From Pixels to Clinical Insight: Computer Vision and Deep Learning for Automated Left Ventricular Segmentation and Ejection Fraction Prediction in Pediatric Echocardiography Videos

By

KONSTANTINOS KOSTIRIS

A thesis submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

Data Science

DEREE - The American College of Greece

2023

THESIS APPROVAL

"From Pixels to Clinical Insight: Computer Vision and Deep Learning for Automated Left Ventricular Segmentation and Ejection Fraction Prediction in Pediatric Echocardiography Videos" a thesis prepared by Konstantinos Kostiris in partial fulfillment of the requirements for the Master of Science (MS) in Data Science was presented December 12, 2023, and was approved and accepted by the thesis advisor, internal examiner and the School of Graduate and Professional Education.

APPROVALS: _____

Dr. George Drakakis, Thesis Advisor

_____

Dr. Dimitrios Vogiatzis, Committee Member

APPROVED BY: _____

Dr. Areti Krepapa

Dean, School of Graduate and Professional Education

Abstract

This study aims to explore the fusion of medical imaging and advanced machine learning in echocardiography, an essential diagnostic tool for heart assessments, especially in the pediatric domain. While echocardiography is widely used, challenges specific to the pediatric population emphasize the need for innovation. Beginning with investigating the heart’s complex structures, the journey of this research dives into reviewing the existing literature on pediatric echocardiography, addressing its advantages and limitations.

Central to the methodology is the first extensive pediatric echocardiographic video dataset available as of early 2023 from Stanford's Center for Artificial Intelligence in Medicine & Imaging, named EchoNet-Pediatric. Expert-annotated echocardiograms lay the foundation for computer vision algorithms and deep learning models on which they are trained and validated.

The results, based on key performance metrics, demonstrate the capabilities of integrating Artificial Intelligence in pediatric echocardiography, showcasing the synergy between traditional medical practices and cutting-edge advancements.

AKNOWLEDGMENTS

I would like to express my gratitude to my thesis advisor, Dr. George Drakakis. I am incredibly grateful for his guidance, patience and invaluable insights throughout our countless meetings that have been instrumental in shaping this thesis. His mentorship has left a profound impact on my academic, personal, and professional growth.

A special thanks to the American College of Greece, the faculty and my fellow students who gifted me with an unforgettable journey into the world of Data Science. Without this experience, my horizons wouldn't have expanded. I sincerely thank them and will always cherish these wonderful memories.

To my relatives and friends, whom I might have unintentionally neglected during my graduate studies. Their understanding and patience have meant the world to me. I truly value our relationship and am committed to making up for the lost time.

I wish to extend my gratitude to Dr. Stamatis Adamopoulos, cardiologist, who was by my side during challenging times as I battled myocarditis and introduced me to the heart complexities. Without him this thesis could not have been possible.

From the bottom of my heart, I want to express my gratitude to my parents and brother. Their unwavering support, guidance, and the foundational values they taught me have been my guiding torch. I am eternally indebted for their unconditional love and everything they provided.

Lastly, I would like to dedicate this thesis to my nieces Eriphyll, Emelleia, and my goddaughter Lydia.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

The ever-evolving world of cardiac diagnostics has experienced significant changes over the recent years due to the latest advancements in medical technology and the evolution of computational methods in Computer and Data Sciences. This introduction is aiming to pave the way for this study of exploring computer vision and deep learning techniques in pediatric echocardiography by addressing the problem statement, emphasizing the significance of echocardiography, acknowledging the limitations of existing approaches, outlining the research goals and hypotheses, justifying the importance of this study, and providing an overview of how this thesis is structured.

How the heart works

Heart Anatomy

The human heart is an extraordinary machine. It beats about 100,000 times a day and the heart's muscular walls pump about 5.6 liters of blood into blood vessels branching throughout the human body three times a minute.

The heart is divided into two sides, the left and the right, each of them consisting of two chambers. On each side the upper part is called atrium while the bottom ventricle. Four valves coordinate the direction of the blood flow, the atrioventricular valves that divide the atria from the ventricles, and the semilunar valves that,

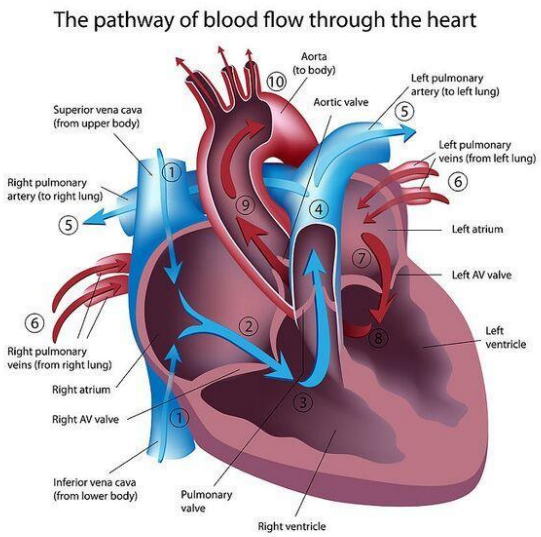


Figure 1: Human heart anatomy and blood flow
© alilamedicalimages.org

when open, allow the blood flow to the two most important human arteries, the aorta and the pulmonary artery.

Non-oxygenated blood arrives to the right atrium via the superior and inferior vena cavae from the circulatory system, while oxygen-rich blood arrives to the left atrium via the pulmonary veins from the lungs. From the right atria the blood flows to the right ventricle via the tricuspid valve, while from the left atria the blood flows to the left ventricle via the mitral valve. The right ventricle pumps the deoxygenated blood to the lungs via the pulmonary valve to the pulmonary artery in order to be oxygenated, while the left ventricle pumps the oxygenated blood via the aortic valve to the aorta in order to reach the body. **(University of Michigan, 2019)**

Cardiac Cycle

The cardiac cycle is a sequence of perfectly orchestrated movements that take place in the normal heart during each heartbeat. There are two types of movements of cardiac chambers: systole and diastole. Systole means contraction and diastole means relaxation. Atria and ventricles move independently; however, movements of the ventricles are clinically more important since it is the ventricles that ultimately pump the blood into the circulation system. Atria act as assistants of ventricles when it comes to pumping, thus, even if not mentioned explicitly, the movements refer to the movements of ventricles. Thus, systole means ventricular systole and diastole means ventricular diastole. **(University of California San Diego, 2017)**

There are four phases of the cardiac cycle.

1. **Ventricular Diastole (Inflow):** During this, the inlet atrioventricular valves are open, and outlet semilunar valves are closed. Inflow follows a relaxation phase, which results in the opening of the atrioventricular valves. Then the blood rushes from the atria into the ventricles due to the lower pressure in the ventricles.
2. **Isovolumetric contraction:** By the end of the ventricular filling, an electrical impulse originated from the sinus node, the natural pacemaker of the heart, starts depolarizing the ventricular myocardium which causes the contraction of the ventricles. This results in a rise of the ventricular pressure, which causes the closing of the atrioventricular valves. During this phase, all the valves are closed and the ventricles are contracting, but there is no flow of blood for a very short period of time. This is the *End Diastole*.

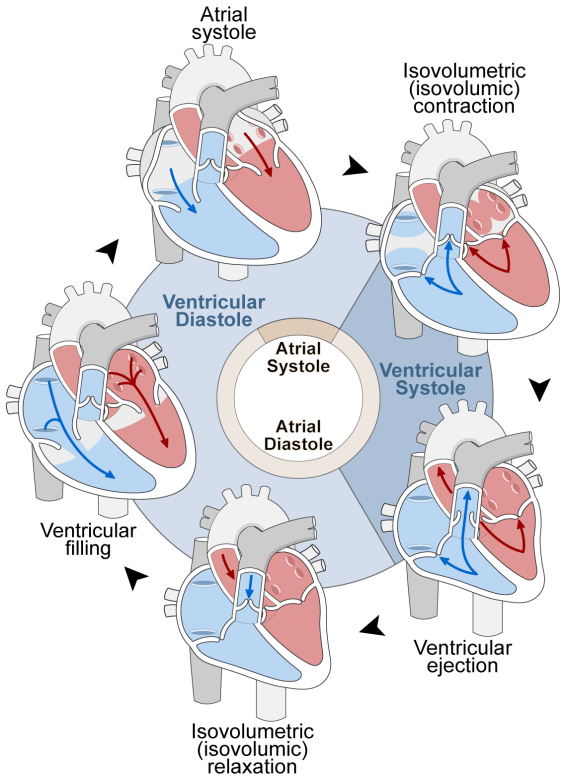


Figure 2: Cardiac Cycle Simulation © humanbiomedia.org

3. **Ventricular ejection (Outflow):** Eventually, the pressure increases above that in the major arteries, causing the semilunar valves to open. During this phase the atria start relaxing and collecting blood. The elevated pressure in the ventricles causes a rapid ejection of blood via the major arteries where oxygenated blood travels to the body via the Aorta and deoxygenated blood travels to the lungs via the pulmonary artery, followed by a slowing ejection.
4. **Isovolumetric relaxation:** Due to the slow ejection at the end of this phase, the ventricles start to dilate. Eventually the blood tends to flow back which causes the closure of the semilunar valves. After this, the ventricles relax with closed chambers for a brief period of time. This is the *End Systole*. The pressure in the ventricles falls below that in the atria and atrioventricular valves open causing the beginning of the next cycle.

Ejection Fraction

The Ejection Fraction compares the amount of blood in the heart to the amount of blood pumped out. It is a percentile measurement and can be calculated for either the left or the right ventricle. The Left Ventricular Ejection Fraction (LVEF) measures how much blood gets pumped from the left ventricle to the body, where the Right Ventricular Ejection Fraction (RVEF) measures how much blood is pumped out of the right side of the heart, to the lungs. Given that the two ventricular volumes are approximately equal, the ejection fractions of the left and right ventricles are typically similar with little variation. A significant discrepancy between the ejection fractions of the left and right ventricles can imply an underlying cardiac issue and the relative elevation or reduction of either ejection fractions can guide healthcare professionals towards a potential diagnosis. **(Mayo Clinic, n.d.)**

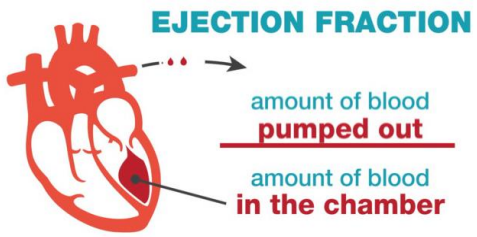


Figure 3: Left Ventricular Ejection Fraction Function
© digitalcommons.otterbein.edu

Due to its technical simplicity compared to calculating the RVEF and the significant impact to the systemic circulation, LVEF is more commonly measured and reported than RVEF, especially in echocardiograms. Ejection fraction typically refers to the left side of the heart and is the most used parameter to assess cardiac health in patients. Nonetheless, under certain conditions, assessment of the RVEF is of paramount importance.

According to the American Heart Association Left Ventricle Ejection Fraction is considered:

- Normal between 50% and 70%
- Mildly reduced between 41% and 49%.
- Reduced when 40% or less.

Echocardiography

An Echocardiogram is a first-line test using high frequency sound waves to map out the shape and size of the heart, thus allowing the doctor to see how well the heart pumps blood and look for abnormalities of the heart valves and heart walls (myocardium). While performing an echocardiogram the operator applies a cool gel on the chest and moves a transducer firmly through the gel that sends painless sound waves to the heart. The sound waves bounce off the heart and echo back to the transducer. The returning sound waves are converted to two-dimensional 2D moving images of the heart muscles, chambers and valves on a video screen. **(Josh Hopkins Medicine, n.d.)**

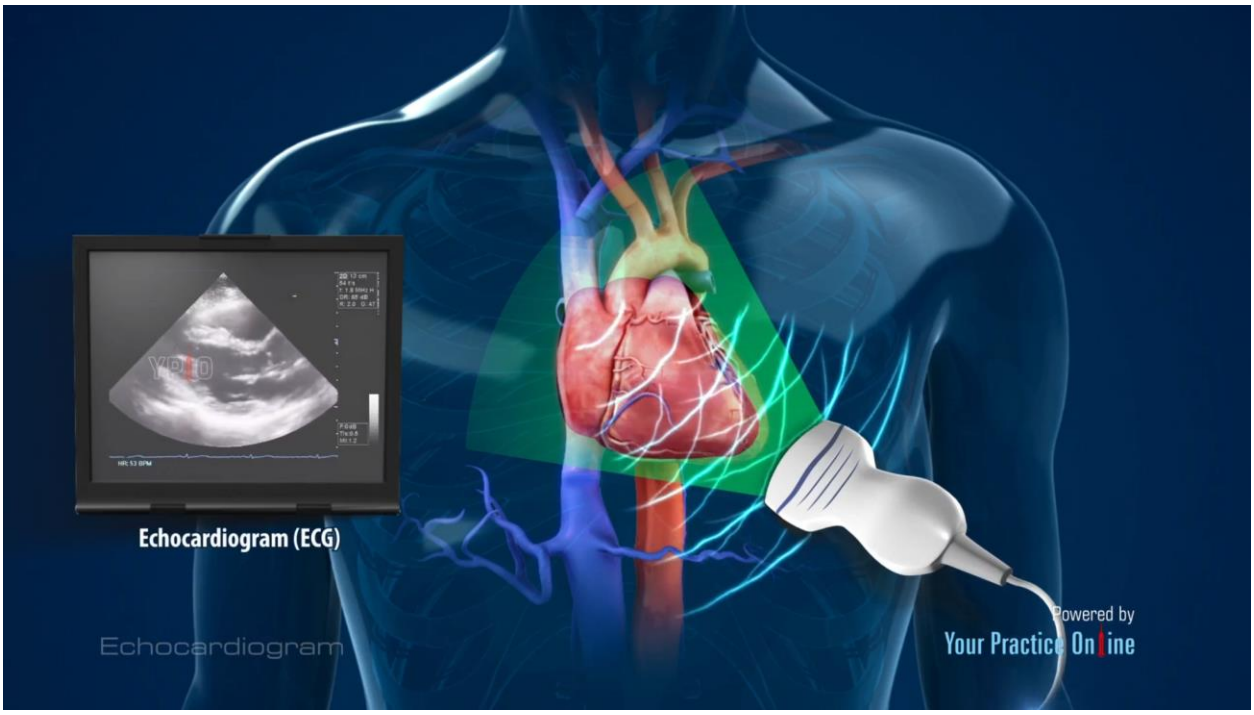


Figure 4: Transthoracic echocardiography © ypo.education

Problem Statement

Cardiovascular Diseases (CVDs) diachronically put immense pressure on the health systems globally. According to the World Health Organization Reports it is the leading cause of morbidity and mortality **(World Health Organization, 2021)**. The persistent dominance of Cardiovascular

Diseases (CVDs), emphasizes the constant need for groundbreaking image diagnostic methods. In the pediatric population, even though Cardiovascular Diseases (CVDs) are not the leading cause, there is a significant rise in the last decade that is connected to obesity due to the modern lifestyle.

