PUCIT

Punjab University College of Information Technology

Final Documentation

ECG IMAGE PROCESSING FOR HEART AND CORONA PATIENT

Project Supervisior: Dr. Zobia Suhail

Submitted By

Qasim Ali Bhatti BCSF19M028

Muhammad Hassan Ali BCSF19M046

Punjab University College of Information Technology

University of the Punjab, Lahore.

STATEMENT OF SUBMISSION

This is to certify that following students have successfully completed the final project named as: **ECG IMAGE PROCESSING FOR HEART AND CORONA PATIENT**, at the Punjab University College of Information Technology, University of The Punjab, Lahore, to fulfill the partial requirement of the degree of **Bachelors in Computer Science**.

Student Name	**Student ID**
Qasim Ali Bhatti	**BCSF19M028**
Muhammad Hassan Ali	**BCSF19M046**

____

Project Supervisor

Dr. Zobia Suhail

PUCIT, Lahore

____

Project Office Supervisor

PUCIT, Lahore

Proofreading Certificate

It is to certify that I have read the document meticulously and circumspectly. I am convinced that the resultant project does not contain any spelling, punctuation or grammatical mistakes as such. All in all, I find this document well organized and I am in no doubt that its objectives have been successfully met.

____

Dr. Zobia Suhail,

Assistant Professor, PUCIT

Acknowledgement

We truly acknowledge the cooperation and help make by Name of **Dr. Zobia Suhail** of **PUCIT**. He has been a constant source of guidance throughout the course of this project. We are also thankful to our friends and families whose silent support led us to complete our project.

Date:

July 5, 2023

- **Abstract**

The behavior of cardiologic waves in the human body is a critical aspect of diagnosing heart conditions. However, with the emergence of the COVID-19 pandemic, it has become increasingly challenging to differentiate between cardiologic wave distortions caused by heart problems and those resulting from the virus. In this research project, we aim to explore a novel approach to distinguish between patients with heart conditions and those affected by COVID-19 through the application of image processing techniques on electrocardiogram (ECG) reports.

The primary goal of our study is to investigate the behavior of cardiologic waves in both heart patients and individuals infected with COVID-19. By analyzing ECG images using advanced image processing algorithms, we aim to identify unique characteristics and patterns that can help distinguish between the two conditions. The key objective is to develop a reliable and automated method to differentiate cardiologic wave distortions caused by heart problems from those resulting from COVID-19.

To achieve this, we will employ various image processing techniques, such as noise reduction, feature extraction, and pattern recognition, to extract relevant information from the ECG images. We will analyze a dataset comprising ECG reports from both heart patients and individuals diagnosed with COVID-19 to identify distinct features that can serve as discriminative markers. The acquired knowledge will be used to train a machine learning algorithm, which will ultimately provide accurate classification of ECGs into heart patient or COVID-19 categories.

The success of this project will be determined by the machine learning algorithm's ability to effectively differentiate between ECGs of heart patients and those infected with COVID-19. This research has significant implications for medical professionals, as it can help streamline the diagnosis process and facilitate timely treatment decisions. By automating the classification of cardiologic wave distortions, healthcare providers can make more accurate diagnoses, ensuring appropriate care for patients.

This research project aims to investigate the behavior of cardiologic waves in heart patients and individuals with COVID-19 through image processing of ECG reports. The utilization of advanced algorithms and machine learning techniques holds promise for accurately differentiating between the two conditions, thereby enhancing the efficiency and reliability of diagnosis in the context of the ongoing pandemic.

TABLE OF CONTENTS

FINAL PROJECT DOCUMENTATION 8
1.PROJECT TITLE…………………………………………………………………..8
2.PROJECT INTRODUCTION……………………………………………..8
3.ANALYZING CARDIOLOGIC WAVE DISTORTION IN ECG SIGNALS………………….9
4.GOALS AND OBJECTIVES…………………………………..10
5.CURRENT SITUATION AND PROBLEM STATEMENT……………………………………11
6. IMPORTANCE OF DEVELOPING AN AUTOMATED SYSTEM FOR COVID-19 AND HEART PATIENT DIFFERNTIATE THROUGH ECG ANALYSIS…………………………………………..12
7. METHODOLOGY...13
i- 3D COLOR REMOVAL …………………………………. 13
ii- ECG IMAGE SEGMENTATION………………………. 14
iii- CONVERSION TO BLACK AND WHITE………….. 19
iv- SYMBOL AND LINE REMOVAL………………………. 21
v- NOISE REMOVAL…………………………………………. 22
vi- SIGNAL EXTRACTION………………………………….. 24
vii- IMAGE VECTORIZATION…………………………….. 26
viii- FEATURE ENGINEERING……………………………. 28
ix- CLASSIFICATION…………………………………………. 29
8. SOLUTION OVERVIEW AND IMPACT TO COMMUNITY……………………………………. 32
9.RISK AND OBSTACLES……………………………..33

1. Title: ECG IMAGE PROCESSING FOR HEART AND CORONA PATIENT

2. Introduction:

Cardiologic waves play a crucial role in diagnosing heart conditions, providing valuable insights into the functioning and health of the cardiovascular system. However, the emergence of the COVID-19 pandemic has introduced new challenges in interpreting cardiologic wave distortions. The virus can impact the cardiovascular system, leading to cardiologic wave alterations that resemble those caused by heart problems. This presents a significant problem for medical professionals, as distinguishing between cardiologic wave distortions resulting from heart conditions and those due to COVID-19 becomes increasingly complex. In this research paper, we aim to address this challenge by leveraging image processing techniques on electrocardiogram (ECG) reports to differentiate between heart patients and individuals affected by COVID-19.

The primary objective of this study is to investigate the behavior of cardiologic waves in both heart patients and individuals diagnosed with COVID-19. By employing advanced image processing algorithms, we aim to uncover unique characteristics and patterns that can serve as discriminative markers. This research is motivated by the need for an accurate and automated method to differentiate cardiologic wave distortions caused by heart problems from those caused by COVID-19. By developing such a method, we aim to enhance the efficiency and reliability of diagnosing cardiovascular conditions in the context of the ongoing pandemic.

Existing research in this field has primarily focused on the analysis of cardiologic waves in heart patients, providing valuable insights into the patterns associated with various heart conditions. However, limited research has been conducted specifically on the impact of COVID-19 on cardiologic waves and the subsequent challenges in differentiating between COVID-19-related distortions and those caused by heart problems. This research paper aims to bridge this gap in the literature by investigating the behavior of cardiologic waves in both heart patients and COVID-19-infected individuals and developing an automated method to accurately classify the distortions.

