence interval increases, suggesting that the increased data augmentation impacts the model's classification behavior. The more extensive augmentation potentially creates a more efficient model by introducing variability in the training data, which helps the model learn more generalized features and become more efficient in classifying coronary artery disease (CAD). However, the increased selectivity also suggests that the augmented model is developing a more refined decision boundary, potentially trading some recall for improved precision. Lastly, in the precision-recall graph the curve area is 0.93, which is extremely close to the maximum area score. These areas range from 0 to 1, where a score of 1 demonstrates that the model is perfect in achieving perfect precision and recall across all thresholds. However, in the graph there is a divot towards the beginning. This divot likely indicates slight faults in our data. Collectively, these performance metrics demonstrate that CAD is relatively sufficient in detecting CAD in MPV scans.

Limitations

Even though our model can evaluate the condition of the patients relatively accurately, it needs large, high-quality datasets to improve its accuracy on a general level. As a result, it goes without saying that the dataset may have contained several potential sources of errors such as inconsistent labeling by medical professionals during annotation, varying image quality and acquisition protocols between medical centers, demographic biases in patient representation, and incomplete documentation of medical conditions in the training data. These data quality issues could negatively impact the model's diagnostic reliability when applied to proper clinical practice. It can be seen that data quality directly impacts patient safety and treatment outcomes in several ways. For example, errors in the dataset could lead to detrimental impacts such as misdiagnosis resulting in inappropriate treatment plans, delayed intervention in critical cases due to false negatives, unnecessary medical procedures from false positives, and systematic biases in diagnosis across different demographic groups. Due to these risks, we would need to ensure that all data being tested goes through rigorous validation protocols and continuous monitoring of model performance across diverse patient populations to prevent any risk to the patient's life or their quality of life. Additionally, we encountered issues such as overfitting and data constraints during the development of the model. As a result, CADD's accuracy was significantly higher on the training dataset than on the testing dataset. This suggests that the model was adapting to only specific patterns instead of generalizing. Although methods

such as drop-out regularization, data augmentation, and early stopping were implemented to deal with overfitting, they were not sufficient to eradicate the issue altogether. These issues prove the necessity of ensuring that the data being implemented into a CNN model is extremely high quality and diverse. Additionally, we found that the model training time for CADD was extremely long (more than 2 hours). In the medical setting, this means that it will take a substantial amount of time for our model to be trained upon medical imaging scans of a patient. Lastly, while our CNN model shows success, it is not yet implementable for accessible use of doctors in the medical setting since there is not a sufficient application to store our model.

Future Direction

In the future, to make our model more implementable in the medical setting, we hope to develop a front-end for our model. This will allow for a more user-friendly interaction, allowing doctors to easily upload medical imaging scans and interact with our model without advanced technical knowledge; ensure a visually appealing model, where the printed results of our model can be more organized; improve navigation for doctors to access patient data, previous scans, and results; ensure responsiveness to guarantee that our model works consistently across different devices, which will allow doctors to access our model with ease; and improve accessibility, such as implementing features such as zoom-in tools, screen readers, and diverse language support. Implementing a front end on our model and on future models can potentially revolutionize disease diagnoses methods and allow for our model to be easily implemented in hospitals all around the world. Furthermore, as we traversed through the training dataset, we found that data preparation was extremely time-consuming. Having to classify and label thousands of medical data was tedious. However, the annotation tool allowed us to speed up the preparation process by making use of InceptionV3's uncertainty estimates. As a result, this would repeatedly refine our dataset's tags and help improve our data from a general point of view. Taking this concept into perspective, we would see similarities to neural ordinary differential equations (ODE), which show much potential for increasing the interpretability of our dataset with respect to our model. With the progression of continuous depth functions, an increased possibility of a more accurate CAD detection is possible. As neural ODEs allow for more diverse data inputs placed at different times of CAD progression, our model would be able to understand the progression of CAD within the body. To assist in the damage control of memory, we used the implementation of regions of interest (ROI) pooling while still managing to keep the essential indicators of CAD. If we were to pair this with automated detection algorithms, we would be able to easily remove much of the overhead that the model normally must complete to be able to provide solutions. In addition to using stronger methods to

help reduce overhead processes, we could merge our image data with medical records such as patient history, genetic history, and biomarkers. Using other sources of concentrated data can help provide a bigger picture of a patient's diagnosis and thus allow for a more nuanced observation with additional factors that could influence a disease. Models could take text and integer values based on patient demographics and history while aligning medical screening. Such a multi-modal approach can assist in achieving greater validation accuracy as well as a more comprehensive view of a patient's health. Not only will this allow for a more accurate approach in evaluating the likelihood that CAD is present, but it could also allow for the detection of disease like CAD. A simple script could also be made to analyze and clean data within the dataset to remove faulty or corrupted data. Since neural networks learn based on what they are given, inputting faulty data would render the model faulty and inaccurate. Lastly, to reduce model training time, future models can be run on more advanced hardware, such as on a sophisticated GPU or TPU.