The statistics on pediatric and congenital heart diseases are disheartening and they call for immediate action. 1 out of every 100 children born globally has a heart defect, known as congenital heart disease (CHD). 25% of these children will need medical intervention within their first years of life to survive. The survival of a child with a heart condition depends on their birth location and the availability of heart care treatment during their lifetime. Other heart conditions that develop during childhood, known as acquired heart diseases (AHD) also significantly add to the global burden of pediatric and congenital heart disease. **(World Heart Federation, 2023)**

Taking into consideration the perplexity of congenital heart defects as well the differences in anatomy and physiology while children grow up, echocardiography has emerged as the primary diagnostic tool, especially for evaluating pediatric cardiac diseases that accurate assessment is of vital importance **(Lai *et al.*, 2006)**. Quick and accurate diagnosis of the Left Ventricle Ejection Fraction (LVEF) is essential for early detection of cardiac disorder, effective treatment and improved outcomes of the affected children.

Incorporating cutting-edge technology such as machine learning and deep learning into pediatric echocardiography is creating new possibilities. These advancements seem to have the potential to tackle and even eliminate the existing gaps and strengthen the capabilities of pediatric cardiac care **(Slostad *et al.*, 2023)**. Researchers from around the world are actively pursuing these integrations aiming to transform the echocardiography and enhance the quality of life for the affected children.

Importance of Pediatric Echocardiography compared to other imaging diagnostic tools

Echocardiography serves as a fundamental pillar in the field of cardiology. It provides real-time, relatively easily acquired, non-invasive imaging of the heart using the harmless ultrasound wave technology. These capabilities allow healthcare professionals to observe the heart, evaluate ventricular functionality, and identify any abnormalities in either the valves or the myocardium

with substantial detail. Thus, showcasing its critical role in cardiac image diagnostics. **(Otto, 2019)**.

Especially in the field of pediatric cardiology, echocardiography is of great importance. Both congenital and acquired heart diseases of children present significant challenges in order to be diagnosed. The anatomical structures in pediatric patients are smaller, their heart rates are faster, and the presence of congenital anomalies can be more common **(Penny *et al.*, 2021)**.

The significance of echocardiography in pediatric patients is even more important when compared to alternative imaging techniques such as computerized tomography (CT) scans and Magnetic resonance imaging (MRI) **(Opfer, 2018)**. For the youngest patients, undergoing CT scans and MRIs can be particularly troublesome. In order to ensure image clarity, the patient must remain still. Something rather difficult for the youngest children and almost impossible for infants. Thus, it is often necessary to administer anesthesia to ensure no movement of the children. However, anesthesia not only introduces additional side effects but also emphasizes the stress and discomfort experienced by both the affected children and their families **(Bhargava *et al.*, 2013).**

Furthermore, cardiac MRIs require administration of contrast agents, which, even though are used to enhance image quality, have the potential for adverse reactions and side effects, particularly in pediatric populations with varying degrees of sensitivity **(Bhargava *et al.*, 2013)**.

Moreover, the inevitable exposure to ionizing radiation from CT scans, raises significant concerns for pediatric patients. The reason is that their fastly developing cells are more susceptible to radiation-induced harm which might lead to long-term risks, thus strengthening the importance of non-radiating alternatives such as echocardiography **(Chodick *et al.*, 2009).**

All things considered, the flexibility of echocardiography, which does not involve those risks along with the fact that it is able to deliver accurate assessments, is important for early detection and treatment that can significantly improve the prognosis and life quality for these young patients.

Additionally, echocardiography serves as a fundamental tool for monitoring the evolution of heart conditions and examining the effectiveness of treatment plans. It helps cardiologists to define the severity of the disease, plan surgical tactics if needed, and provide invaluable insights for clinical decision-making, therefore leading to personalized patient care **(Ash & Chowdhury, 2023)**.

Additional relatively recent advancements in echocardiographic technology have made significant contributions to its importance in cardiology. The introduction of techniques such as three-dimensional (ED) echocardiography, speckle-tracking and contrast echocardiography have expanded the capabilities of this image diagnostic method allowing clinicians to run more comprehensive evaluations of the heart **(Kamel *et al.*, 2022)**. These innovations have played a critical role in improving the precision of echocardiography as a diagnostic tool, particularly in pediatric cardiology.

Taking everything into account, echocardiography has been the preferred method and continues to constitute the most efficient image diagnostic tool in pediatric cardiac diagnoses.

Limitations in Existing Methods

Despite the advantages compared to other image diagnostic tools, the substantial progress made in echocardiography and its invaluable contributions to pediatric cardiology, there still exist considerable drawbacks. One major challenge is the fact that conventional echocardiography heavily relies on trained observers who are specifically familiar with the specific equipment they use. Additionally, pediatric echocardiography demands a high level of expertise that involves sufficient knowledge of the normal growth and development of children, the plethora of congenital and acquired heart defects and the principles of ultrasound physics **(Benavidez *et al.*, 2008).**

In addition, even though traditional 2D echocardiography is broadly accepted and applied, its most important drawback is that it depends greatly on the operator. Apart from the need of them being highly skilled, it introduces the margin of variability in image acquisition and interpretation between different operators that can potentially compromise the diagnostic accuracy and reliability **(Morbach *et al.*, 2018)**. This fact is known as inter-observer variability and is particularly worrying given the fact that when it comes to children the need for prompt and precise heart disease detection is critically important.

Furthermore, the manual analysis and interpretation of echocardiographic data can be time consuming and labor intensive. This is exceptionally difficult in pediatric healthcare where timely and accurate diagnoses are imperative. **(Ash & Chowdhury, 2023)**.

Moreover, pediatric echocardiography faces difficulties related to patient cooperation and image quality. It can be challenging for young children and infants to fully cooperate during the examination process, making it rather difficult for the operator to obtain clear imaging views. Poor image quality furthers the complexity of diagnosis that indicates the need for development of sophisticated techniques and technologies to improve image acquisition and analysis in pediatric populations where capturing the echocardiographic videos should last as briefly as possible **(Law *et al.*, 2017)**.

Considering all the above constraints, it is important to explore solutions and technological advancements that will help overcome the difficulties associated with pediatric echocardiography. The incorporation of automation, computer vision and deep learning, offers potential ways for improving the effectiveness and precision of echocardiographic assessments in pediatric cardiology **(Madani *et al.*, 2018)**.

Research Objectives and Hypotheses

The objective of this research is to employ computer vision and deep learning techniques in order to overcome the limitations that currently exist in pediatric echocardiography. The main goals of this study are the following:

1. Incorporate computer vision and machine learning techniques to analyze the dataset and extract important insights.
2. Utilization of the OpenCV Library to explore the echocardiographic video dataset and evaluate its effectiveness in processing and analyzing echocardiographic imagery.
3. Development of a Graphical User Interface to bridge the gap between complex computational techniques and practical clinical use.
4. Development of Automated Left Ventricular Segmentation Model: Develop and validate a deep learning model that can automatically segment the left ventricle in pediatric echocardiography videos.
5. Prediction of Ejection Fraction: Building on the left ventricular segmentation model, predict the ejection fraction (Regression Task) aiming to improve diagnostic accuracy.
6. Evaluation of Model Performance: Assess how well the developed models perform using appropriate metrics and a comparative analysis with existing models.

Hypotheses

H1: A deep learning model can accurately automatically segment the left ventricle in pediatric echocardiography videos.

H2: A deep learning model can predict the left ventricle ejection fraction (LVEF) in pediatric patients that is comparable to or even better than current methodologies.

H3: Integrating computer vision and deep learning in pediatric echocardiography can improve diagnostic accuracy and efficiency.

Justification for the Study

The primary reason for conducting this research is the existing limitations of echocardiography that plays a vital role in identifying and treating congenital and acquired Cardiovascular Diseases (CVDs) in children. In addition, the collateral risks of other imaging diagnostic methods, such as CTs and MRIs. However, under no circumstances does this research attempt to undermine the significant importance of those methods.

This study is both timely and relevant. On the one hand, recent advancements of artificial intelligence are being incorporated in medical diagnostic equipment. On the other hand, as of early 2023, the availability of an extensive pediatric echocardiographic dataset from Stanford University presents a unique opportunity to develop and validate advanced deep learning models specifically designed for pediatric echocardiography. **(Reddy C., 2022)**.

Ultimately, this research aims to support the overall objectives to improve, while minimize the current shortcomings, the ultrasound imaging diagnostic tools for the delicate age group of pediatric patients and to contribute to the advancements of utilizing artificial intelligence in healthcare equipment.

Thesis Structure

This thesis is structured into well-organized chapters, each serving a specific purpose in explaining the application of computer vision and deep learning in pediatric echocardiography. In Chapter 1

Chapter 1: Introduction

The first chapter introduces the reader to the characteristics of the heart and pediatric echocardiography that are referred to throughout this study, while outlining the current limitations as well as this research's significance, goals, and structure.

Chapter 2: Literature Review

Reviews the history of cardiac diagnostic methods, from obsolete techniques to current technologies. It discusses the integration of machine learning, emphasizes the importance of automation in echocardiography, and highlights the gaps and potential of using deep learning in diagnosing heart defects in children.

Chapter 3: Methodology

This chapter details the methodology followed in this study. Focused on the EchoNet-Pediatric dataset from Stanford, it explains the data acquisition process, provides an analysis of the dataset through multiple visualizations and dives into the methods for the task of left ventricular segmentation using traditional, machine learning and deep learning techniques. The chapter concludes by reaching the final goal of building a deep learning model capable of accurately predicting the left ventricle ejection fraction.

Chapter 4: Results

Reports the research outcomes, showcasing the final model's performance metrics while performing both a quantitative and a comparative analysis.

Chapter 5: Discussion

Interprets the results, emphasizing the role of computer vision and deep learning in enhancing pediatric echocardiography. It discusses the limitations and challenges faced throughout this study while providing areas for further exploration.

Chapter 6: Conclusion

Summarizes the key findings and contributions of the research, strengthening the potential of computer vision and deep learning in pediatric echocardiography. It also reflects on the broader implications of the study's results for the field and provides recommendations and considerations for future research endeavors in this domain.

Lastly, the study features a list of References for further reading, and Appendices containing supplementary material.

Chapter 2

Literature Review

Evolution of Cardiac Diagnostic Methods

The evolution of cardiac diagnostic methods portrays a fascinating journey with tremendous advancements in the last decades, that reflects the broader technological progressions of diagnostics in Medical Science. From retiring invasive techniques to non-invasive elementary diagnostic tools such as stethoscopes, leading to the new era of sophisticated methods such as advanced imaging technologies. Cardiac diagnostics have come a long way that not only allowed more accurate diagnosis, but have also minimized the risks associated with invasive procedures. The future is just around the corner. Machine learning has already been integrated into such tools, but deep learning has yet to prove that it is altering the ways cardiologists and radiologists diagnose heart defects.

Introduction to Cardiac Diagnostics

The interest of research on cardiovascular diseases has been undiminished since the development of modern medicine. The primary methods for the evaluation of heart conditions were invasive and often carried associated risks. Traditional diagnostic tools like stethoscopes and electrocardiograms (ECGs) have served as the basis in cardiology diagnosis and remain the primary evaluation tools for every cardiac check-up **(Kligfield *et al.*, 2007)**. However, as the understanding of cardiovascular diseases has progressed, so has the range of diagnostic tools available. Now, medical experts have a plethora of tests and imaging techniques at their disposal to obtain a thorough view of cardiac health. The evolution of these diagnostic methods has had an extreme impact not only on the early detection and management of heart-related conditions, but also in a decrease of mortality rates over the years.

Role of Imaging in Cardiology

The limitations and risks of traditional diagnostic methods have made them obsolete, making necessary the integration of advanced imaging technologies into cardiology. Technologies such as Echocardiography, Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) have changed the field of Cardiology irreversibly. They provide detailed anatomical and functional insights that have been greatly important in the diagnosis of simple to complex heart conditions **(Saeed *et al.*, 2015).** In addition, advances in imaging technology have set the basis to minimally invasive cardiac interventions with real-time guidance during procedures reducing the risks associated with traditional surgical methods **(Prempeh *et al.*, 2020)**.

Advent of Echocardiography

The introduction of echocardiography in the 1950s **(Edler** & Lindström, **2004)** was a significant moment for the field of Cardiology. Unlike methods, like chest x rays and invasive angiography, echocardiography provided a safe, non-invasive way to diagnose heart conditions that allowed the visualization of the structure and function of the heart in real time. Echocardiography enabled cardiologists and radiologists to assess the heart anatomy, evaluate ventricular function and identify myocardial abnormalities. This breakthrough was not a small improvement. It completely transformed the landscape of cardiac care. Thus, made echocardiography an essential diagnostic tool, particularly in situations where early intervention can be lifesaving. **(Keller *et al.*, 2023)**

Pediatric Echocardiography

Pediatric echocardiography is an area of expertise that focuses on diagnosing and treating heart conditions in children whether they are present from birth or acquired later. Unlike diagnosing heart problems in adults, pediatric cases come with their set of unique challenges that require customized approaches. Due to the size of their anatomical structures, their faster heart rates and the frequent presence of congenital anomalies, specialized equipment and expertise are necessary for accurate diagnosis and treatment. **(Lai *et al.*, 2006)**

Role of Automation in Echocardiography

In today's rapidly advancing world, there's an undeniable rise in the demand for imaging, propelling the necessity for diagnostic procedures that are not only efficient but also highly accurate. To address these challenges automation in echocardiography has emerged. Advanced software algorithms have recently been integrated to ultrasounds that can now automate echocardiographic processes, such as image acquisition and quantitative analysis (**Asch *et al.*, 2019**). However, they still depend on the experts to confirm their predictions, thus there is plenty of room for more study. This automation speeds up the diagnostic process, reduces the chances of human error and inter-observer variability. In addition, further advances have the potential to make quality cardiac diagnostics more accessible, especially for underdeveloped countries, by reducing costs and reliance on highly specialized personnel. **(Vidal-Perez R *et al.*, 2023)**

Need for Automation in Pediatric Cardiology

The particularly alarming statistics, especially for children, and their increased over time diagnostic accuracy on multiple cardiovascular diseases, resulted in escalated demand for advanced cardiac image diagnostic methods. Even though traditional echocardiography has been proven invaluable for evaluating heart related problems, its existing limitations for pediatric, but not limited to, patients can significantly impact the accuracy and reliability of the diagnoses. In addition, the various diseases that rely their diagnosis on imaging methods have put a significant burden on healthcare systems worldwide. As technology continues to advance in the healthcare industry, automated ultrasounds can play a significant role in addressing these challenges by speeding up diagnoses, reducing the risk of human error, reducing inter-observer variability, and ultimately improving patient outcomes. **(Karatxa *et al.*, 2022)**

Early Attempts at Automation

Automating cardiac diagnostics has been a topic of interest from researchers all over the world for quite some time. Initially the focus was on automating data collection and extracting insights. Although these early efforts were simple, they laid the foundation for advanced applications. Initially there was skepticism surrounding these endeavors due to their capabilities and high costs.