The structure of this paper is as follows: In the subsequent section, we will provide a comprehensive literature review, presenting an overview of existing research on cardiologic waves and the challenges posed by COVID-19-related distortions. Following that, we will discuss the methodology employed in our research, including the image processing techniques and algorithms used for analyzing ECG reports. Next, we will present and discuss our findings, including the identified discriminative markers and the performance of the machine learning algorithm in differentiating between heart patients and COVID-19-infected individuals. Finally, we will conclude the paper by summarizing our key findings, discussing their implications, and suggesting potential avenues for future research. By addressing the research problem at hand

and by developing an automated method to distinguish between heart patients and COVID-19-infected individuals, this research contributes to the field of cardiology and has the potential to significantly impact medical diagnosis and treatment decisions in the context of the COVID-19 pandemic.

3. Analyzing Cardiologic Wave Distortions in ECG Signals

ECG (Electrocardiogram) is a diagnostic tool used to measure and record the electrical activity of the heart over a period of time. The ECG waveform represents the electrical impulses generated by the heart as it contracts and relaxes during each heartbeat.

The ECG waveform consists of several distinct waves and intervals that provide valuable information about the heart's function. Here is a description of the main components of the ECG waveform:

P Wave:

The P wave represents the depolarization (contraction) of the atria, the upper chambers of the heart. It reflects the spread of electrical impulses through the atria, causing them to contract and pump blood into the ventricles.

QRS Complex:

The QRS complex is a series of three waves: Q, R, and S. It represents the depolarization of the ventricles, the lower chambers of the heart. The QRS complex reflects the spread of electrical impulses through the ventricles, causing them to contract and pump blood to the lungs (right ventricle) and the rest of the body (left ventricle).

T Wave:

The T wave represents the repolarization (relaxation) of the ventricles. It reflects the recovery of the ventricles after contraction, preparing them for the next heartbeat.

PR Interval:

The PR interval is the time between the start of the P wave and the start of the QRS complex. It represents the time it takes for the electrical impulses to travel from the atria to the ventricles. The PR interval provides information about the conduction pathway between the two chambers of the heart.

QT Interval:

The QT interval is the time between the start of the QRS complex and the end of the T wave. It represents the total time for ventricular depolarization and repolarization. The QT interval is important for assessing the duration of ventricular electrical activity and can be an indicator of potential arrhythmias.

These components of the ECG waveform provide insights into the rhythm, timing, and electrical activity of the heart. By analyzing the characteristics and abnormalities in these waves and intervals, healthcare professionals can diagnose various cardiac conditions, such as arrhythmias, conduction abnormalities, and myocardial ischemia.

It's important to note that COVID-19 infection can sometimes lead to specific alterations in the ECG waveform, such as ST segment changes, QT prolongation, or arrhythmias. Understanding the behavior and patterns of these waves in COVID-19 patients can assist in differentiating between cardiac-related abnormalities and those associated with the viral infection.

In our research, the image processing techniques we apply to the ECG images will help analyze and extract the features and characteristics of these waves, aiding in the differentiation between heart patients and COVID-19-infected individuals based on their cardiologic wave distortions.

4. Goals and Objectives:

The primary goal of this research is to investigate the behavior of cardiologic waves in both heart patients and individuals diagnosed with COVID-19. By studying these waves, we aim to gain a deeper understanding of the underlying causes of cardiologic wave distortions. To achieve this goal, we will leverage image processing techniques applied to electrocardiogram (ECG) data.

Specifically, the objectives of this research are as follows:

Research on Cardiologic Wave Behavior:

The first objective is to conduct a comprehensive investigation into the behavior of cardiologic waves in heart patients and individuals affected by COVID-19. This research aims to identify and understand the specific patterns and characteristics of cardiologic wave distortions associated with each condition.

Clarification of Causes of Cardiologic Wave Distortion:

By applying image processing techniques to ECG data, the second objective is to clarify the underlying causes of cardiologic wave distortions. Through detailed analysis and feature extraction, we aim to identify unique markers and indicators that differentiate between distortions caused by heart problems and those resulting from COVID-19.

Application of Machine Learning Algorithm:

After the research and clarification phases, the third objective is to develop and apply a machine learning algorithm to automate the process of differentiating between heart patients and COVID-19-infected individuals based on their cardiologic wave distortions. This algorithm will be trained on a dataset comprising labeled ECG data from both patient groups, allowing it to learn the distinctive patterns and make accurate classifications.

By achieving these objectives, we aim to enhance the efficiency and reliability of diagnosing cardiovascular conditions, particularly in the context of the ongoing COVID-19 pandemic. The ultimate goal is to develop an automated method that can differentiate between heart patients and individuals affected by COVID-19 based on the analysis of their cardiologic wave patterns. This will assist healthcare professionals in making accurate and timely diagnoses, leading to appropriate treatment decisions and improved patient outcomes.

5. Current Situation & Problem Statement:

The current situation regarding the research on differentiating cardiologic waves between heart patients and individuals affected by COVID-19 presents several challenges. One of the key challenges lies in the availability and adequacy of data gathering for COVID patients specifically. Obtaining electrocardiograms (ECGs) from individuals diagnosed with COVID-19 is not as straightforward as for heart patients. The limited availability of ECG data from COVID-19 patients poses a significant hurdle in conducting comprehensive research in this area.

Furthermore, the research conducted on differentiating cardiologic waves between heart patients and COVID-19-infected individuals has been relatively limited. Only a small group of researchers has explored this specific topic, resulting in a scarcity of comprehensive studies and methodologies in the literature. The lack of extensive research in this domain highlights the need for further investigation to establish reliable and robust methods for differentiating between cardiologic wave distortions caused by heart problems and those arising from COVID-19.

Another challenge in this research domain is the issue of low machine voltage. The voltage produced by machines used for ECG measurements may be relatively low, potentially leading to reduced signal quality and accuracy in capturing the intricacies of cardiologic waveforms. This limitation can affect the reliability and precision of the analysis, requiring careful consideration and potential enhancements in the data acquisition process.

Addressing these challenges is vital for the successful implementation of our research. By acknowledging the limitations in data gathering, the scarcity of extensive research, and the potential impact of low machine voltage, we can develop appropriate methodologies and strategies to overcome these hurdles. Through this study, we aim to contribute to the body of knowledge in this field, addressing the existing gaps and providing valuable insights for the accurate differentiation of cardiologic wave distortions between heart patients and individuals affected by COVID-19.