Closing Thoughts

Despite limitations and problems encountered, the results shown by CADD offer tangible evidence on machine learning's positive impact on improving the medical setting. CNNs show promise with its capabilities and implications with not only detecting coronary artery disease but also a wider range of illnesses. If experts expand on the concept of machine learning in analyzing medical imaging scans, like CADD, a more accurate, reliable, and timely means of disease diagnoses is possible. The future of machine learning shows much hope in improving the wellbeing and care of patients all over the world.

Datasets and Methods:

This paper presents an experimental machine learning study. In this study, we developed and tested a CNN model, named CADD, to correctly classify MPV scans of coronary arteries from patients with and without CAD. We manipulated variables such as our model architecture and through data augmentation to maximize the model's performance. The study utilizes a dataset of thousands of MPV scans to train the model upon which was collected at a single point in time.

Data Collection

To begin the development of CADD, we needed to collect data to train this model upon. CNNs require high-quality data to be able to produce accurate results. Because of this, we needed at least 1,000 accurate scans of coronary arteries to train CADD. Among these scans, it was required that over 500

possess CAD and the other half does not. Collecting high-quality data ensures that our CNN model produces the most accurate and precise results when trained. To search for data, we traversed through numerous dataset libraries. Eventually, we located a structured dataset with 4,768 images consisting of only MPV scans at the Mendeley Datasets. This significantly exceeded our initial expectations. The dataset included a split of roughly 50% positive and 50% negative CAD scans and consisted of three subsets: one for training the model, one for testing the model, and one for validating the model. These subsets were then further divided into 3 categories following an approximate 3:1:1 ratio for training, testing, and validation, respectively. The training data set images originally consisted of 300 total patient scans; however, these images were augmented 6-fold to significantly multiply the amount of data and end with 2,364 images. This augmentation was explicitly performed on the training case of the dataset to strengthen modeling and dataset balance, and ultimately improve our model's accuracy with more images to learn off of. Augmentation was not performed on the normal component of the training set, the entire validation dataset, and the entire testing dataset. Moreover, in the validation dataset, only one artery was selected at random per normal case (50 images) and per diseased case (50 images) for balance maintenance. The Images were all rendered as a PNG with dimensions of 299 x 299 pixels Demirer *et al.*¹⁴. The reason why we chose to utilize a dataset consisting solely of MPVs over the more common angiogram is because MPVs would be more effective to implement in the case of machine learning. Since an MPV in this dataset projects a 2D view of the examined artery lumen of an analyzed patient along a horizontal axis, giving the appearance of lines, each scan will be relatively similar. This makes them perfect to be utilized in CNNs, as the MPVs will provide an effective visual for the artery's path and ultimately enhance our model's accuracy to evaluate any potential complications in the images by learning from diverse perfusion patterns. This standardization process ensures that regardless of the original anatomical arrangement, the resulting MPV presents vessels in a consistent format, making it more suitable for automated analysis compared to traditional angiograms, where vessel arrangements vary significantly based on patient anatomy. Note that all CCTA image-dataset utilization was retrospective, performed locally under Institutional Review Board approval (including HIPAA compliance) with the waiver of patient consent¹⁴. Moreover, factors such as patient demographics, age, sex, and disease severity were not available as they were kept confidential due to the rules set forth by the HIPAA.

Data Augmentation

While substantial data augmentation was already performed on the original dataset we located, we implemented even more

extensive data augmentation on our data to maximize the diversity and strength of the training dataset to ultimately improve the effectiveness of CADD. Data augmentation aims to expand the diversity of the training dataset³¹. Data augmentations were exclusively applied to the training dataset to maintain similarity with real-world scans in the testing and validation datasets. The specific data augmentations utilized include shifts along the width and height of the images; 40-degree rotation, which would further increase the accuracy by accounting for different orientations of the input; scaling the images by a factor of 1/255 to normalize the values of the pixel to either 0 or 1; and zooming and flipping to simulate varying distances and mirror-image variations. Applying augmentations, or transformations, to the images in the dataset exposes the model to a wide range of variations in the input images, like real-world scenarios. This process also reduces overfitting, which can significantly affect the training accuracy while also hindering the testing accuracy. After augmenting the data, the strength of our data significantly increased (multiplied by a factor of 36). We believed that with more augmentation performed on the data, the more accurate our model would be in detecting CAD.