Nevertheless, as technology advanced these attempts were improved, leading to the development of sophisticated automated systems that are now considered essentials in modern cardiac diagnostics. **(Slomka *et al.*, 2017)**

Machine Learning in Echocardiography

Machine learning has made significant advancements in the field of automated echocardiography. Machine learning models employ algorithms that learn from existing data and are able to identify patterns and make highly accurate predictions. This progress has opened up opportunities for research and clinical applications in echocardiography, especially in automatically detecting cardiac irregularities and providing real time guidance to medical professionals. Such algorithms are already integrated into the latest ultrasound systems. The ability of machine learning to handle big amounts of image data and generate valuable insights has revolutionized the field and paved the way to more efficient cardiac care. **(Zhou *et al.*, 2021)**

Challenges and Limitations

Even though remarkable progress has been made in automating echocardiography, it has its limitations. A profound challenge is known as the "black box" problem. These algorithms can be quite complex for healthcare professionals to understand, which can undermine the trust in automated systems.

In addition, the accuracy and reliability of machine learning applications depend heavily on the quality and quantity of the training data. If the data used is inaccurate or biased it can lead to inaccurate diagnoses that jeopardize patient outcomes.

Those are the reasons behind the fact that modern ultrasound systems rely heavily on the operator's confirmation of the produced predictions. Lastly, the implementation of automated systems raises ethical considerations such as concerns regarding data privacy and potential uncertainties that may alter the healthcare industry. **(Vandenberk *et al.*, 2023)**

Deep Learning in Medical Imaging

The relatively recent technological advancements and availability of computational resources allowed the exponential rise of deep learning which has revolutionized various industries. This groundbreaking development has completely transformed how we analyze, understand and utilize complex data such as imaging. In the healthcare industry deep learning has made a tremendous impact, especially in the field of medical imaging.

Introduction to Deep Learning

Deep learning, a subfield of machine learning, has attracted attention due to its ability to analyze complex and high dimensional datasets using neural networks with multiple layers, an approach that imitates the functioning of the human brain neural network. Based on the principles of artificial intelligence, deep learning has enabled an unprecedented revolution in various areas, such as natural language processing and computer vision.

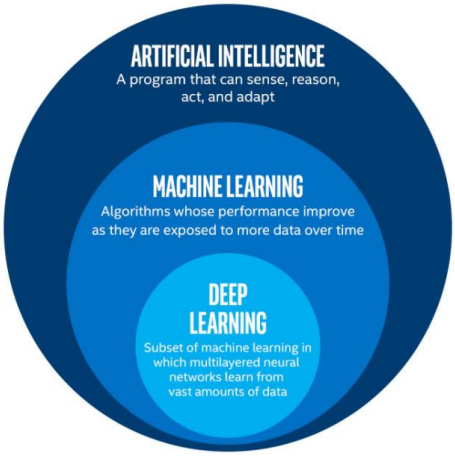


Figure 5: AI vs ML vs DL ©TowardsDataScience

Its effectiveness is also being acknowledged in the field of Medical Science, where imaging diagnostic tools are of great importance. As a result, it seems to be the start of a new era which may potentially completely transform modern medicine as we know it.

Deep Learning in Medical Imaging

Deep Learning has had a great impact on the domain of image analysis over the recent years; especially through the use of Convolutional Neural Networks (CNNs). CNNs are feed forward neural networks that have achieved great recognition for their ability to identify patterns in images. In medical imaging, CNNs have transformed diagnostic tools such as X rays, MRIs and CT scans by enabling accurate object recognition and semantic segmentation. Compared to traditional machine learning algorithms, CNNs consistently outperform them both in terms of speed and diagnostic accuracy, making them highly valuable. It is worth noting that deep learning has gained

ground across various fields but its impact on healthcare stands out as it benefits human health. **(Kim *et al.*, 2019)**

Deep Learning in Echocardiography

The use of deep learning in echocardiography is a rapidly developing area of research held for tasks such as image segmentation and functional assessments. U-Net architectures which were originally designed for segmenting biomedical images, have exhibited exceptional results in accurately identifying cardiac structures in echocardiographic videos **(Ronneberger *et al.*, 2015)**.

In addition, advanced techniques called spatial transformer networks have been utilized to predict both right and left ventricular ejection fraction with impressive accuracy and represents a significant advancement in the automated assessment of cardiac function. These networks are adept at capturing both spatial and temporal features in echocardiographic videos, which are crucial for accurately evaluating the dynamic nature of cardiac activities. The Echonet-Dynamic model **(Ouyang *et al.*, 2020),** for example, is a pivotal resource in this domain. These models are specifically tailored to parse the intricate movements and morphological changes of the heart over time, crucial for precise Ejection Fraction computation. The spatial component allows the detailed mapping of cardiac structures, while the temporal aspect captures the rhythmic contraction and relaxation cycles of the heart. This combination is crucial for understanding the nuanced variations in cardiac function, particularly in the context of Ejection Fraction, which is a key indicator of cardiac health.

These advancements are greatly significant in echocardiography, where measurements of high accuracy and timely diagnoses can have a huge impact on both patients and doctors. The utilization of Deep Learning and its capabilities has initiated a new era of echocardiography research where there will be limited dependance on human operators. **(Krittanawong *et al.*, 2023)**

Performance Metrics

To determine the effectiveness of deep learning models in echocardiography, it is imperative to use specific performance metrics that depend on the task at hand. While basic measures like accuracy may be suitable for certain tasks, sensitivity (aka Recall, True Positive Rate) and specificity (True Negative Rate) are more suitable for classification assignments. Specifically in

medical research they are used to evaluate the effectiveness of the alternative test compared to the "gold standard" **(Hariton & Locascio, 2018)**. For image segmentation tasks, the Dice coefficient is greatly accepted and it measures how well the segmented images match the ground truth images **(Zhu *et al.*, 2004).** When evaluating models that predict the ejection fractions (regression), the coefficient of determination or r^2 provides the most reliable criterion. Carefully selecting the appropriate performance metrics not only ensures the comprehensive validation of the deep learning models, but also paves the way for successful implementation within clinical settings.

Research Gaps and Future Prospects

As the field of echocardiography begins to explore the possibilities offered by machine and deep learning advancements, it becomes important to explore existing research gaps and possible avenues for further development. **(Lopez-Jimenez *et al.*, 2020)**

Identifying the Gaps

Lack of Large-Scale, Diverse Pediatric Datasets

A notable drawback in echocardiographic research is the lack of datasets specifically focused on pediatric cases. Most available datasets either come from adult cases **(Ouyang *et al.*, 2020)** or lack the necessary diversity to effectively develop and evaluate robust deep learning models. This limitation makes it difficult to generalize the study findings to the field of echocardiography.

Need for More Reliable and Clinically Validated Models

Over the last decade, there has been an increase in the development of such models for echocardiography. However, there is a need for those models to be clinically validated. Although many of these may demonstrate remarkable results on some metrics, they often either fall short in establishing their clinical reliability to real-world case scenarios or lack comparisons with expert opinion in real-world clinical settings. This gap is a barrier to the adoption of automated systems in clinical settings.

Future Prospects

The Potential of Integrating AI Models into Clinical Workflows

We approach the dawn of a new era in healthcare where there is tremendous potential for incorporating artificial intelligence into existing clinical equipment. The integration of machine and deep learning in echocardiography, and medical imaging in general, holds the potential to not only revolutionize this equipment, but also redefine the medical diagnostic procedures. This will give the opportunity to healthcare professionals to receive automated more precise and dependable results with minimal interaction with the equipment, while allowing them to focus their attention on other more complex diagnostic challenges. Ultimately, Artificial Intelligence (AI) in healthcare has the ability to enhance patient outcomes, reduce healthcare expenses and establish more effective diagnostic procedures into the healthcare system.

Importance of Interdisciplinary Collaboration

The future of echocardiography research goes beyond technological advancements. It also involves the promotion of interdisciplinary synergies **(Basu *et al.*, 2020)**. As this field advances the collaboration between data scientists, clinicians, radiologists and cardiologists becomes increasingly vital. These partnerships form the foundation for developing solutions that are not only technologically innovative, but also reliable in a clinical setting. The union of technology, expertise and practice can have a catalytic impact on medical research findings that hold the potential of unprecedented imaging diagnostic methods.

Chapter 3

Research Methodology

Acquiring the Dataset

The dataset used in this study was acquired from Stanford's Center for Artificial Intelligence in Medicine & Imaging (AIMI). The Stanford AIMI Center was established in 2018 and its mission is to empower artificial intelligence research for solving clinically important problems.

The specific dataset acquired is called EchoNet-Pediatric and it includes the first large pediatric video dataset of echocardiograms with human expert annotations for computer vision research as of today. It was published in Stanford's AIMI Shared Datasets in January 2023. The size of the dataset is 2.51GB and it includes avi and csv files.

The dataset is Licensed under Stanford University Dataset Research Use Agreement **(Appendix A).** Thus, the author of this thesis has acquired access by registering to the platform, accepted the terms and is personally fully bound by them.

The choice of this dataset was motivated by several factors. The fact that it is the first extensive collection of labeled pediatric cardiac videos makes it ideal for training robust machine learning models, while providing a great opportunity for unique research in the form of the thesis that incorporates a variety of knowledge subjects acquired in the scope of the master's in Data Science at the American College of Greece. Additionally, the fact that it involves the delegate pediatric population where early and accurate diagnosis is of critical importance. Lastly, on a personal note, my personal struggles with myocarditis have deepened my familiarity with the cardiac structures that fueled my passion for this research.

Dataset Description

The EchoNet-Peds database consists of 7,643 labeled echocardiogram videos from pediatric patients imaged between 2014 and 2021 as part of routine clinical care at Lucile Packard Children's Hospital at Stanford. All videos underwent cropping and masking to exclude text and irrelevant data outside the scanning sector. Subsequently the processed images were then down sampled using cubic interpolation into standardized 112x112 pixel videos.

The typical echocardiography procedure includes a series of videos and images that are extracted by visualizing the heart from different angles based on the position of the transducer. However, this dataset focuses on videos from two specific views, subsequently it is divided into two parts. 3,176 Apical 4-Chamber (A4C) echocardiography videos and 4,424 Parasternal Short Axis (PSAX) echocardiography videos.

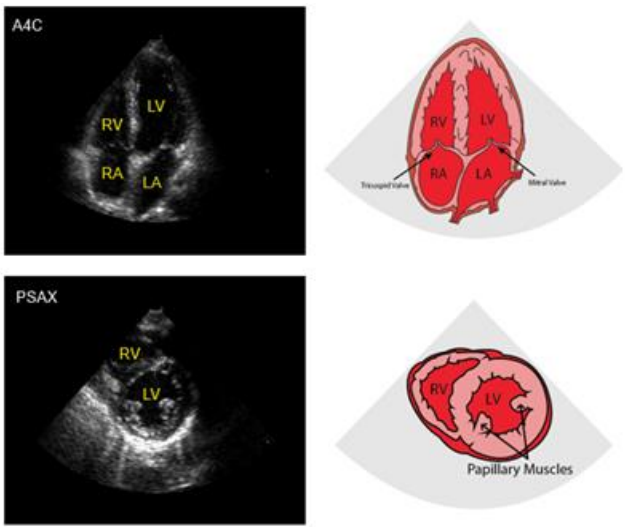


Figure 6: Apical 4-Chamber and Parasternal Short Axis angles and corresponding echocardiograms ©pocus101.com

For each video category (A4C and PSAX) the dataset presents two csv files named FileList.csv and VolumeTracings.csv each containing different human expert annotations (measurements, tracings, and calculations). Each csv file contains the file name of the video, thus allowing the

merging of varied data within the dataset. The primary contains demographics of the pediatrics patient: Sex, Age, Weight, Height as well as the EF annotated by the experts, while the latter contains coordinates and frame number denoting the areas of two time points (frames) representing the End Systole and the End Diastole.

Technical acquisition of data.

Every patient's study is associated with clinical measurements and calculations obtained by a registered sonographer and verified by an expert physician echocardiographer as part of the regular clinical procedure. The method employed to determine the ejection fraction in the pediatric dataset is the “Bullet Method” or the “5/6 Area Length Method,”. This method calculates left ventricular volumes using both the apical and parasternal short axis perspectives. In the dataset, for each video, the left ventricle is traced at the endocardial border at two distinct moments: End Systole and End Diastole. Each tracing is used to estimate ventricular volume by integration of the ventricular area over the length of the major axis of the ventricle. The expert tracings are represented by a collection of paired coordinates corresponding to each human tracing.

Working Environment

Throughout this thesis, the programming language used was Python, version 3.10. Python is a versatile, high-level programming language that is widely used in data analysis, web development, artificial intelligence, and many other fields.

For the development and execution of code, PyCharm Integrated Development Environment (IDE), specifically version 2022.3.2 of the Professional edition was used. PyCharm, developed by JetBrains, provides a comprehensive set of tools for Python developers, including intelligent code completion, code analysis, and integrated testing.

Pre-processing.

The dataset is well organized in a tree model. The folder pediatric_echo_dataset contains two subfolders A4C and PSAX, that each contains a subfolder named videos and the two csv files, FileList.csv and VolumeTracings.csv **(Appendix B)**

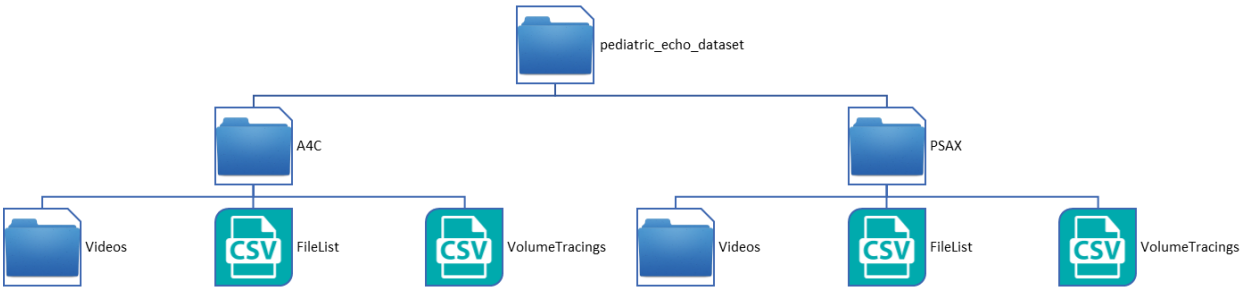


Figure 7: File System Hierarchy

Data Organization and Cleaning

Navigating through the dataset it is apparent that it contains complexities that rely on the fact that for each context addressed in this thesis different sub-datasets are needed. Each research question or analytical task demanded the unification of different parts of the dataset depending on the task at hand. Such an approach ensured that for every analytical task there was a sub-dataset tailored to its specific needs. However, it is important to note that the A4C video angle was the primary focus of the study due to its significance for evaluating the Left Ventricle Ejection Fraction.