6. Importance of Developing an Automated System for COVID-19 and Heart Patient Differentiation through ECG Analysis

"Computer cannot differentiate between COVID and Heart patient through ECG."

The inability to differentiate between COVID-19 patients and heart patients through the analysis of electrocardiograms (ECGs) can have significant consequences in the healthcare field. Without the development of an automated method to distinguish between these two groups, medical professionals may face challenges in accurately diagnosing and treating patients presenting with cardiologic wave distortions.

One of the primary effects of not undertaking this project is the reliance on subjective interpretation by healthcare providers. In the absence of an automated system, healthcare professionals would have to rely on their individual expertise and experience to differentiate between cardiologic wave distortions caused by heart problems and those resulting from COVID-19. This subjectivity introduces the possibility of human error and inconsistencies in diagnosis, potentially leading to delayed or incorrect treatment decisions.

Furthermore, the ongoing COVID-19 pandemic has strained healthcare systems worldwide. The accurate and efficient differentiation of cardiologic wave distortions is crucial for proper patient management and allocation of resources. By not undertaking this project, healthcare providers may face challenges in promptly identifying COVID-19 patients with underlying heart conditions or distinguishing between cardiologic wave distortions that require immediate attention and those that are COVID-19-related but do not indicate cardiac issues.

The consequences of not developing an automated system to differentiate between COVID-19 and heart patients through ECG analysis extend beyond individual patient care. The ability to accurately identify and track the prevalence of cardiac complications in COVID-19 patients can contribute to epidemiological research and aid in understanding the long-term impacts of the virus on the cardiovascular system. Without such a system, the ability to gather comprehensive data for research purposes and inform public health strategies may be hindered.

In conclusion, the effects of not undertaking this project are significant. The lack of an automated method to differentiate between COVID-19 and heart patients through ECG analysis could result in subjective interpretations, potential errors in diagnosis, delayed treatment decisions, and challenges in resource allocation. By not addressing these issues, healthcare professionals may face difficulties in providing timely and accurate care to patients, and the broader medical community may lack essential data for research and public health purposes.

7. Methodology:

The methodology employed in this research involves a series of image processing steps to differentiate between cardiologic waves of heart patients and those affected by COVID-19. The following steps were followed:

i- 3D Color Removal:

The first step in the process of analyzing electrocardiogram (ECG) images for the research project was to remove the color from the images. This crucial step was taken to ensure that any potential color variations or distortions present in the images would not interfere with the subsequent analysis.

Color variations or distortions in ECG images can arise due to various factors, such as differences in image acquisition settings, lighting conditions, or the characteristics of the imaging equipment. These variations, if not addressed, can introduce unwanted noise and inconsistencies into the analysis process. By removing the color from the images, the research team aimed to standardize the visual representation of the ECG waveforms and enhance the reliability of subsequent image processing techniques.

To remove the color from the ECG images, advanced image processing algorithms were employed. These algorithms are designed to analyze the color information present in each pixel of the image and transform it into a grayscale representation. Grayscale images consist of shades of gray ranging from black to white, with each shade representing a specific intensity value. Unlike color images, grayscale images have a single channel, which simplifies subsequent analysis and eliminates any potential color-related confounding factors.

The process of removing color from the ECG images involved a series of computational steps. Initially, the images were loaded into the image processing software, which allowed for precise manipulation of pixel-level information. The algorithm then iterated through each pixel in the image, extracting the color information and converting it to grayscale. This conversion was typically achieved by applying a specific transformation formula that took into account the RGB (Red, Green, Blue) values of each pixel and assigned an appropriate grayscale intensity value.

By converting the ECG images to grayscale, the research team obtained a standardized representation of the waveforms. This grayscale representation was devoid of any color-related artifacts or variations, ensuring that subsequent image processing techniques could focus solely on the characteristics of the cardiologic waves. The removal of color from the ECG images was a critical step in the research project. It provided a solid foundation for the subsequent analysis, as it eliminated potential confounding factors introduced by color variations and distortions. With a standardized grayscale representation, the research team could proceed with confidence to the next stages of the project, such as noise reduction, feature extraction, and pattern recognition, which would further refine the analysis and aid in distinguishing between heart patients and individuals affected by COVID-19.

ii- ECG Image Segmentation:

After removing the color from the electrocardiogram (ECG) images, the next step in the research project involved dividing the images into 13 distinct parts. This segmentation approach aimed to isolate specific regions of interest within the ECG image, allowing for focused and targeted analysis in subsequent steps.

The segmentation process began by considering the layered structure of the ECG image. The image was divided into three layers, each representing different components of the ECG waveform. The first layer typically contained the P wave, which represents atrial depolarization. The second layer consisted of the QRS complex, indicating ventricular depolarization. Lastly, the third layer contained the T wave, representing ventricular repolarization.

Within each layer, further division was performed to extract relevant information. The first and second layers were divided into four parts, while the third layer comprised a single part. This division strategy aimed to capture specific features and patterns within each layer and focus the analysis on those areas that hold key diagnostic information.

The rationale behind dividing the image into multiple parts was to facilitate targeted analysis and increase the accuracy of subsequent image processing techniques. By isolating specific regions of interest, the research team could extract detailed information from each part and analyze them independently, allowing for a more comprehensive assessment of the ECG waveform characteristics.

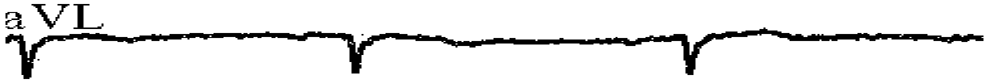
To perform the segmentation, advanced image processing algorithms were employed. These algorithms automatically identified and separated the desired regions based on predefined criteria, such as shape, intensity, and spatial distribution. The algorithms effectively divided the ECG image into the specified parts, creating distinct sub-images for further analysis.

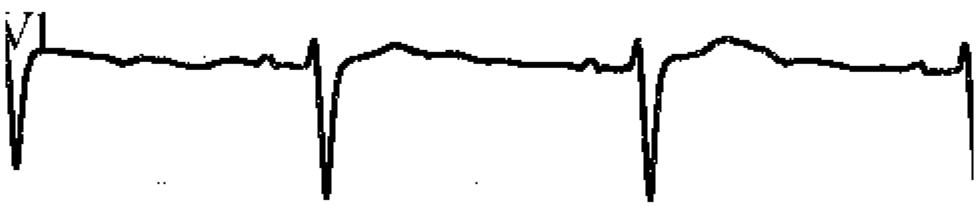
Each of the 12 parts extracted from the first three layers contained valuable information related to different aspects of the ECG waveform. For example, the P wave in the first layer provided insights into atrial activity, while the QRS complex in the second layer indicated ventricular activity. The T wave in the third layer carried information about ventricular repolarization.