Model Construction

TensorFlow resources³² and Python 3.9 Keras were utilized to help construct and improve CADD. InceptionV3, a pre-trained model trained on ImageNet, was also implemented to greatly assist in the development of CADD. InceptionV3 leverages pre-trained features and transfer learning capabilities from another domain, allowing us to build upon established computer vision patterns rather than building our own CNN from scratch. This ultimately simplified the construction of our model and allowed us to focus on augmentation based on our own parameters instead. InceptionV3 is a CNN that has 48 layers. It is often used for loading pre-trained models that are then customized as necessary³³. Moreover, inceptionV3 includes several improvements over previous architecture such as Label smoothing, factorized 7 x 7 convolutions, and the use of an auxiliary classifier to propagate label information lower down the network³⁴. 50 total layers were implemented, which included 48 customized layers, 0 dropout layers, 1 fully connected layer, and 1 Softmax layer. The customized layers are designed to perform a specific task, the fully connected layers are the decision-making layers of the CNN, and the Softmax layers classify an image via binary classification in our case. The network has the capability to classify images into over 1,000 categories. To implement data augmentations in our training dataset, we used TensorFlow's "ImageDataGenerator"³². Performing these data augmentations on the training dataset increased CADD's rigor and validation accuracy, which is how accurately the model can diagnose CAD on a different dataset based on the augmented training

Table 3 Hyperparameter Settings for CADD

Hyperparameter	Value	Description
Learning Rate	0.001	Initial learning rate for training.
Batch Size	32	Number of images processed in each batch.
Epochs	10	Number of training epochs.
Dropout Rate	0.3	Rate of dropout applied to reduce overfitting.
Optimizer	Adam	Optimization algorithm used.

dataset. Validation accuracy is displayed after running through 10 epochs (ten iterations that would start at a different location within the given images) in the training dataset. The drop out ratio implemented into our model was approximately 0.3. The dropout ratio is a parameter that represents the percentage of neurons that are randomly deactivated during phase training to prevent overfitting. Overfitting negatively affects the accuracy, and the dropout ratio directly associated with the accuracy. To optimize the model and ensure its maximum performance by working with a substantial dataset, we utilized the Adam Optimizer, or Adaptive Moment Estimation. Adaptive Moment Estimation is an optimization algorithm for gradient descent. This algorithm is extremely effective when dealing with a lot of data or parameters³⁵. Table 3 below displays the exact model settings utilized in CADD.

Mathematics Involved in CNNs

The mathematics behind a CNN are a group of matrices that look for similarities between what our model already knows (from the training) and the input MPV³⁶. To arrive at a decision, our model uses multiple methods such as the Rectified Linear Function (ReLU), gradient descent through backpropagations, and SoftMax to classify the given images. The ReLU function acts as the starting activation point³⁷. It works by taking the maximum value of a given range (from x, or in this case, U), where the program will be looking at the input image in sections and determining it either as zero or a value greater than zero. In our case, the model applying the sequential approach (the ReLU function) will figure out parts that are unnecessary. For example, on an epoch the image will first look from left to right to read whether there were plaque blockages to figure the nonessential starting points to reduce the possibility of being “stuck within a curve” and allows for our model to move away from a linear perspective to more realistic representation. Then, the convolutions that the program implements act as a focal point by simplifying the runtime of our neural network when it works between the multiplication of two functions while simultaneously editing the parameters as the calculated output moves through the network. The pooling layers (GlobalAveragePooling2D) help reduce the dimensions of our program as it runs through the training dataset. Our project uses the SoftMax function to help classify the image by taking

$$ConvolutionalLayer : (I \cdot K)(x, y) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} I(x+i, y+j) \cdot K(i, j)$$

$$Activation\ Function\ (ReLU) : \max(0, x)$$

$$Pooling\ Layer : P(x, y) = \max_{i=0}^{p-1} \max_{j=0}^{p-1} F(x \cdot s + i, y \cdot s + j)$$

$$Fully\ Connected\ Layer : z = W(x) + b$$

$$Softmax\ function\ (for\ final\ classification) : \sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^N e^{z_j}}$$

Fig. 10 Heavily utilized mathematical formulas in CNNs

the input (in the form of an exponent with its base as Euler’s number) over the output (in the form of an exponent with its base as Euler’s number). The ADAM optimizer (Adaptive Moment Estimation), which works to help increase the accuracy of our program by taking the “exponential weighted average” of the gradients, helps decrease the overall run time³⁸. Fig. 10 below displays the heavily utilized mathematical equations utilized in CNNs.