To comprehensively process, clean and structure the appropriate data set in order to extract insightful knowledge and create statistics, the integration of the demographic information that is contained in the "Filelist.csv" and both Apical 4-Chamber (A4C) and Parasternal Short Axis (PSAX) videos were used.

The data preprocessing began with the goal of creating a thorough dataframe for demographics and video observations. Starting with defining utility functions, one to extract the patient ID from each filename, others to compute the number of data frames and the duration of the videos incorporating appropriate cv2 library's functions. The datasets from both A4C and PSAX folders were loaded and processed separately and the information of each was stored into two distinct data frames that subsequently merged in one on the Patient ID column.

To comprehensively explore the dataset, it was essential to examine all the possible scenarios. This was achieved by employing the union operation of set theory which in python is incorporated as a full outer join. Thus, the resulting merged data frame captures all the possible outcomes; instances where A4C videos correspond with PSAX videos and vice versa, as well as cases where a video in one category lacks a corresponding video in the other.

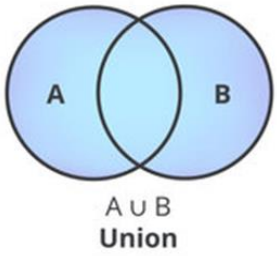


Figure 8: Union of Sets

After merging the datasets, the next step involved a thorough examination of missing values in the final dataframe. As anticipated, there was a discrepancy between the number of A4C and PSAX videos. Additionally, the demographic details from both csv files were cross checked. This served a dual purpose. On the one hand to check the correctness of the merge operation of all the sub datasets, and on the other, to ensure that the videos corresponding to each patient ID were accurately paired.

Specifically, videos without the Apical 4-Chamber (A4C) view were identified and excluded from the dataset, given the primary focus of this thesis on Apical 4-Chamber (A4C) videos. To maintain dataset integrity, data rows with missing weight (9 data points), height values (15 data points), or gender values that were neither male [M] nor female [F] (12 data points) were omitted. Additionally, a discrepancy involving duplicate PSAX videos for identical patient IDs was resolved. The data frame was sorted in descending order based on the number of frames for both A4C and PSAX videos, grouped by patient ID, with only the first entry (longest video) retained for subsequent tasks.

As a result, out of the initial 3176 A4C videos the final data frame contained 3149. All insights were exported to an excel file for further verification. **(Appendix C)**

Visualizations and Data Exploration

With the dataset now cleaned and organized, the focus shifts to detailed analysis. Utilizing statistical methods and visualization tools, key insights and patterns within the pediatric cardiac data will be identified. Subsequent sections will present these findings in a structured manner.

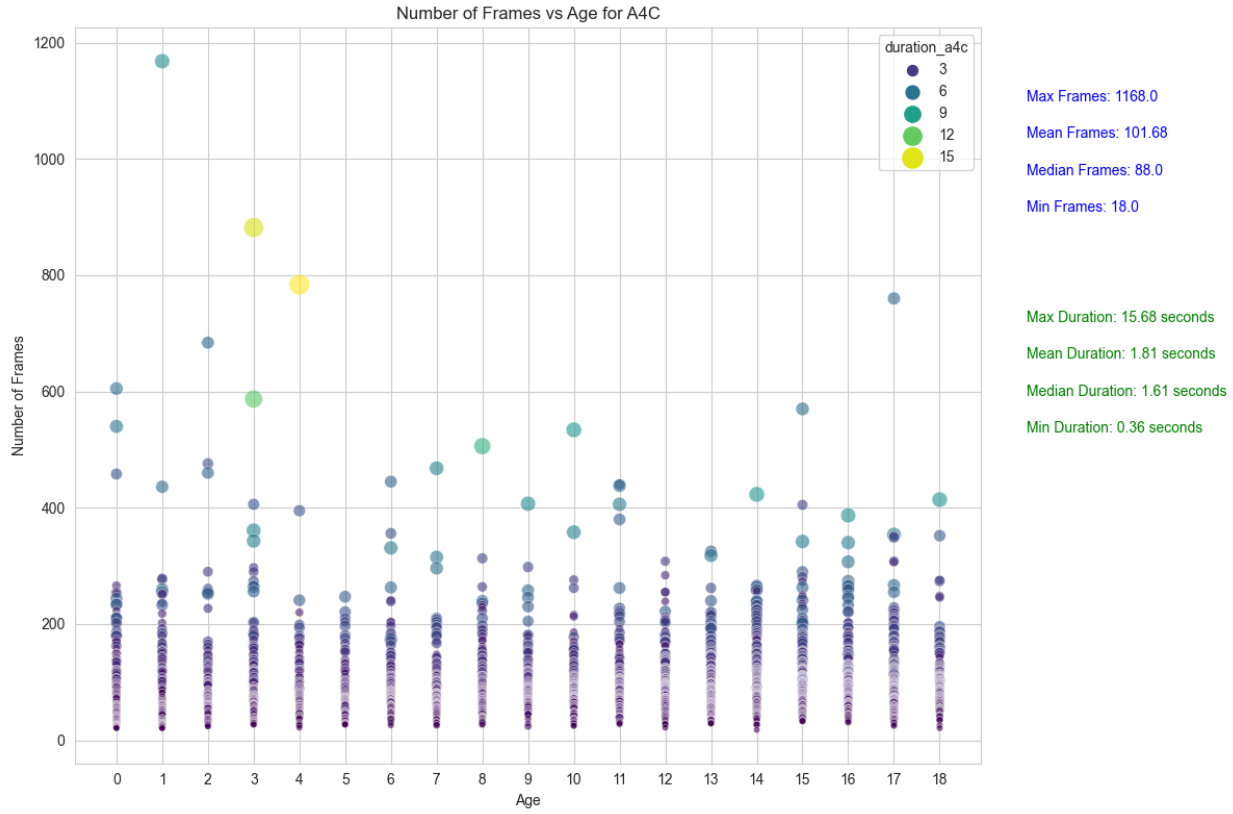


Figure 9: Frames per video per age and their corresponding duration

The data spans 18 distinct age groups, from newborns (age 0) to adolescents (age 18). The side legends provide useful statistical insights, such as the mean and maximum frames and duration. The number of frames varies widely and subsequently the duration, since they are directly related, across different ages. This suggests that the number of frames of the echocardiographic videos is not consistent, something that will play an important role later in this study.

Moreover, it is profound that the higher the age the longer the video duration which is directly linked to the patient's cooperation. In addition, there is a dense cluster of data points in the lower age range, particularly between ages 0 and 5, indicating that the dataset contains more data for younger ages, possibly due to congenital heart defects. Lastly, there are several outliers across different ages, particularly noticeable at ages 1, 3, 4, and 17 that are significantly longer than the typical videos for that age group.

Pediatric Ejection Fraction Distribution with Critical (Below 40%) and Cautionary (40%-50%) Thresholds

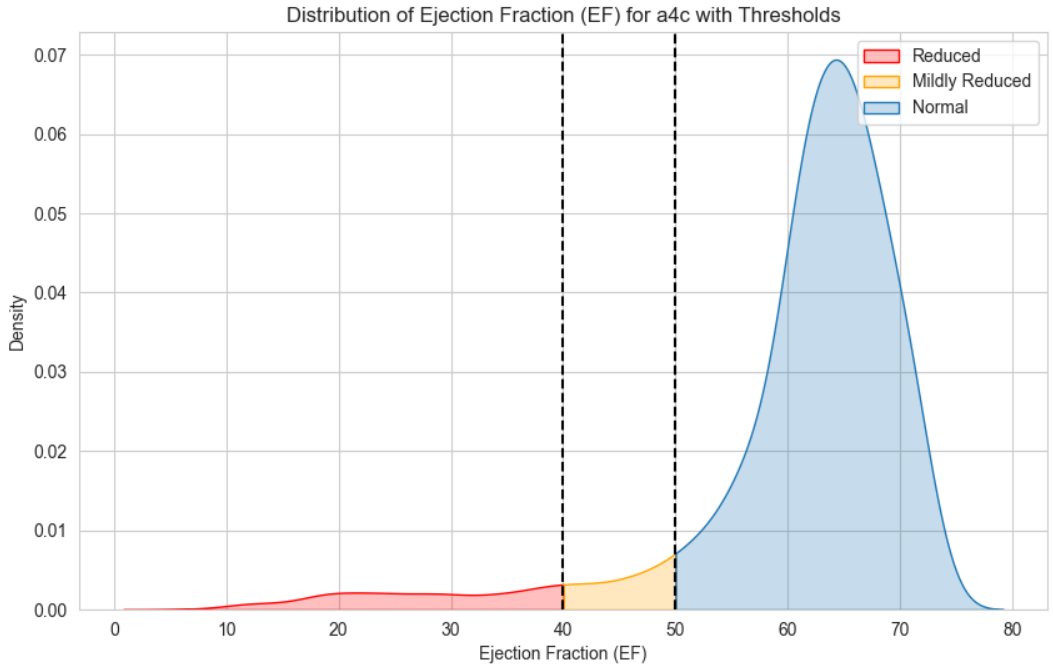


Figure 10: Distribution of Ejection Fraction, color categorized and Cautionary Thresholds

The plot has two vertical dashed lines that represent the two clinical thresholds. EF values that are below 40%, red-shaded area, indicate a severe heart defect where the heart cannot effectively pump oxygenated blood to the body. Values between 40% and 50%, orange-shaded area, indicate a mild reduction, and lastly values above 50%, blue-shaded area, that are considered normal.

The distribution plot shows a notable concentration of normal EF values in contrast to the other two regions showcasing that the dataset is skewed to the normal EF region. Even though a substantial portion of the dataset indicates a general healthy cardiac function among the observed children, there is a non-negligible presence of a subset of the population that exhibits not normal cardiac function.

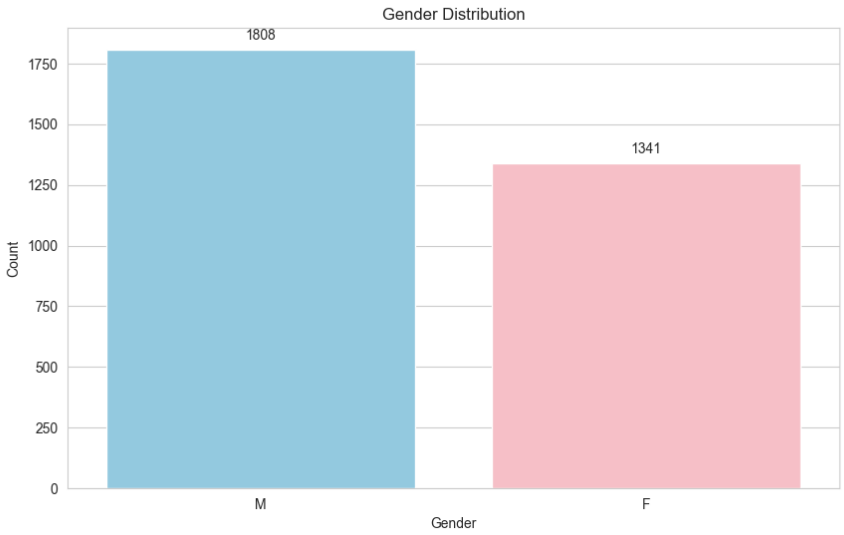
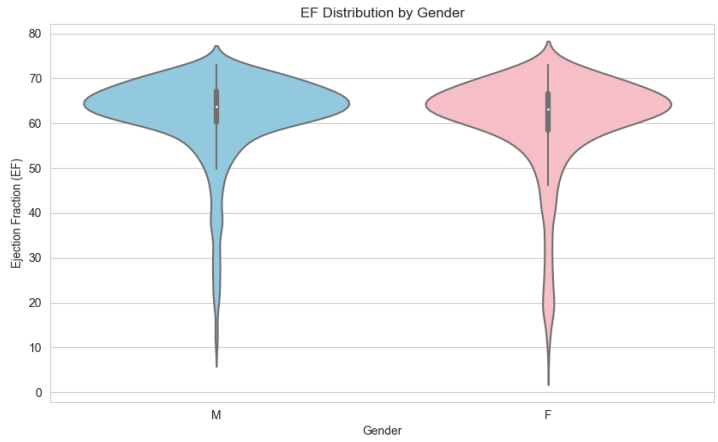


Figure 11: Gender Distribution Plot

There is a noticeable imbalance in the number of male (1808 data points) and female (1351 data points) pediatric patients, with more data available for males. The imbalance in gender distribution could potentially introduce bias in subsequent analyses.



F

Figure 12: Gender-wise EF distribution in a violin plot

Overall, both genders exhibit similar EF distributions, with a significant skew towards higher EF values. The median appears to be quite similar between the two genders, indicating comparable

central tendencies, however the male distribution appears slightly more compact around the median compared to females.