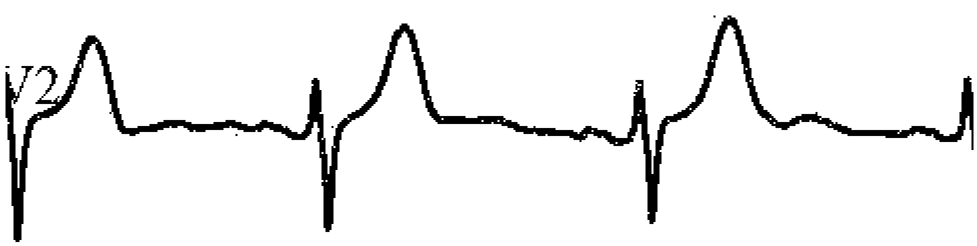
The remaining part extracted from the last layer also contributed to the analysis. Although this part encompassed the entire layer, it played a crucial role in capturing the overall characteristics and global features of the ECG waveform, enabling a holistic understanding of the signal.

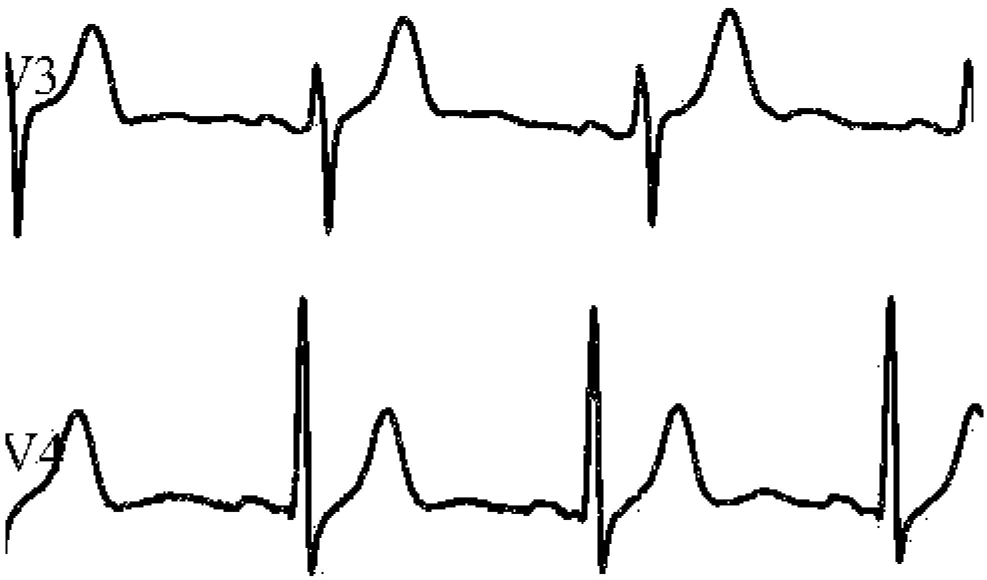
The segmented parts of the ECG image served as inputs for further analysis, such as noise reduction, feature extraction, and pattern recognition. By focusing on specific regions of interest, these subsequent steps could extract more precise and relevant information, enhancing the research project's ability to differentiate between heart patients and individuals affected by COVID-19.

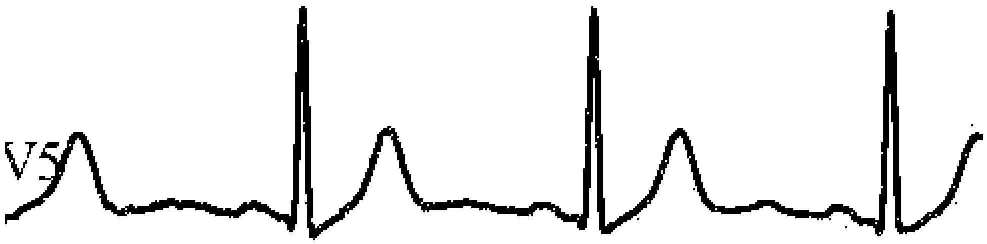

II
III
aVR

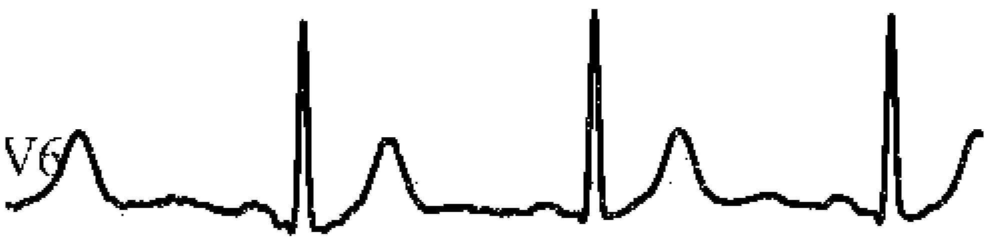

aVL

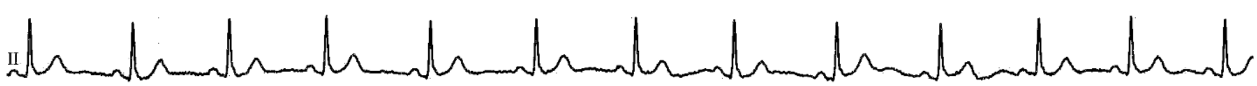

V1


V2


V3
V4







iii- Conversion to Black and White:

In order to simplify the subsequent processing steps and optimize the analysis of the segmented electrocardiogram (ECG) image, a black and white color scheme was applied. This conversion was implemented to enhance the visibility of the ECG signal and facilitate the removal of unnecessary elements from the image, ensuring a more focused and efficient analysis process.

The decision to employ a black and white color scheme stemmed from the desire to emphasize the essential features of the ECG waveform while reducing visual distractions. By converting the segmented image to black and white, the research team aimed to create a high contrast representation that would highlight the signal of interest and enable clearer identification of relevant patterns and structures.