Data Analysis

In the MPV, the main thing put into consideration was the presence of any divots or constriction along the lines of the 2D projected artery lumen. That was the major varying factor between each image. In our dataset, the images were previously classified with positive and negative CAD. So, we simply trained our model thoroughly with each set of respective images so it can evaluate the image and increasingly recognize any distinct patterns that make an image positive or negative. When training our model on the augmented training dataset, our model processes these images and evaluates them utilizing its various implemented layers and mathematical functions. Then, it will predict which category (positive or negative) this image belongs via binary classification by comparing the resemblance of the patterns of the current image to the patterns of positive and negative CAD scans contained in the training dataset. To visualize what the process of evaluating an image looks like in CADD, we generated a feature map of an input MPV scan with resources from TensorFlow³² and Keras 3.9. Utilizing a feature map allows us to visualize how our model is processing the input image and evaluating for any subtle variations in it. These subtle variations will then allow our model to conclude whether CAD is present or not via binary classification. To produce this feature map, the convolutional layer of our model applies numerous filters to our input image. The images from these feature maps are then subjected under an activation function, such as ReLu, to emphasize significant features. Next, the pooling layers of

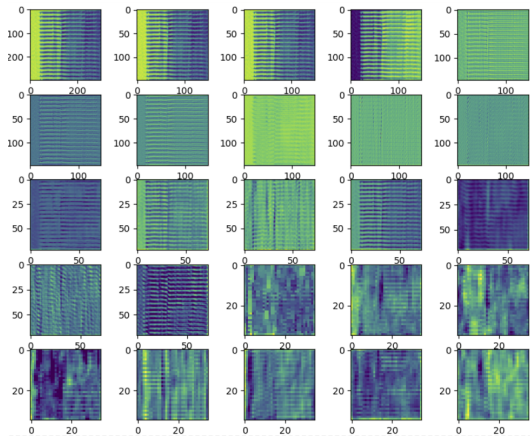


Fig. 11 Representation of visual patterns and features in the input image

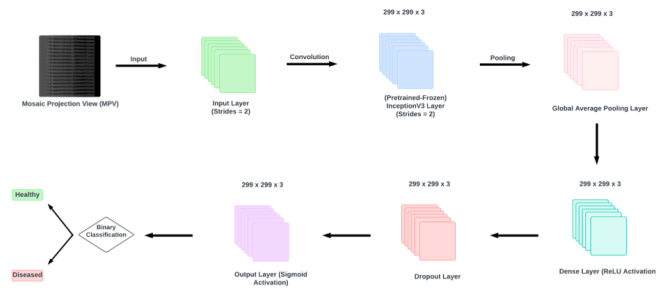


Fig. 12 Model overview for CADD

our model reduce the computational complexity of the images while still retaining critical information. This process allows our network to learn and represent complex patterns, making it extremely accurate in identifying any subtle changes in the images and thus classifying them accordingly via binary classification. Fig. 11 below displays the feature map for an input MPV.

Overview of Model Structure

Utilizing the mathematics shown in fig.10 above and pairing this with TensorFlow’s multiple methods, our program is able to evaluate a substantial dataset swiftly. After evaluating the contents of the dataset, our program comes to the robust binary classification of 1 or 0, or, in other words, positive or negative, of CAD. The overall process of our model, from the input MPV scan to the output depicting if CAD is present or not, is shown in fig.12 below.