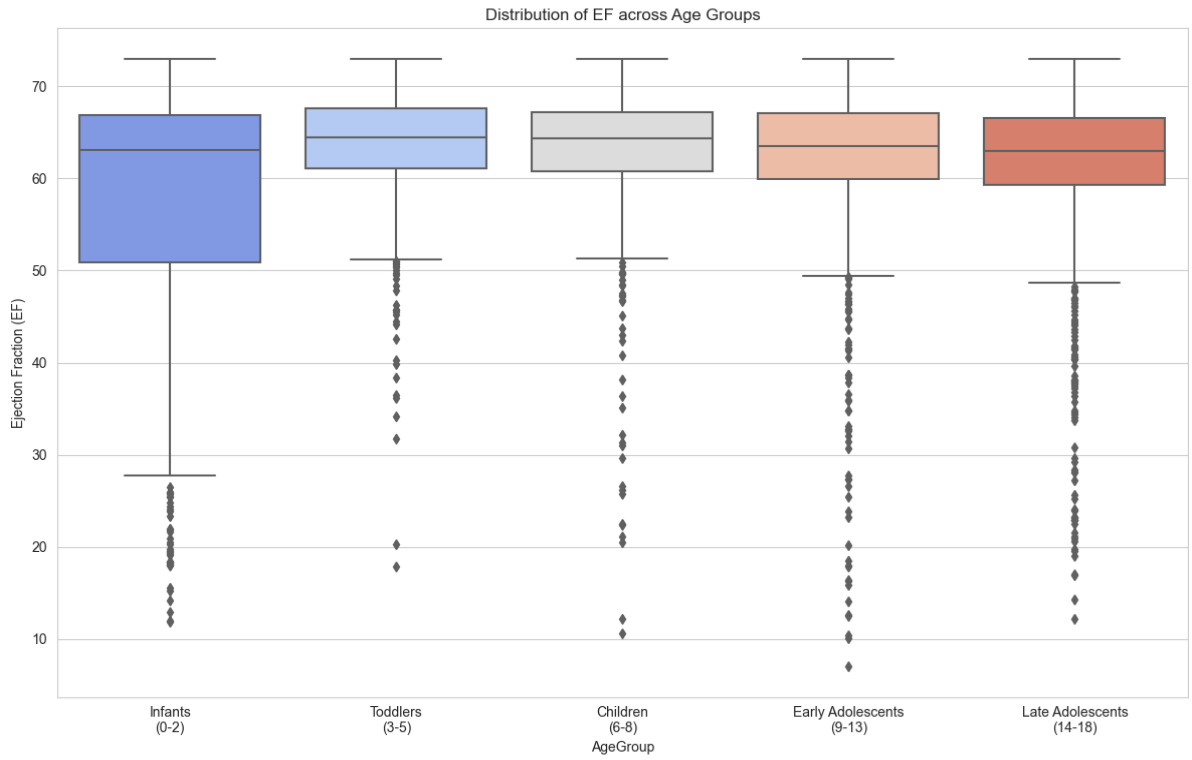


Figure 13: EF Distribution in different age groups in box plots

The boxplot shows a considerable variability in EF ranges within each age group while all age groups present outliers, however infants show notably wider interquartile range. This showcases the impact of congenital cardiac defects. The median EF values appear relatively stable across different age groups, indicating consistent central tendencies in cardiac function despite age variations.

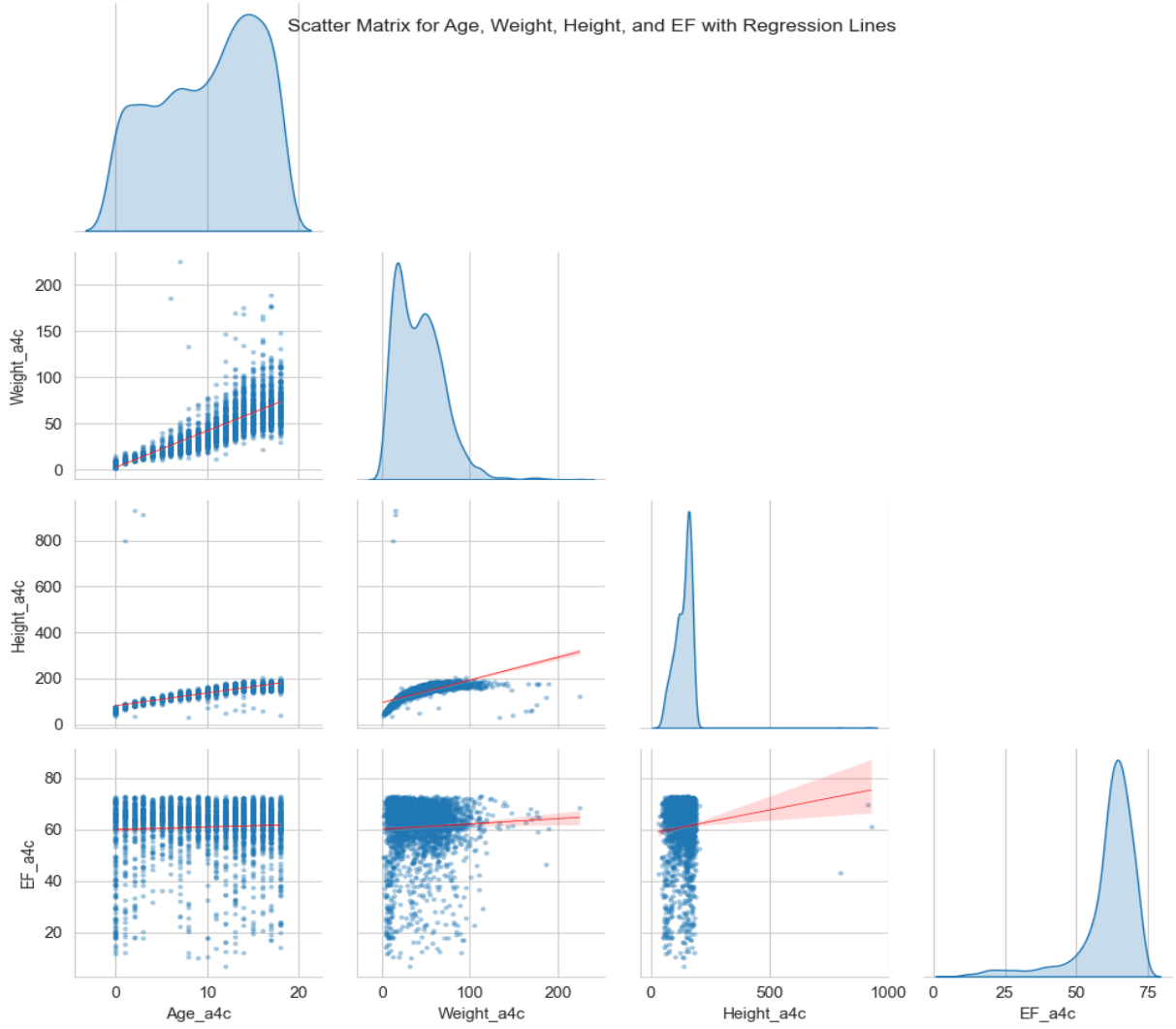


Figure 14: Scatter matrix of demographics and EF, showing linear correlations.

The scatter matrix **(Seaborn, n.d.)** provides a visual exploration of potential relationships between pairs of features and can provide further insights using multivariate analysis techniques. It is profound that there are relatively linear relationships both between Weight and Height, and between Height and Age, aligning with anticipated child growth and development. On the contrary, the relationship between Age and Weight is lacking linearity, especially for ages above seven, potentially highlighting the influence of childhood obesity on cardiac function. Furthermore, in certain subplots, such as Ejection Fraction (EF) versus Weight, definite outliers

are observed. Taking everything into account, by investigating all three features (age, weight, height) compared to the Ejection Fraction (EF) none of them demonstrate a clear linear relationship, indicating that the Ejection Fraction (EF) is not linearly predictable solely based on these variables.

Video Preprocessing Part 1

Introduction

Echocardiography provides a visual window into the heart that offers important details about the heart's structure and functioning. The Apical 4-Chamber (A4C) view is particularly important since it allows the operator to observe all four chambers of the heart and assess the functionality of the atrioventricular valves.

The objective of the video preprocessing stage was to explore the video data using traditional computer vision techniques and explore the likelihood of automatically segmenting the left ventricle.

Managing a dataset with video data presents challenges due to its heterogeneity, especially when compared to more uniform datasets, such as those in Excel, CSV, or JSON formats. To navigate through these challenges efficiently, a graphical user interface (GUI) has been employed, enabling easy, quick, and effective examination of the dataset by incorporating multiple techniques from the OpenCV libraries.

Developing an Interactive Platform through a GUI

The exploration of the echocardiographic video data began with the design and development of a graphical user interface (GUI) by taking advantage of the capabilities of the PyQt6 library **(Riverbank Computing, n.d.)**. PyQt6 is a set of Python bindings of the Qt toolkit which is a robust cross-platform framework that provides multiple tools for graphic user interface (GUI) creation. This strategy was selected not only to provide a user-friendly platform that can allow users to interact with the Apical 4-Chamber (A4C) videos, but also serves as a holistic analytical environment for video preprocessing tasks.

Figure 15: Graphical User Interface based on PyQt6.

GUI Components and User Interaction

The graphical user interface (GUI) is architecturally structured to include various widgets, such as video display labels, buttons, sliders, and checkboxes, each meticulously incorporated to facilitate fluid user interaction with the echocardiographic data.

The interface is divided into three distinct frames, each designed for a specific purpose. The left frame, or layout as denoted in PyQT6, allows the user interactivity with multiple image processing functionalities. Meanwhile the right frame is partitioned into the top and bottom layouts, where the former is dedicated to always display the original video and the latter to dynamically display the processing results of the echocardiographic videos based on the filters chosen by the end user

in real time. This structure offers an immediate visual comparison and enables the exploration of the impact of various computer vision processing techniques on the visual data.

Utilizing OpenCV for In-Depth Analysis of Cardiac Morphology

OpenCV is pivotal to multiple methodologies in this study **(OpenCV, n.d.)** due to the dataset. Abbreviated from "Open-Source Computer Vision", OpenCV is an open-source computer vision and machine learning framework that offers a plethora of functionalities enabling the exploration and analysis of visual data through various image processing techniques. It is widely used across diverse domains for real-time image processing tasks; however, the rise of deep learning tends to minimize its usage due to the fact that certain tasks can be time consuming and sensitive to changes especially when it has to deal with not so homogenous datasets. Nevertheless, it is the de facto way for certain tasks such as image/video reading, data augmentation, tracking objects and other simple computer vision implementations.

In the context of echocardiographic video data, OpenCV is employed to conduct a meticulous analysis of the anatomical structures observable in the video frames, since the library allows the application and fine-tuning of numerous filters and morphological operations on structural elements of the heart as well as an attempt to segment the left ventricle of the heart echocardiographic data.

Images and videos: A Sequential correlation.

The term image processing refers to the analysis and manipulation of a digitized image, while a video represents a sequence of images over time. Mathematically, images are constructed as arrays of numerical values, ranging between 0 and 255, where each value corresponds to a pixel’s intensity. For color images, there are 3 primary channels (most encountered RGB representing red, green and blue). Each channel encodes the intensity of its respective color for any pixel in an image.

In the context of this study on echocardiographic data, each video provided consists of a sequence of frames (images), capturing the dynamic movements and changes in the heart's chambers and valves over time. This allows the understanding not only of spatial features and patterns within each individual frame but also of sequential patterns and changes occurring across frames.

Given the fact that the video data size is 112x112, which is considerably small, a conscientious decision was made to resize the videos to 400x400 to ensure proper depiction in the graphic user interface (GUI). By using OpenCV's resizing functionality, which augments the pixel count in both the horizontal and vertical dimensions, allowed more spacious visuals without compromising image quality, allowing more effective inspections.

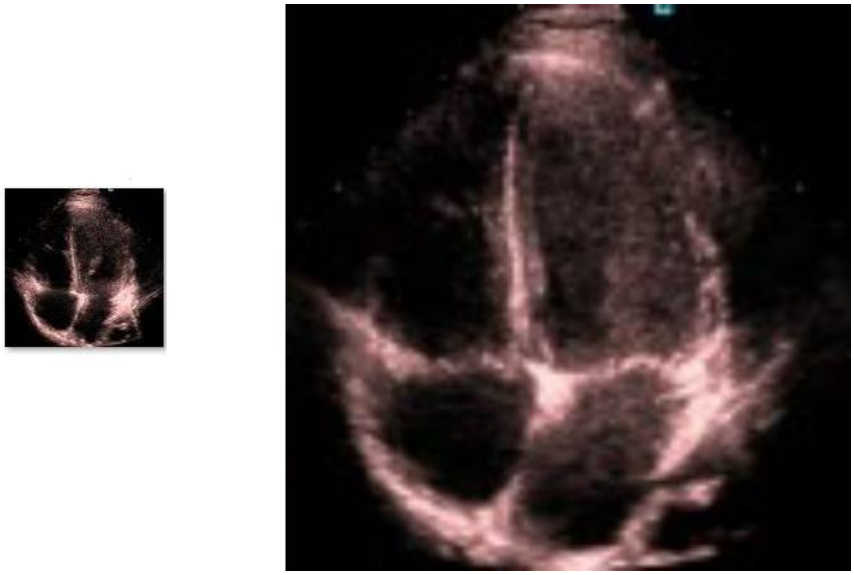


Figure 16: Original 112x112 size (left) vs Magnified 400x400 size (right)

Interactivity, Grayscale Conversion, Histogram and Enhancement:

The process of selecting a new echocardiographic video is facilitated through a 'Select Video' button, which triggers a file dialog window, allowing the user to navigate through the filesystem and select the desired video.

The EchoNet-Pediatric dataset consists of 3-channel videos. Given the complexity and high-dimensionality associated with processing three-channel data and the need to homogenize the dataset, the initial processing step requires the conversion of the frames into grayscale.

Figure 17: 3-channels videos

When this is selected within the graphical user interface (GUI), a real time histogram of the grayscaled video frames is displayed in the bottom-left section, providing a visual representation of the pixel intensity distribution across the images. This frame-by-frame image histogram allows insights about the brightness levels of the sequential images.

Figure 18: Histogram of a single frame

To further enhance the visual understanding of the echocardiographic data, histogram equalization is employed as an optional mechanism to amplify the contrast within the images. This method essentially redistributes the dominant intensity values across the image, enhancing the contrast between lighter and darker areas. However, in the context of echocardiography this method did not prove beneficial since it introduced a considerable amount of noise into the frames, making it almost impossible for additional OpenCV techniques to perform properly.

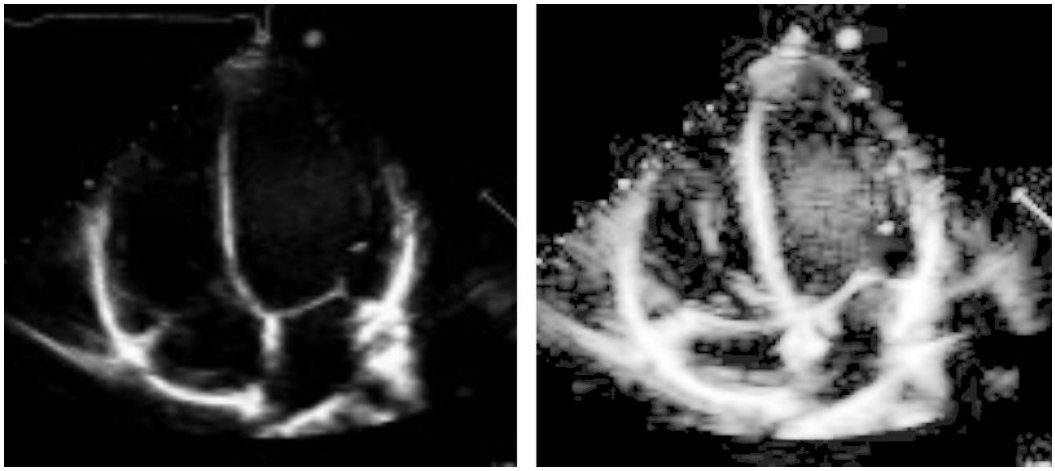


Figure 19: Original (left) vs Histogram Equalization (right)

Figure 20: Their corresponding Histograms

Filter Application and Morphological Operations.