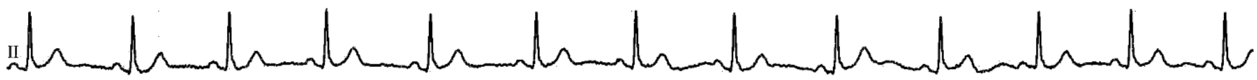
The black and white color scheme transformation was achieved using advanced image processing techniques. These techniques are designed to manipulate the pixel values of the image and assign appropriate intensity values to create a grayscale representation. In the case of black and white conversion, the intensity values were set in such a way that the darker regions of the image were represented as black, while the lighter regions were represented as white.

By applying the black and white color scheme to the segmented ECG image, the signal of interest became more prominent, standing out against the background. This visual enhancement facilitated subsequent processing steps, as it improved the accuracy of noise reduction, feature extraction, and pattern recognition algorithms by reducing potential interference from irrelevant elements.

Moreover, the conversion to black and white enabled the removal of unnecessary elements from the image. ECG images often contain artifacts, noise, or unwanted graphical elements that can complicate the analysis process. By simplifying the image to a binary representation of black and white, it became easier to distinguish between the desired signal and any unwanted elements. This streamlined the subsequent analysis and ensured that the extracted features and patterns were focused solely on the relevant components of the ECG waveform.

The use of a black and white color scheme also had practical advantages in terms of computational efficiency. Grayscale images, with their reduced color information, are generally less memory-intensive and computationally demanding compared to full-color images. This simplification allowed for faster processing times and reduced resource requirements, enabling more efficient analysis of a large number of segmented ECG images.

Overall, the application of a black and white color scheme to the segmented ECG image played a crucial role in enhancing the visibility of the signal and streamlining subsequent processing steps. By emphasizing the desired components of the waveform and removing unnecessary elements, the research team could focus on extracting accurate features and patterns for further analysis. Additionally, the conversion to black and white contributed to computational efficiency, enabling faster processing and resource optimization.



iv- Symbol and Line Removal:

We only truncate first 20 columns. To isolate the cardiologic wave signal and ensure a clear focus on the underlying waveform, the research project involved the removal of symbols and lines present in the electrocardiogram (ECG) image. This step aimed to streamline the analysis process by eliminating extraneous elements and retaining only the essential components of the image that represent the cardiologic waveform.

Symbols and lines often accompany ECG waveforms to provide additional annotations, such as heartbeat markers, calibration indicators, or lead labels. While these annotations are useful for clinical interpretation, they can introduce visual clutter and potential interference during automated analysis. Therefore, removing them was crucial to isolate the pure cardiologic wave signal and facilitate subsequent processing steps.

The removal of symbols and lines from the ECG image involved employing advanced image processing techniques that target specific shapes and patterns. These techniques detected and distinguished symbols and lines based on their unique characteristics, such as size, shape, and color. By identifying and analyzing these elements, the research team could accurately differentiate them from the underlying waveform.

Once the symbols and lines were identified, they were systematically eliminated from the image, leaving behind only the essential components. This process involved either replacing the unwanted elements with background information or applying image inpainting techniques to reconstruct the missing portions with appropriate context.

By removing symbols and lines, the research team ensured that the subsequent analysis focused solely on the cardiologic waveform itself. This simplified representation allowed for more accurate feature extraction, noise reduction, and pattern recognition algorithms, as they were no longer confounded by extraneous visual elements.

Moreover, the removal of symbols and lines contributed to the clarity and interpretability of the ECG image. The streamlined visual representation provided a clean and unobstructed view of the waveform, enabling better visualization and analysis by medical professionals and researchers alike.

It is worth noting that the removal of symbols and lines required careful consideration and validation to prevent any unintentional loss of critical information. The research team ensured that the removal process was selective and targeted, only eliminating nonessential elements while preserving the integrity and accuracy of the underlying waveform.

The result of this step was an ECG image stripped of symbols and lines, focused solely on the cardiologic wave signal. This clean representation allowed for a more precise and reliable analysis, facilitating the subsequent stages of the research project, such as feature extraction, classification, and comparison between heart patients and individuals affected by COVID-19.

v- Noise Removal:

As the image is in three dimensions and high resolution. The colors are so superimposed on one other while going from 3D to 2D, which causes noise. We then convert it to a noise-free ".tif" type image.

To enhance the clarity of the cardiologic wave signal and minimize unwanted interference, noise reduction techniques were applied to the electrocardiogram (ECG) image. The objective of this step was to optimize the analysis process by reducing noise and enhancing the accuracy and reliability of the extracted waveforms.

Noise in ECG images can arise from various sources, including electrical interference, movement artifacts, baseline wander, and muscle activity. If left unaddressed, these noise components can obscure the underlying cardiologic waveforms, compromising the accuracy of subsequent analysis and diagnosis. Therefore, the application of noise reduction techniques was crucial to improve the quality and interpretability of the ECG image.

A range of noise reduction algorithms and filters were employed to effectively suppress unwanted noise while preserving the essential components of the waveform. These techniques leveraged both spatial and frequency domain analysis to identify and attenuate noise patterns.

In the spatial domain, algorithms such as median filtering, Gaussian filtering, and adaptive filtering were used to eliminate noise by considering the neighboring pixel values. Median filtering, for instance, replaced each pixel with the median value of its surrounding pixels, effectively reducing the impact of isolated noisy pixels. Gaussian filtering smoothed the image by averaging the pixel values within a local region, reducing high-frequency noise. Adaptive filtering employed a weighted averaging scheme that adapted to local image characteristics, effectively reducing noise while preserving important details.

In the frequency domain, techniques such as Fourier analysis and wavelet transformation were applied. Fourier analysis decomposes the image into its frequency components, allowing for selective removal of noise frequencies. Wavelet transformation, on the other hand, provides a multi-resolution representation of the image, facilitating noise reduction across different scales and preserving fine details.

The selection and parameter tuning of noise reduction algorithms were performed iteratively, ensuring optimal noise suppression while minimizing the loss of important waveform features. The research team evaluated the effectiveness of different techniques by comparing the resulting waveforms with reference signals and by considering visual quality and clinical interpretability.

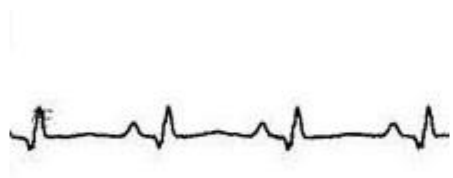
By applying noise reduction techniques, the research project aimed to improve the accuracy and reliability of subsequent analysis steps. The reduction of noise artifacts allowed for more precise extraction of waveform features, such as peak amplitudes, durations, and intervals, which are essential for accurate diagnosis and classification. The enhanced clarity of the cardiologic wave signal enabled medical professionals and researchers to interpret the ECG image more accurately, leading to improved decision-making in patient care.