Conclusion

This paper presents a novel approach to CAD diagnoses through use of machine learning. We developed a CAD detection model that uses MPV scans to reliably identify the existence of coronary artery disease. This model is built using TensorFlow resources and InceptionV3 architecture and is based on a convolutional neural network (CNN). A CNN is a branch of machine learning, where a machine learns based on the data that is input into it. Roughly 5,000 Mosaic Projection View scans with initial dimensions of 299 x 299 pixels served as the main dataset on which we trained our model on. These images were divided into positive and negative CAD scans and further divided into three subsections for training, testing, and validating the model. Various functions within the model enabled it to accurately assess the presence of CAD. However, we had to perform numerous fine-tuning procedures to achieve successful results. This underscores the importance of tuning parameters such as batch size, learning rate, and number of epochs to make sure that a CNN model is successful. The findings shed light on the advantageous implications of machine learning in the medical industry for effectively detecting diseases such as CAD. Similar constructed models intended to detect CAD achieved a relatively lower accuracy than our proposed model, CADD. AI/Machine learning offers numerous advantages over humans, such as efficiency, precision, and pace. These capabilities allow AI to make instantaneous decisions, which will be helpful when taking large amounts of data or treating numerous patients at once. Such models built upon the concept of AI will revolutionize the approach for medical diagnostic practices as it will slowly start allow for resources to be allocated for data collection instead of continuing with dated diagnostic techniques. As with more resources, machine learning algorithms will be able to grow to be greater in size with greater variations in their learning data sets which in turn will allow for greater accuracy and provide inexpensive diagnosis for patients across the world. Furthermore, they will provide a low-cost alternative for disease detection and drastically improve upon current methods. Medical related CNN models will drastically aid populations around the world who do not have proper access to healthcare and are in dire need of health evaluations that are currently deemed unreachable due to the significant operating and transportation costs of current methods. In addition to reaching underserved populations, our program will improve the populations that already have access to healthcare, helping to mitigate the risk of misdiagnosis worldwide. However, further study is needed to reduce model training time and verify the model across a larger range of datasets, such as those with more diverse patient demographics and better-quality imagery to ensure maximum performance. Additionally, development of an advanced front-end for our model is necessary to allow our model to be implemented into the widespread medical setting.

Acknowledgements

We would like to thank Dekyi Mcknight, University of Texas at Dallas, for his guidance on this paper.