The constant movement of cardiac muscles results in ultrasound noise across the frames. The implementation of smoothing filters through OpenCV, specifically Gaussian and Median, were considered beneficial to refine the visual quality of the echocardiographic images and reduce the noise and detail in an attempt to distinguish the myocardium. The Gaussian filter employs a Gaussian-distributed kernel surrounding a pixel and replaces it with the value of its local neighborhood. On the other hand, the Median filter, good at preserving boundaries, replaces each pixel value with the median of neighboring pixels, particularly effective in dealing with salt-and-pepper noise. The user can select the kernel size, with values between 3 and 17 and both filters provide a smoother visual output, with the Gaussian offering a more natural blurring, thus ensuring a denoised and more visually coherent result.

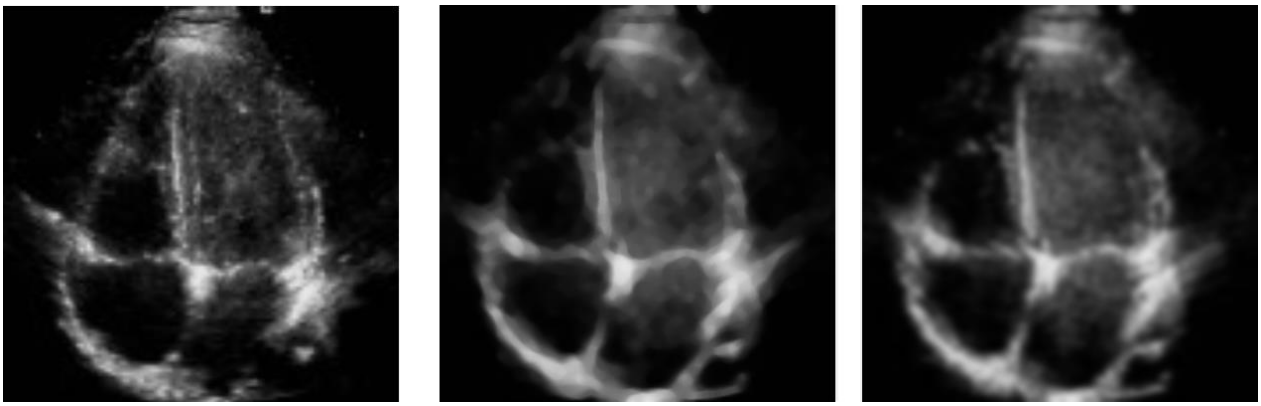


Figure 21: Original Frame (left) vs Median Blur (Middle) vs Gaussian Blur (right) (kernel 11x11)

Morphological operations such as erosion, dilation, opening, and closing are included to give the user more control over the morphological analysis of the echocardiographic images. Erosion makes the edges of objects in the foreground smaller, usually used to reduce object size in an image, while dilation does the opposite, often used to fill in small holes and connect separate areas. Opening, which is erosion followed by dilation, and closing, which is the opposite, are especially

good at removing small noise from the foreground and background respectively. Through the graphical user interface (GUI), users can enable these operations allowing adjustments to further refine the segmentation of structures within the images.

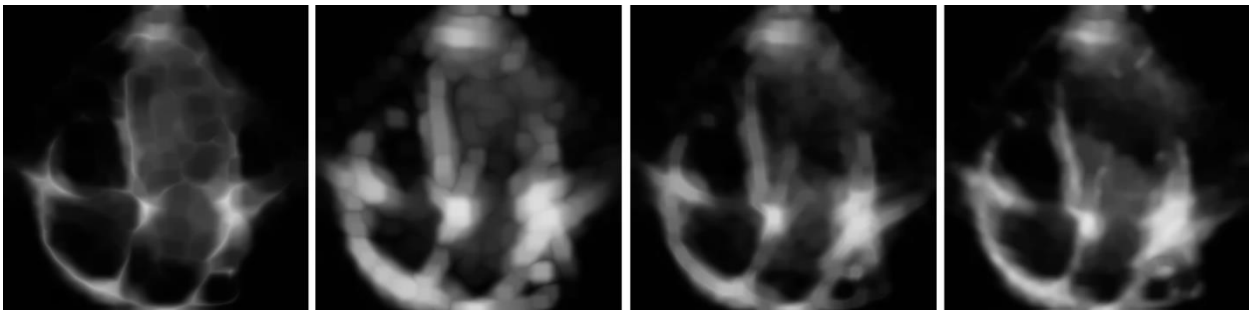


Figure 22: Erosion (left), Dilation (center left), Opening (center right), Closing (right) (kernel: 11x11)

Edge Detection and Structural Analysis

In the pursuit of extracting structural boundaries in order to segment the left ventricle the Canny Edge detection algorithm was employed in order to display important outlines within the echocardiograms. The graphical user interface (GUI) allows the user to set the upper and lower threshold values, which is key to making the algorithm work well with different echocardiograms allowing the edge detection to adjust to different visibility and contrast conditions in different echocardiographic data. Thus, this interaction makes sure that the edge detection is strong and flexible, while efficiently handling the varied and dynamic nature of the echocardiographic images.

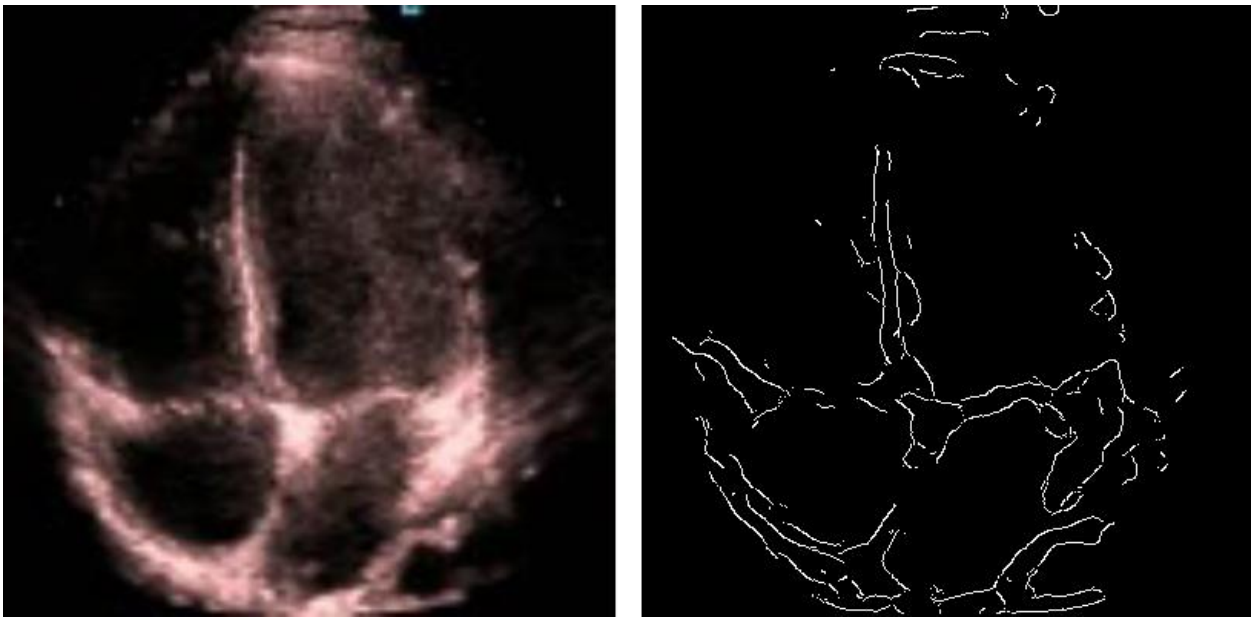


Figure 23: Original Frame (left) vs Canny Edge of the grayscaled and Gaussian blurred (right)

This edge detection method not only shows the structural lines within the images but also creates the base for more analyses, such as finding contours or attempting to segment regions.

Incorporating Volumetric Tracings for Enhanced Analysis

Navigating further through the complexities of the EchoNet-Pediatric dataset, the incorporation of the volumetric tracings enabled a more integrated and functionally insightful exploration. Volumetric tracings are provided in VolumeTracings.csv. Specifically, the file contains, for each video, a pair of frames and coordinates for each frame denoting the polygon vertices of the areas of the left ventricle during diastole and systole. These were implemented as Regions of Interest (ROI) – Contours displayable in the graphic user interface (GUI).

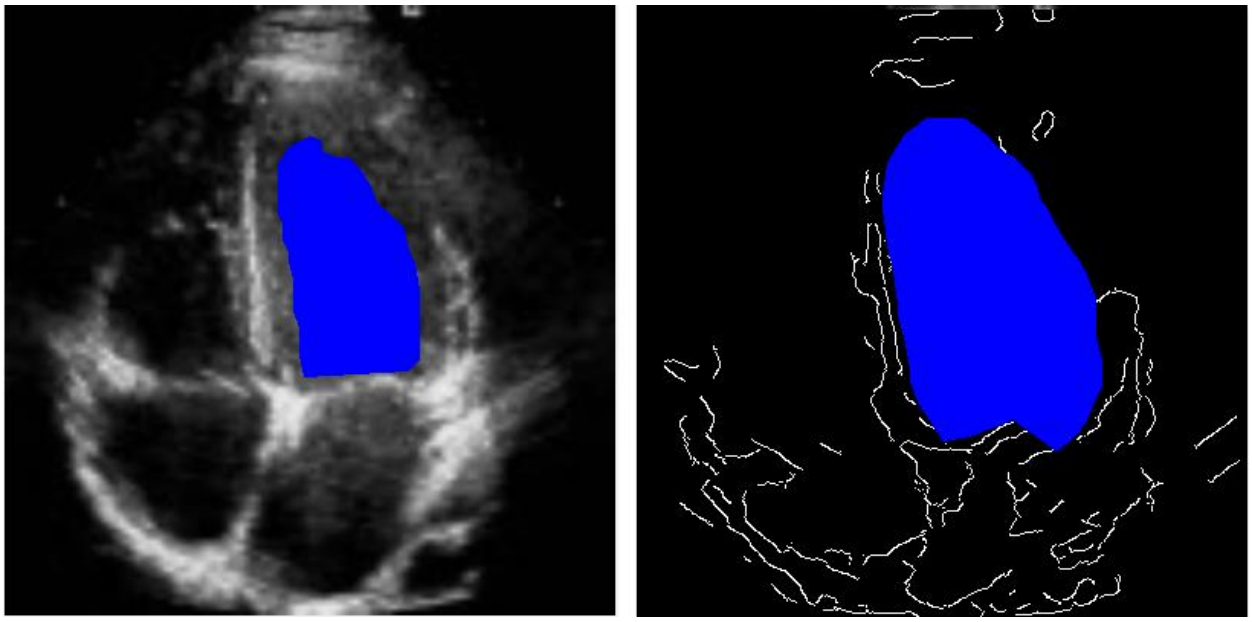


Figure 24: End Systole on grayscaled Frame (left) and End Diastole after Canny Edge

A major aspect of this process was the adjustment of contour coordinates in accordance with the frame resizing, ensuring their consistent and accurate overlay on the echocardiographic images. It is worth noting that the original frame dimensions were retained to enable precise scaling and adaptation of the contour points to the resized frames. Furthermore, as for the visualization process it was important to ensure that the contours were visually apparent and did not distort the underlying echocardiographic data. Thus, the contours were drawn in a distinctive blue color, providing a clear separation from the grayscale frames, thereby ensuring optimal visual clarity.

Consequently, the interface allowed for dynamic interaction with the visual data. The user has the option to toggle the display of contours on and off, facilitated by a designated button on the GUI. This button was constructed to be informative, changing color based on its state (green when active, red when inactive), providing a clear visual indicator to the user regarding the status of the contour display. Moreover, a checkbox within the GUI provided an additional layer of user interaction, enabling the selective display of contours based on user preference.

In essence, the integration and dynamic visualization of contours, meticulously synchronized with the echocardiographic data, provided a comprehensive and insightful analytical platform, enabling a deeper understanding of the diastolic and systolic volumes of the left ventricle.

Segmenting the left ventricle

The early steps into processing and analyzing the EchoNet-Pediatric dataset provided both insights and challenges, especially when attempting to segment the left ventricle. Even though the methodology, which was leaned on traditional image processing techniques using OpenCV, demonstrated success in preprocessing individual frames, its application across subsequent frames was less effective. The noise in the echocardiographic videos and the small, quick movements of the heart muscle made it hard to analyze and segment the left ventricle in the frames, in spite of the meticulous application of filters and edge detection. Recognizing these challenges, a second attempt was made leveraging more sophisticated OpenCV operations.

Video Preprocessing Part 2 - Engaging with Different Views and Advanced Techniques

Introduction

In the previous section, the journey through EchoNet-Pediatric dataset was focused on the Apical 4-Chamber (A4C) view, utilizing traditional image processing techniques. Moving forward, the journey continues by widening the scope to include the Parasternal Short Axis (PSAX) videos while delving deeper into advanced OpenCV techniques and at the same time incorporating the whole EchoNet-Pediatric dataset. While the Apical 4-Chamber (A4C) view displays all four cardiac chambers, PSAX view offers a vertical perspective to the A4C view focused on the mitral valve. Thus, by incorporating both angles the visual inspection through the different echocardiographic angles of the heart is provided integrating the complete extension of the dataset.