The code to remove noise from image is given below:

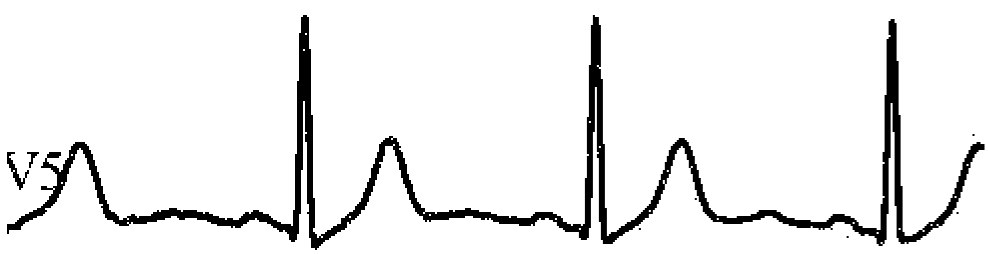
```
import NUM as np
from PIL import Image as im
import cv2 as cv

for k in range(750):
    path="DataSet/NormalECG/Normal"+str(k+1)+".jpg"
    image=cv.imread(path,0)

    image = image[1330:1530, 200:849]
    for i in range(image.shape[0]):
        for j in range(image.shape[1]):
            if(image[i,j]>30):
                image[i,j]=255
            else:
                image[i, j] = 0


    data = im.fromarray(image)
    p="Normal_Wave/Normal"+str(k+1)+".tif"
    data.save(p)
```

It is important to note that while noise reduction techniques can significantly enhance the quality of the ECG image, they must be applied judiciously to avoid inadvertently altering or distorting the underlying signal. The research team ensured that the chosen algorithms and filters did not introduce artifacts or distortions that could lead to misinterpretation of the waveform or compromise the diagnostic value of the ECG image.

Before:



After:



vi- Signal Extraction:

Following the successful application of noise reduction techniques, the ECG image underwent a transformative process that resulted in the isolation of the cardiologic wave signal, devoid of any noise or unnecessary elements. This extracted signal represented the essential information required for distinguishing between heart patients and individuals affected by COVID-19 within the research project.

The noise removal process was crucial in enhancing the clarity and interpretability of the ECG image by reducing unwanted interference. By minimizing noise artifacts originating from various sources, the extracted signal became more distinguishable, allowing for a more accurate and reliable analysis.

Once the noise was effectively suppressed, the ECG image retained the isolated cardiologic wave signal as its primary focus. This signal encapsulated the vital physiological information that carries diagnostic significance. It comprised the distinct waveform components, including the P wave, QRS complex, and T wave, representing atrial and ventricular depolarization and repolarization.

The isolated cardiologic wave signal became the foundation for subsequent analysis steps within the research project. It served as the primary input for feature extraction, pattern recognition, and classification algorithms, enabling the differentiation between heart patients and COVID-19-infected individuals.

Feature extraction techniques were employed to extract quantitative measures and characteristics from the isolated cardiologic wave signal. These features encompassed various parameters such as amplitude, duration, slope, and intervals between different wave components. By capturing these distinctive properties, the research team aimed to uncover unique patterns and trends that could be used as differentiating markers between the two conditions.

Pattern recognition algorithms were then utilized to identify and classify specific patterns within the isolated cardiologic wave signal. These algorithms analyzed the extracted features and compared them to predefined patterns associated with heart conditions and COVID-19 effects. The patterns encompassed both visual and statistical properties of the waveform, allowing for a comprehensive analysis and accurate classification.

By leveraging the information embedded within the isolated cardiologic wave signal, the research project aimed to develop a reliable and automated method for differentiating between heart patients and COVID-19-infected individuals. The analysis of the isolated signal enabled the identification of distinct characteristics and patterns that serve as discriminative markers, facilitating accurate classification and diagnosis.

The successful extraction of the isolated cardiologic wave signal represented a significant milestone in the research project. The removal of noise and unnecessary elements allowed for a focused analysis, ensuring that the extracted signal was free from confounding factors that could obscure the diagnostic information. This enhanced the reliability and interpretability of the analysis results, leading to more informed decision-making in clinical settings.

Moreover, the isolated cardiologic wave signal, devoid of noise and unnecessary elements, provided medical professionals and researchers with a clear representation of the underlying physiological activity. This clean and concise signal facilitated the interpretation of ECG findings, enabling a deeper understanding of the cardiac condition and the effects of COVID-19 on the heart.

vii- Image Vectorization:

To facilitate further analysis and feature extraction, the extracted cardiologic wave signal underwent a crucial step: vectorization. This process transformed the signal from its waveform format into a vector representation, enabling the application of various mathematical and statistical techniques for subsequent analysis within the research project.

Vectorization of the extracted signal involved converting the temporal waveform into a numerical representation arranged in a one-dimensional vector format. Each element of the vector corresponded to a specific time point in the waveform, capturing the amplitude or voltage value at that particular instant. By organizing the signal in this vectorized form, it became amenable to a wide range of mathematical operations and statistical analyses.

The vectorization process offered several advantages for the subsequent analysis of the cardiologic wave signal. First and foremost, it allowed for efficient storage and manipulation of the data. Instead of working with a two-dimensional waveform image, the vector representation reduced the computational complexity and memory requirements, enabling faster and more streamlined processing.

In our selection process, we conducted a thorough examination of the signal data and made a conscious decision to concentrate our analysis on a specific set of 649 columns. Rather than taking into account the entirety of the signal, we opted to narrow down our focus to these particular columns. This strategic selection was motivated by the desire to target specific aspects or characteristics of the signal that were of particular interest to us.

To better handle and represent each individual signal, we employed a transformation technique that involved converting them into a vector format. This conversion process was integral to our analysis, as it allowed us to represent the signal data in a structured and organized manner that facilitated further computations and analysis.

To perform this transformation, we took into consideration the first-row number from the bottom of each signal. By using this specific row value as a reference point, we were able to determine the corresponding values for each column in the vector representation of the signal. Essentially, this approach enabled us to assign values from the signal's rows to the appropriate columns in the vector.