References

- 1 National Heart, Lung, and Blood Institute (NIH), *What Is Coronary Heart Disease?*, 2023, <https://www.nhlbi.nih.gov/health/coronary-heart-disease>.
- 2 Yale Medicine, *Coronary Artery Disease (CAD)*, 2022, <https://www.yalemedicine.org/conditions/coronary-artery-disease>.
- 3 Mayo Clinic, *Coronary artery disease - Symptoms and causes*, 2022, <https://www.mayoclinic.org/diseases-conditions/coronary-artery-disease/symptoms-causes/syc-20350613>.
- 4 American Heart Association, *What is Atherosclerosis?*, 2024, <https://www.heart.org/en/health-topics/cholesterol/about-cholesterol/atherosclerosis>.
- 5 Cleveland Clinic, *Coronary Artery Calcification: Causes, Symptoms & Treatment*, 2022, <https://my.clevelandclinic.org/health/diseases/22953-coronary-artery-calcification>.
- 6 American Heart Association, *Coronary Artery Disease - Coronary Heart Disease*, 2024, <https://www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-disease/coronary-artery-disease>.
- 7 J. M. Edwards, *Family History of Coronary Heart Disease? It Might Be Your Genetics*, 2022, <https://www.healthline.com/health/is-coronary-artery-disease-genetic>.
- 8 R. McPherson and A. Tybjaerg-Hansen, *AHA Journals*, 2016, **118**, 564–578.
- 9 T. A. Gaziano, A. Bitton, S. Anand, S. Abrahams-Gessel and A. Murphy, *Science Direct*, 2010, **35**, 72–115.
- 10 O. C. H. G. Ozcan, Deleskog, *European Heart Journal*, 2018, **4**, 25–35.
- 11 Better Health Channel, *ECG test*, 2020, <https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/ecg-test>.
- 12 T. Cox, *Tomography - an overview | ScienceDirect Topics*, 2020, <https://www.sciencedirect.com/topics/medicine-and-dentistry/tomography>.
- 13 Better Health Channel, *Coronary angiogram*, 2012, <https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/coronary-angiogram>.
- 14 M. Demirel, V. Gupta, M. Bigelow, B. Erdal, L. Prevedello and R. White, *Image dataset for a CNN algorithm development to detect coronary atherosclerosis in coronary CT angiography*, 2019, <https://data.mendeley.com/datasets>.
- 15 C. W. Wong, J. Tafuro, Z. Azam, F. Z. Ahmed, C. Mallen and C. S. Kwok, *Journal of Cardiac Failure*, 2021, **27**, 925–933.
- 16 M. M. Wilson, *The Dangers of Medical Misdiagnosis*, 2017, <https://wilsonlaw.com/blog/the-dangers-of-medical-misdiagnosis/>.
- 17 HILLROM & WELCH ALLYN, *Three Factors That Could Be Impacting Your ECG Interpretation*, 2024, <https://www.hillrom.eu/en/knowledge/article/three-factors-that-could-be-impacting-your-ecg-interpretation/>.
- 18 M. Kodeboina, K. Piayda, I. Jenniskens, P. Vyas, S. Chen, R. J. Pesigan, N. Ferko, B. P. Patel, A. Dobrin, J. Habib and J. Franke, *MDPI*, 2023, **20**, 5633–5633.
- 19 R. Yamashita, M. Nishio, R. K. G. Do and K. Togashi, *Springer Open*, 2018, **9**, 611–629.
- 20 Aswathisidharan, *Support Vector Machine Algorithm*, 2024, <https://www.geeksforgeeks.org/support-vector-machine-algorithm/>.
- 21 S. Saeedbaksh, M. Sattari, M. Mohammadi, J. Najafian and F. Mohammadi, *Advanced Biomedical Research*, 2023, **12**, year.
- 22 V. Gupta, M. Demirel, M. Bigelow, K. J. Little, S. Candemir, L. M. Prevedello, R. D. White, T. P. O'Donnell, M. Wels and B. S. Erdal, *Journal of Digital Imaging*, 2019, **33**, 431–438.
- 23 Robinson, *Decoding the '95% Accuracy' Standard: Ensuring Consistent Quality Metrics in Medical Coding*, 2024, <https://www.codametrix.com/medical-coding-ai-95-percent-accuracy-standard/>.
- 24 P. Huilgol, *Precision and Recall in Machine Learning*, 2024, <https://www.analyticsvidhya.com/blog/2020/09/precision-recall-machine-learning/>.
- 25 K. P. Shung, *Accuracy, Precision, Recall or F1?*, 2018, <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>.
- 26 P. Singh and A. Singh, *Confusion Matrix - an overview | ScienceDirect Topics*, 2021, <https://www.sciencedirect.com/topics/engineering/confusion-matrix>.
- 27 N. Sharma, *Understanding and Applying F1 Score: AI Evaluation Essentials with Hands-On Coding Example*, 2023, <https://arize.com/blog-course/f1-score/>.
- 28 GPTutorPro, *F1 Machine Learning Essentials: Optimizing F1 Score with Threshold Tuning*, 2024, <https://gpttutorpro.com/f1-machine-learning-essentials-optimizing-f1-score-with-threshold-tuning/>.
- 29 L. Yu and S. Liu, *International Journal of Advanced Computer Science and Applications (IJACSA)*, 2023, **14**, year.
- 30 D. Steen, *Precision-Recall Curves*, 2020, <https://medium.com/@douglaspsteen/precision-recall-curves-d32e5b290248>.
- 31 CCS Learning Academy, *What is Data Augmentation? Techniques, Examples & Benefits*, 2024, <https://www.ccslearningacademy.com/what-is-data-augmentation/>.
- 32 TensorFlow, *Convolutional Neural Network (CNN) | TensorFlow Core*, 2019, <https://www.tensorflow.org/tutorials/images/cnn>.
- 33 Keras Team, *Keras documentation: InceptionV3*, 2024, <https://keras.io/api/applications/inceptionv3/>.

-
- 34 C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, *Papers with Code - Rethinking the Inception Architecture for Computer Vision*, 2017, <https://paperswithcode.com/paper/rethinking-the-inception-architecture-for>.
- 35 J. Brownlee, *Gentle Introduction to the Adam Optimization Algorithm for Deep Learning*, 2017, <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>.
- 36 TensorFlow, *tf.keras.applications.InceptionV3 | TensorFlow v2.16.1*, 2024, https://www.tensorflow.org/api_docs/python/tf/keras/applications/InceptionV3.
- 37 A. Rosebrock, *Convolutional Neural Networks (CNNs) and Layer Types*, 2021, <https://pyimagesearch.com/2021/05/14/convolutional-neural-networks-cnns-and-layer-types/>.
- 38 Prakharr0y, *Adam Optimizer | GeeksforGeeks Contributions*, 2024, https://www.geeksforgeeks.org/adam-optimizer/?itm_source=auth&itm_medium=contributions&itm_campaign=articles.