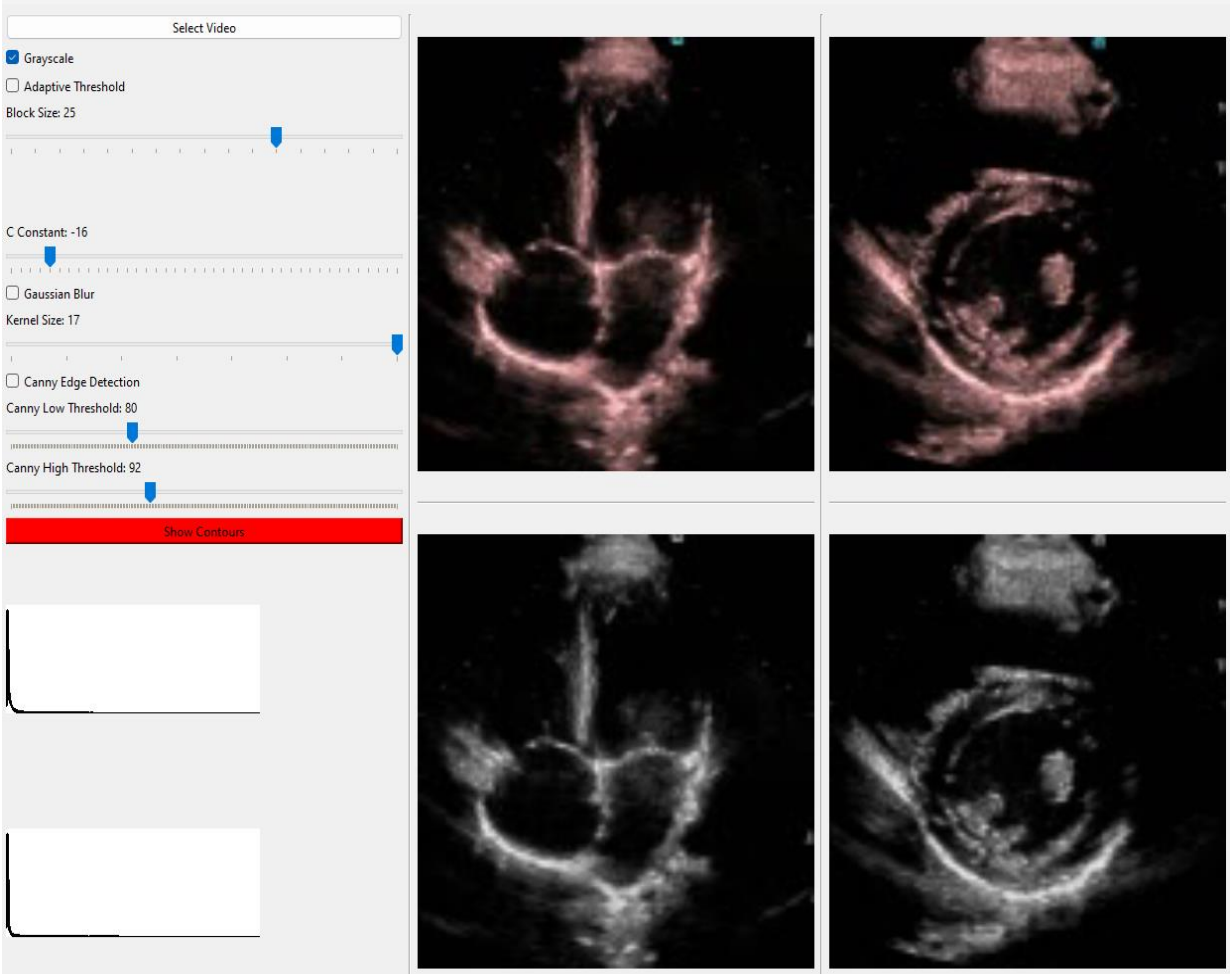


Figure 25: Second Graphical User Interface with both A4C and PSAX vies

The new graphical user interface (GUI)

The new graphical user interface (GUI) consists of three layouts. The left, the middle and the right. The primary contains the user options and the histograms of both views, the middle displays the Apical 4-Chamber (A4C) view of both the original and the filtered frame, while the last the Parasternal Short Axis (PSAX) respectively.

OpenCV Techniques in Focus: Exploring Beyond Basics.

Based on the challenges of the basic preprocessing and visualization step a new extended program was created aiming to dive into more advanced OpenCV techniques while at the same time trying to maintain a minimalistic approach to available filters in order to ensure consistent application

across multiple frames and videos. This philosophy aligns eloquently with Leonardo da Vinci's saying: "Simplicity is the ultimate sophistication."

Adaptive Thresholding and Gaussian Blurring: Highlighting Essential Details while Denoising Preserving Structures

Firstly, adaptive thresholding is implemented with OpenCV, which involves determining the threshold for a pixel based on a small region around it. Thresholding is identical to deciding a rule: "if a part of a frame is brighter than a certain level, color it white; if it's darker, color it black." This allows to display shapes more clearly by turning the frame into a simple black and white image where significant shapes stand out. But, with just one rule for the whole frame, it might be possible to lose details in the too-dark or too-bright parts. That is where "adaptive thresholding" comes into play. Essentially it introduces incorporation of different brightness rules for different parts of the frame by exploring small regions of the frame deciding a suitable rule for each one, thus helping to keep important details. This way, even if a part of the frame is in shadow or is too bright it preserves essential details.

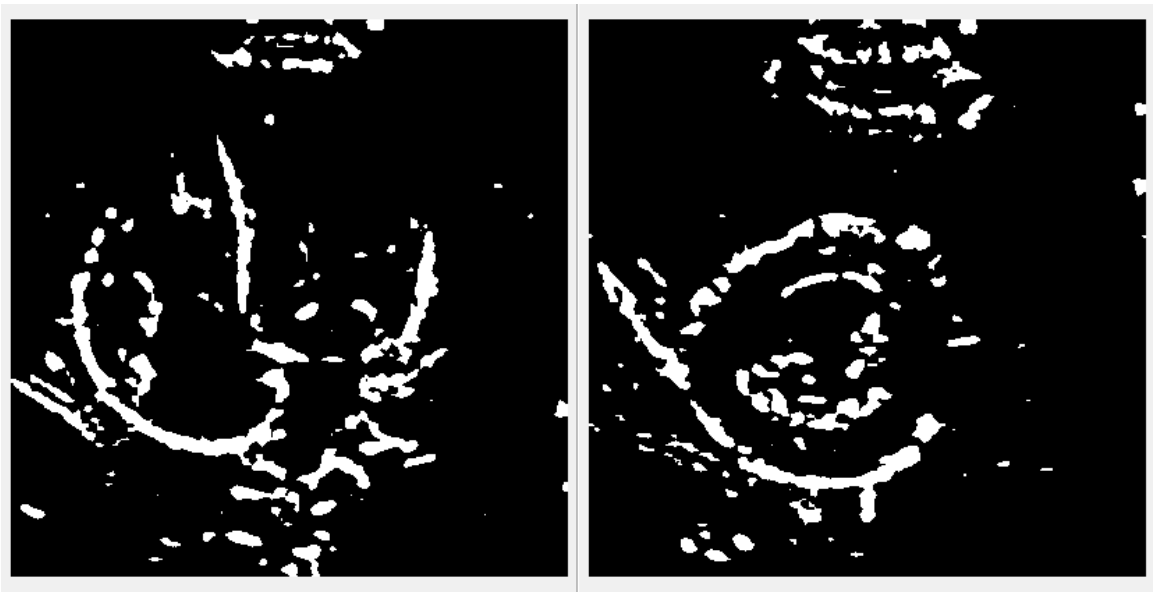


Figure 26: Adaptive Thresholding in both A4C and PSAX views

In the context of echocardiographic videos, where light and shadows may vary considerably across different regions, adaptive thresholding preserves important details that a global threshold might obscure. Adaptive Thresholding coupled with Gaussian blur, which is used to smoothen the image by reducing the characteristic ultrasound noise, proved especially beneficial. Gaussian blurring in

contrast to other blurring methods was chosen since it tends to preserve edge structures, which is crucial when analyzing anatomical structures such as the chambers and valves of the heart.

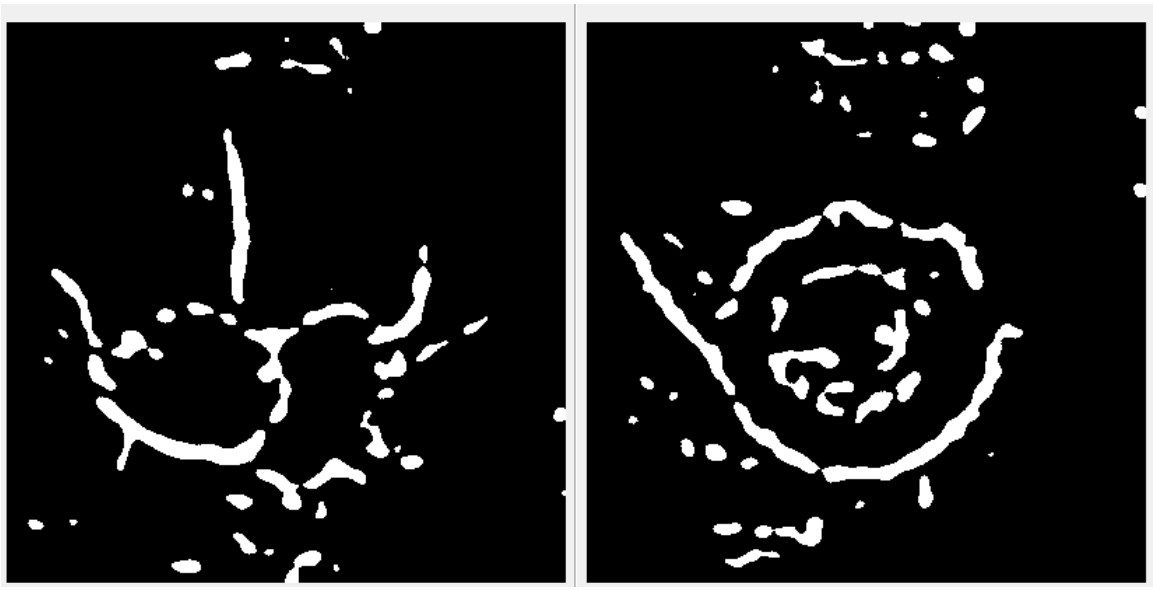


Figure 27: Gaussian Blurring and Adaptive Thresholding in both A4C and PSAX views

Canny Edge Detection: Identifying Boundaries

The use of the Canny Edge Detection algorithm was kept due to its ability to detect edges even under noisy conditions such those in echocardiographic frames that might be helpful in facilitating the segmentation of the left ventricle.

Embracing Simplicity Revealed New Challenges.

By strategically selecting advanced techniques and limiting the tuning parameters, the methodology not only became computationally efficient but also more robust and applicable to the rather complex EchoNet-Pediatric dataset. However, it was during this exploration that the data set , particularly the Apical 4-Chamber (A4C) view, revealed its additional complexities.

Through trial and error across multiple videos, it was unveiled that the Apical 4-Chamber (A4C) sub-dataset is not as homogenous as initially anticipated. Encountering multiple views, with various rotations, some displaying only the left ventricle, others being excessively noisy, and a few featuring significant elements added by the ultrasound machine, introduced the necessity of clustering the dataset. This is where machine learning principles were integrated to manage these discrepancies, aiming to divide the dataset into homogeneous parts.

Machine Learning: Clustering for Homogeneous Analysis

I. Feature Extraction from Apical 4-Chamber (A4C) and Demographics Datasets

The initial phase of the analysis aimed at meticulously extracting appropriate features from the two principal datasets, the Apical 4-Chamber (A4C) videos and the demographic data. This thorough extraction was pivotal for the analysis and clustering of the dataset into an appropriate number of sub-datasets that would be homogenous enough for the subsequent task of potentially segmenting the left ventricle from the cardiac ultrasound videos using OpenCV. The extracted features not only include the critical visual aspects of the ultrasound data but also the crucial demographic details, attempting to incorporate relevant futures across the dataset.

1. Computation of the Average Image

Through OpenCV library, each frame of the video was converted into a grayscale format to minimize the computational complexity and emphasize the structural aspects of the images and each video in the dataset was processed to yield an average image computed by calculating the mean across all frames, which was subsequently used as a representative snapshot of the video for further analysis.

2. Extraction of Features via GLCM and Histogram Techniques

Following the acquisition of the average image, the feature extraction process was executed by employing the Gray Level Co-Occurrence Matrix (GLCM) and histogram techniques. The GLCM **(scikit-image, n.d.)** is a statistical method that allowed the extraction of texture features that consider the spatial relationship of pixels, thus allowing to obtain features such as contrast, dissimilarity, homogeneity, Angular Second Moment (ASM), energy, and correlation. It is important to note that there has been substantial research utilizing GLCM texture analysis in ultrasound data (**Yang X, 2012**) In addition, higher-order statistical features commonly used in image analyses such as entropy (a measure of randomness or disorder), energy, and smoothness were derived to provide a more in-depth understanding of the image textures. (The first-order involves the mean, median and mode, while the second-order the variance and standard deviation). Simultaneously, histogram features were extracted which provided a statistical representation of the pixel intensity distribution of the images.

3. Integration with Demographic Data

After the feature extraction, the derived features were merged with the existing demographic data, resulting in a comprehensive dataset that also includes vital clinical and demographic details. This mixture of data ensured a rich dataset, providing a robust foundation for the following phases of analysis and clustering. This dataset was stored in a structured format excel file **(Appendix D)**, serving as the needed input for the following phases of Principal Component Analysis (PCA) and clustering.

II. Dimensionality Reduction and Clustering of the Feature Space

The feature dataset was intentionally high-dimensional, including multiple factors that characterize each video, in order to enable more accurate clustering through the process of dimensionality reduction. This process is suitable for focusing on the most important information, facilitating easier data grouping into clusters while also allowing more interpretability of the results.

1. Principal Component Analysis (PCA)

Principal Component Analysis PCA **(scikit-learn, n.d.)** was employed as a mathematical technique to reduce the feature space, with a meticulous emphasis on retaining maximal variance and thus, preserving the essence of the original data. This technique transforms the original features into a new set of uncorrelated variables, called principal components, which are linear combinations of the original variables. It is important to note that after importing the file it was crucial to scale the features using the Standard Scaler **(scikit-learn, n.d.)** ensuring that all of them contributed equally to the Principal Component Analysis (PCA).

2. Analysis of Explained Variance

A thorough investigation into the explained variance was executed to determine the optimal number of principal components to be retained for subsequent analyses. The analysis included evaluating the percentage of the total variance of each principal component “explained”. Then, a plot of the cumulative explained variance against the number of components was utilized to visually determine a suitable cutoff.

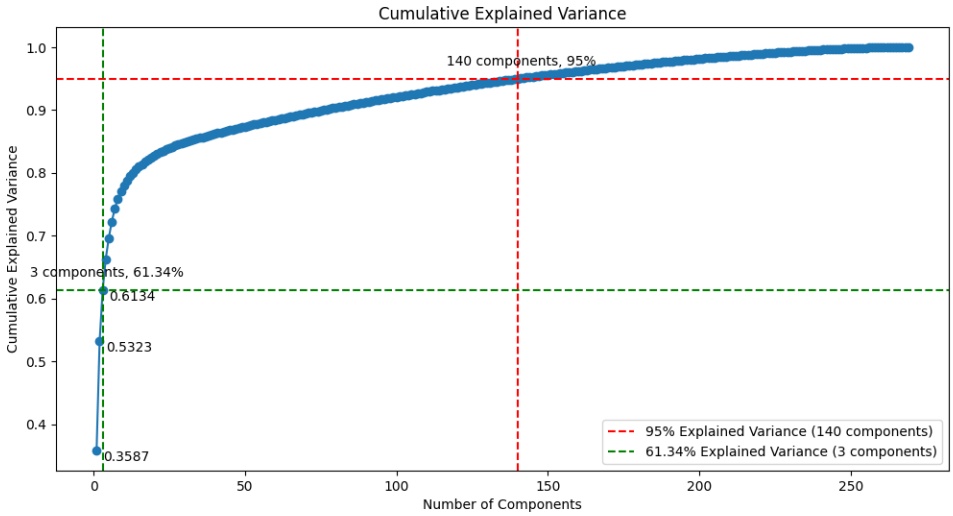


Figure 28: Cumulative Explained Variance with infographics to determine the optimal number of components

The cumulative explained variance plot revealed a crucial insight into the proportion of total variance encapsulated as the number of principal components increased. Particularly, the first three principal components encapsulated approximately 61.34% of the total variance in the dataset, suggesting a significant reduction from the original 277-feature space. The decision to select n=3 principal components was made by several strategic considerations. Firstly, choosing three components allowed an intuitive and interpretable graphical representation of the clusters. Secondly, the choice ensures computational efficiency and model simplicity in subsequent analyses, providing a careful symmetry between reducing dimensionality while still maintaining a substantial amount of the original information.