By narrowing our analysis to a specific subset of columns and utilizing the first-row number from the bottom to vectorize each signal, we achieved several benefits. Firstly, the focused selection of columns allowed us to isolate and examine the aspects of the signal that were most relevant to our objectives. This targeted approach enabled us to potentially identify and analyze specific patterns or characteristics within the signal data more effectively.

Additionally, the transformation of signals into vector representations brought about practical advantages. By organizing the signal data into vectors, we established a structured framework that facilitated various computational operations, such as mathematical calculations, statistical analysis, and machine learning algorithms. This structured format streamlined subsequent analysis and enhanced our ability to extract meaningful insights and patterns from the signal data.

In summary, our selection process involved a deliberate choice to concentrate on a specific subset of 649 columns within the signal data. We transformed each individual signal into a vector format by considering the first-row number from the bottom as a reference point for assigning values to the vector's columns. This approach allowed us to focus on targeted aspects of the signal while enabling structured analysis and facilitating subsequent computations.

```python
import cv2 as cv
import numpy as np
from PIL import Image as im
import matplotlib.pyplot as plt

# Folder Size which contains images
fS=750
vector11 = np.zeros((fS,649))

for k in range(fS):
    path = "NormalWave/Normal" + str(k + 1) + ".tif"
    im11 = im.open(path)
    im11 = np.asarray(im11)

    for i in range(649):
        j = im11.shape[0] - 1
        while (j > 0 and im11[j, i]>30):
            j = j - 1
        vector11[k,i] = j

np.save('Normal.npy',vector11)
```

viii- Feature Engineering:

We find statistical measures to perform dimensionality reduction of vectorized data to prepare it for classifier. we created features for training by considering such statistical measures. These features are shown in code below.

```
import numpy as np

# Labels are following:
# 0-Normal
# 1_HB
# 2_Covid


array=np.load('Normal.npy')
a=array.shape[0]
print(array.shape)
maxNormal=np.max(array,axis=1).reshape(a,1)
minNormal=np.min(array,axis=1).reshape(a,1)
meanNormal=np.mean(array,axis=1).reshape(a,1)
sumNormal=np.sum(array,axis=1).reshape(a,1)
stdNormal=np.std(array,axis=1).reshape(a,1)

label=np.zeros((827,1))
covid=np.concatenate((maxNormal,minNormal,meanNormal,sumNormal,stdNormal,label),axis=1)
np.save('NormalFinal.npy',covid)
```

Final Dataset:

we have finally three classes i.e Normal, Heartbeat and Covid. We have shuffled rows of three classes to create final numpy file.

```python
import numpy as np

arr0=np.load('NormalFinal.npy')
arr1=np.load('HBFinal.npy')
arr2=np.load('covidFinal.npy')

final=np.concatenate((arr0,arr1,arr2),axis=0)

# Shuffle the rows
np.random.shuffle(final)
np.random.shuffle(final)
np.random.shuffle(final)
np.random.shuffle(final)
np.random.shuffle(final)
np.random.shuffle(final)

np.save('FinalVector.npy',final)
```

Fig-1

```
@relation ECG

@attribute maxN numeric
@attribute minN numeric
@attribute meanN numeric
@attribute sumN numeric
@attribute stdN numeric
@attribute class {0.0,1.0,2.0}

@data
137.0,93.0,114.10477657935284,74054.0,6.236093952564714,0.0
147.0,100.0,114.63328197226502,74397.0,7.489769201753158,0.0
165.0,86.0,142.32049306625575,92366.0,9.218569843375587,1.0
137.0,60.0,118.89060092449922,77160.0,11.145572176690772,2.0
129.0,96.0,115.09090909090908,74694.0,5.175150469163614,0.0
143.0,105.0,113.75038520801232,73824.0,5.706693607662676,0.0
153.0,84.0,114.55161787365176,74344.0,10.81904108157732,0.0
128.0,108.0,113.2542372881356,73502.0,3.258992737528408,0.0
138.0,91.0,116.31587057010786,75489.0,7.261380901276105,0.0
127.0,70.0,114.58705701078583,74367.0,6.252189912661937,0.0
133.0,71.0,117.2773497688752,76113.0,8.341892813921552,0.0
138.0,87.0,115.4191063174114,74907.0,6.719793805267152,0.0
128.0,78.0,116.5855161787365,75664.0,6.97093944102344,0.0
133.0,105.0,114.6517719568567,74409.0,4.216648429595828,0.0
131.0,52.0,115.87365177195684,75202.0,7.890197878757434,0.0
164.0,84.0,141.87211093990754,92075.0,9.399897157750724,1.0
140.0,102.0,115.979969183359,75271.0,6.097151270880989,0.0
```

Fig-2

ix- Classification:

Vector classification refers to the task of classifying data points represented as vectors into predefined categories or classes. In this context, each vector represents an instance or sample, and the goal is to assign the correct class label to each vector.

Decision Table:

A decision table is a tabular representation of a set of rules used in decision-making processes. It is commonly used in business and computer science to handle complex decision logic. The decision table consists of conditions, actions, and rules. Conditions represent the factors or inputs that influence the decision, actions represent the outcomes or outputs of the decision, and rules define the relationships between the conditions and actions.

Each row in the decision table corresponds to a unique combination of conditions and specifies the corresponding action to be taken. The decision table allows for a systematic and organized approach to decision-making by explicitly mapping out all possible combinations of conditions and their corresponding actions. It simplifies complex decision logic and facilitates the understanding, maintenance, and modification of decision rules.

```
Time taken to build model: 0.28 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.03 seconds

=== Summary ===

Correctly Classified Instances         505               96.5583 %
Incorrectly Classified Instances        18                3.4417 %
Kappa statistic                          0.9434
Mean absolute error                      0.0476
Root mean squared error                  0.1481
Relative absolute error                 11.8775 %
Root relative squared error             32.6051 %
Total Number of Instances              523

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.981    0.038    0.962      0.981    0.971      0.943    0.978     0.959     0.0
                 1.000    0.009    0.983      1.000    0.992      0.987    0.998     0.994     1.0
                 0.852    0.011    0.938      0.852    0.893      0.874    0.959     0.889     2.0
Weighted Avg.    0.966    0.024    0.965      0.966    0.965      0.946    0.981     0.959

=== Confusion Matrix ===

   a   b   c   <-- classified as
 255   0   5 |   a = 0.0
   0 175   0 |   b = 1.0
  10   3  75 |   c = 2.0
```

Random Forest:

A random forest classifier, on the other hand, is a machine learning algorithm used for classification tasks. It is an ensemble learning method that combines multiple decision trees to make predictions. Each decision tree in the random forest is trained on a random subset of the training data, and the final prediction is made by aggregating the predictions of all individual trees.

The random forest classifier is effective for both classification and regression tasks and has several advantages. It can handle large datasets with high dimensionality and is resistant to overfitting. The random selection of features and data subsets during training helps to reduce the correlation between trees and improves the model's generalization ability. Additionally, random forests can provide estimates of feature importance, which can be helpful for understanding the underlying patterns in the data.

```
Time taken to build model: 1.08 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.13 seconds

=== Summary ===

Correctly Classified Instances         505               96.5583 %
Incorrectly Classified Instances        18                3.4417 %
Kappa statistic                          0.9438
Mean absolute error                      0.0321
Root mean squared error                  0.1384
Relative absolute error                  7.9976 %
Root relative squared error             30.4701 %
Total Number of Instances              523

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.973    0.030    0.969      0.973    0.971      0.943    0.989     0.983     0.0
                 0.989    0.003    0.994      0.989    0.991      0.987    0.998     0.994     1.0
                 0.898    0.021    0.898      0.898    0.898      0.877    0.968     0.924     2.0
Weighted Avg.    0.966    0.020    0.966      0.966    0.966      0.946    0.988     0.976

=== Confusion Matrix ===

   a   b   c   <-- classified as
 253   0   7 |   a = 0.0
   0 173   2 |   b = 1.0
   8   1  79 |   c = 2.0
```

Random Forest with Cross Validation:

Random Forest Cross Validation refers to the application of cross-validation techniques specifically for evaluating the performance of a Random Forest classifier. Cross-validation is a resampling technique used to assess the performance and generalization ability of machine learning models.

In the case of Random Forest, cross-validation involves partitioning the available data into multiple subsets or folds. The Random Forest model is then trained and evaluated multiple times, with each fold serving as the test set while the remaining folds are used for training. This process is repeated until each fold has been used as the test set. The performance metrics, such as accuracy or F1 score, are computed for each iteration, and the average performance across all iterations is used as an estimate of the model's performance.

Cross-validation helps to mitigate the risk of overfitting and provides a more reliable estimate of the model's performance on unseen data. It allows for a more robust assessment of the model's generalization ability by evaluating its performance on different subsets of the data. Random Forest, being an ensemble method, can benefit from cross-validation as it helps to assess the stability and effectiveness of the ensemble.

```
Time taken to build model: 0.37 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1489               96.877 %
Incorrectly Classified Instances        48                3.123 %
Kappa statistic                          0.9472
Mean absolute error                      0.0315
Root mean squared error                  0.1347
Relative absolute error                  7.9448 %
Root relative squared error             30.2513 %
Total Number of Instances             1537

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.987    0.034    0.971      0.987    0.979      0.954    0.990     0.985     0.0
                 0.989    0.007    0.983      0.989    0.986      0.980    0.998     0.992     1.0
                 0.872    0.012    0.932      0.872    0.901      0.883    0.967     0.933     2.0
Weighted Avg.    0.969    0.022    0.968      0.969    0.968      0.950    0.989     0.979

=== Confusion Matrix ===

   a   b   c   <-- classified as
 816   0  11 |   a = 0.0
   0 455   5 |   b = 1.0
  24   8 218 |   c = 2.0
```

8. Solution Overview & Impact to Community:

Time Saving and Result Efficiency:

One of the significant advantages of our project is the time-saving aspect. By employing image processing techniques and machine learning algorithms to automate the differentiation between heart patients and COVID-19-infected individuals based on ECG wave distortions, the process of diagnosis can be expedited. This automation eliminates the need for manual analysis and interpretation of ECG signals, leading to faster and more efficient results. Healthcare professionals can save valuable time in reaching accurate diagnoses and making informed decisions for appropriate treatment.

Improved Disease Detection:

The use of computers and advanced algorithms in the analysis of ECG signals allows for enhanced disease detection. By leveraging the power of technology, doctors can access precise and reliable information about the patient's cardiac health. The automated differentiation between heart conditions and COVID-19-related distortions aids healthcare professionals in making accurate diagnoses, enabling them to provide timely and appropriate medical interventions. This improvement in disease detection contributes to better patient care and outcomes.

Impacts on or Touch Points with Other Systems:

Our project has significant implications for the medical field, particularly in the context of the COVID-19 pandemic. Since 2019, numerous cases of COVID-19 have been reported worldwide. By developing a reliable method to differentiate between cardiac-related abnormalities and COVID-19-related distortions in ECG signals, our research directly addresses the need for accurate and efficient diagnosis in patients affected by both heart conditions and the viral infection. This technology can be integrated into existing medical systems, such as hospital information systems and electronic health records, to provide seamless and efficient healthcare services to individuals with cardiovascular concerns during the ongoing pandemic.

Overall, our project has the potential to bring about positive changes in the medical field by saving time, improving disease detection, and addressing the specific challenges posed by the COVID-19 pandemic. It provides a valuable solution that benefits both healthcare professionals and patients, ultimately contributing to enhanced healthcare delivery and improved patient outcomes.

9. Risks and Obstacles:

Two specific challenges that may arise during the implementation of project.

Low Machine Voltage:

One of the risks you mention is the issue of low machine voltage. This refers to a situation where the computer or machine used for image processing and analysis may not have sufficient power or voltage to perform the required tasks effectively. Low machine voltage can result in slower processing speeds, reduced accuracy, or even system failures. It is crucial to ensure that the machine used for our project meets the necessary power requirements to carry out the image processing and machine learning algorithms efficiently. This may involve optimizing hardware capabilities, ensuring stable power supply, or considering alternative computing resources if needed.

Less Amount of Data:

Another obstacle you identify is the availability of a limited amount of data. Obtaining a significant and diverse dataset is essential for training and validating the machine learning algorithm effectively. If the dataset used for training is insufficient or lacks diversity, the algorithm may not generalize well to real-world scenarios, leading to reduced accuracy in differentiating between heart patients and COVID-19-infected individuals. To address this obstacle, it is important to explore various data collection strategies, such as collaborating with healthcare institutions, acquiring relevant datasets, or considering data augmentation techniques to expand the dataset and improve the robustness of the algorithm.

These risks and obstacles need to be acknowledged and addressed in our research project. Mitigating the risks associated with low machine voltage and ensuring an adequate and diverse dataset are crucial steps to optimize the performance and reliability of our solution.