To further explore the feature importance in the three principal components, the following plot was employed.

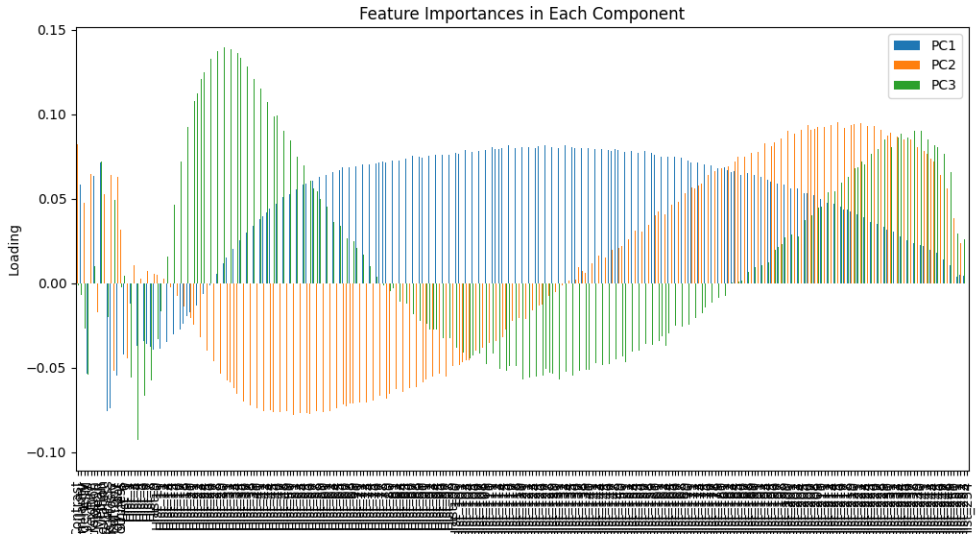


Figure 29: Feature Importance in Each Components to draw insights related to the initial features.

Each bar represents a feature from the original dataset, and its height signifies the loading of that feature in the respective principal component (PC1, PC2, and PC3). The loading essentially shows how much a feature contributes to that component. Positive and negative values suggest the direction of the correlation between the original feature and the principal component.

The following features emerge as pivotal across the first three principal components:

PC1 (Principal Component 1): The dominant contributors include "Mean", "Std Deviation", "Smoothness", along with the histogram features "Hist_133" and "Hist_127".

PC2 (Principal Component 2): This component is mostly influenced by several histogram features: "Hist_215", "Hist_221", "Hist_216", "Hist_222", and "Hist_212".

PC3 (Principal Component 3): The leading contributors are histogram features: "Hist_30", "Hist_31", "Hist_33", "Hist_32", and "Hist_29".

These insights suggest that the first principal component (PC1) encapsulates general statistical properties of the images, while the second and third principal components (PC2 and PC3) capture variations in the pixel intensity distributions from different parts of the histogram.

Upon reduction of the feature space, the aim transitioned towards partitioning the dimensionally reduced feature space into coherent clusters, each representing distinct variations within the A4C video data.

1. Determining the Optimal Number of Clusters.

To decide on the optimal number of clusters for the clustering algorithms, both the elbow method and the dendrogram analysis were employed.

Elbow Method **(Analytics Vidhya, 2021)**: A technique used to identify the optimal number of clusters by plotting the inertia of different numbers of clusters. The inertia is the sum of squared distances of samples from their closest cluster center. The point where the decrease in inertia begins to slow down ("elbow" point) is considered an indicator of the optimal number of clusters.

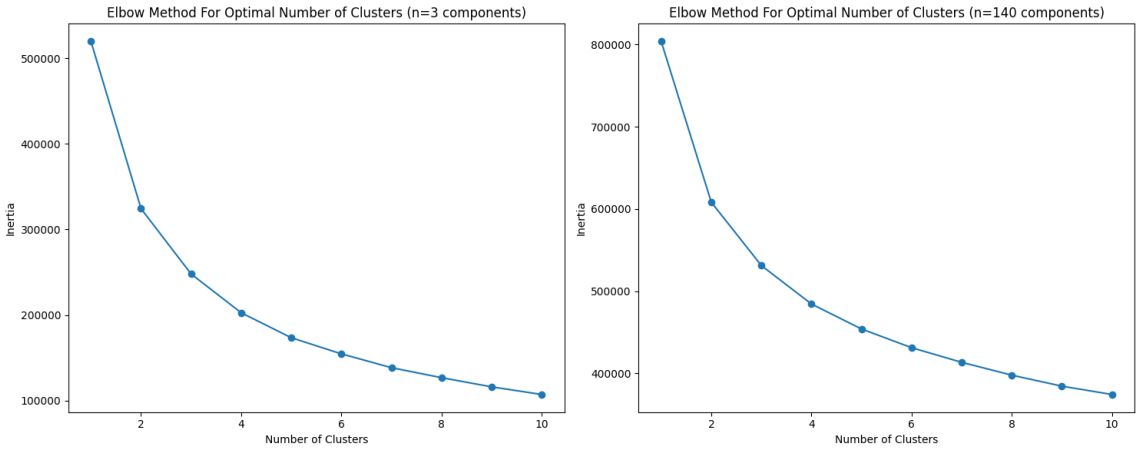


Figure 30: Elbow Method for PCA for both n=3 and n=140 components to determine the optimal number of clusters.

Dendrogram Analysis **(SciPy, n.d.)**.: Dendrograms are tree-like diagrams that visually represent the results of hierarchical clustering. They display how data points are grouped together step-by-step in a tree structured way. By examining the lengths of the branches, which indicate the dissimilarities or distances between clusters, one can decide on the optimal number of clusters. Typically, a longer branch length indicates a larger dissimilarity, and cutting the dendrogram at a specific height allows for the determination of distinct clusters.

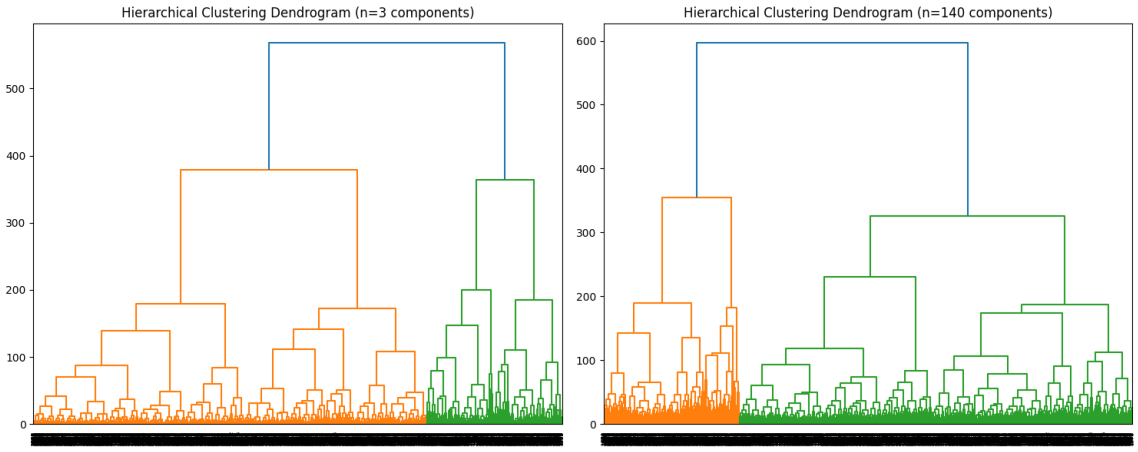


Figure 31: Dendrograms to confirm the optimal number of clusters

Notably, the analyses via both the elbow method and dendrogram were performed on the data in two different PCA component spaces (3 and 140) in order to ensure the number of clusters. Remarkably, both indicated that the optimal number of clusters is four. This was particularly important and showed that despite the great difference in dimensionality, the clustering of the data remained consistent, showcasing the robust clustering within the data.

2. K-Means Clustering

The K-Means clustering **(scikit-learn, n.d.)** algorithm was utilized. This method breaks down data into ‘k’ groups, or 'clusters', based on their similarity. The center of each cluster is found by averaging all its data points. The goal is to make sure data points in the same cluster are close to each other. The algorithm keeps adjusting the groups and their centers until the best arrangement is found.

3. Agglomerative Clustering

Agglomerative clustering **(towardsdatascience.com, 2019)**, a hierarchical clustering technique, was also employed. This method starts by treating each data point as a single cluster and then gradually combines them based on how similar they are. The similarity is often measured by how close the points or clusters are to each other. As the points come together, a tree-like diagram called a dendrogram is created.

4. Bisecting K-Means Clustering

Bisecting K-Means **(*medium.com*, 2020)** clustering was employed as an additional clustering strategy, which recursively partitions the data into two clusters at each level until a specified number of clusters is reached. This method works by continually splitting data into two parts, while at each step ensuring that the most similar data points stay together. It keeps going until it reaches the number of groups needed. This method offers a balance between the simplicity of K-Means and the hierarchical detail of agglomerative clustering.

5. Bisecting K-Means Clustering with Manhattan Distance

A variant of the bisecting K-Means algorithm was also explored, where instead of the typical Euclidean distance, the Manhattan **(Science Direct, n.d.)** distance metric was employed during the clustering process. This distance metric calculates the absolute differences between the points, potentially offering a different perspective on data similarity and cluster formation.

C. Comparative Analysis

To determine the best clustering algorithm, a meticulous comparative analysis of the clustering methods previously presented was performed by examining the formed clusters of the original A4C median frame with the objective to reveal any patterns or associations that might be helpful for the left ventricle segmentation task. Both validation tools and visual representations of their structure were utilized in an attempt to have both mathematical and visual criteria.

1. Silhouette Analysis

Silhouette Analysis **(scikit-learn, n.d.)** was used to evaluate the quality of the clusters that provides a metric that quantifies the efficiency of the clustering results. The silhouette score, ranging from -1 to +1, offers a measure of how similar an observation is to its own cluster compared to other clusters by computing the mean inter-cluster distance and the mean intra-cluster distance, thus validating if the formed clusters are distinct and appropriate. The silhouette plots below display a measure of how tightly grouped the samples in the cluster are for each clustering algorithm.

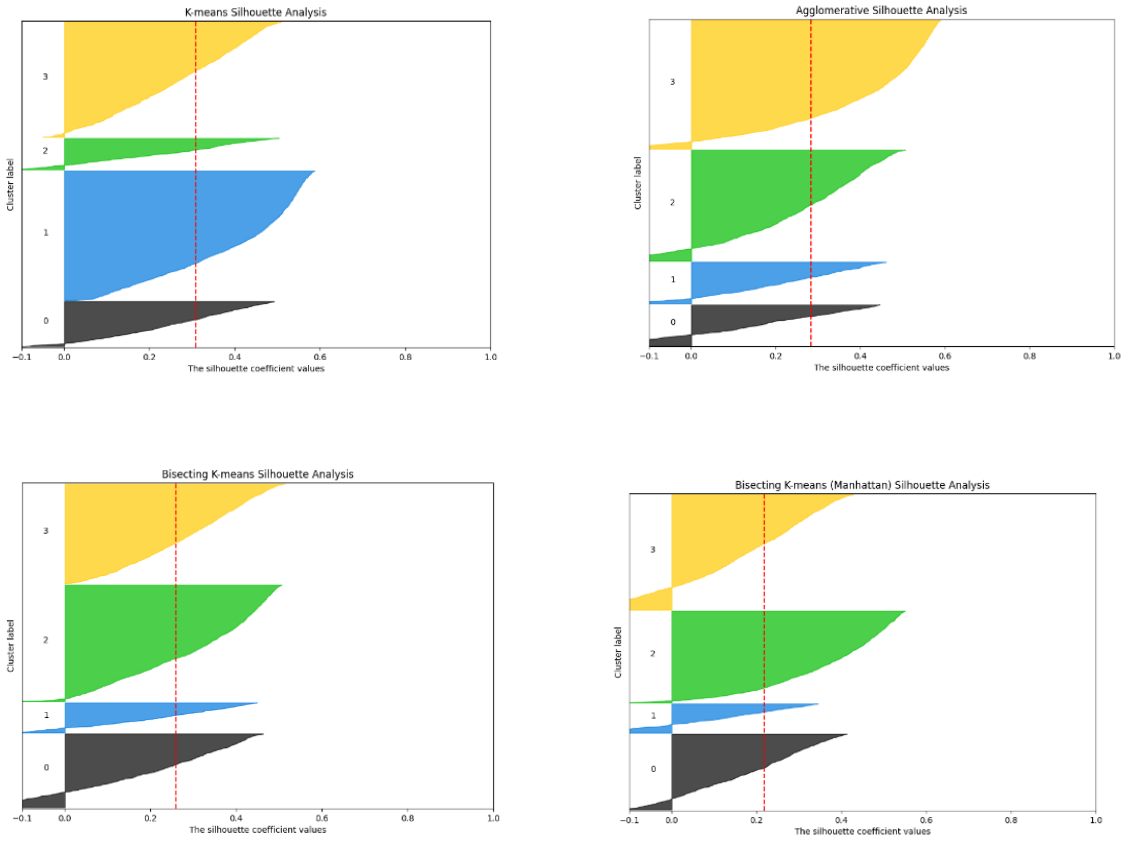


Figure 32: Silhouette Plot Analysis for all four methods

From the above graphs it is apparent that the k-means clustering algorithm produces a more accurate result since fewer data points have silhouette score < 1. To further prove the point the average silhouette score for each method was computed confirming the visual inspection. (KMeans: 0.31, Agglomerative: 0.28, Bisecting K-means: 0.26, Bisecting K-means Manhattan: 0.22)

2. 2D and 3D Visualization

The dataset was visualized using both 2D and 3D renderings of the principal components, allowing an insightful exploration of the created clusters within the data. The 2D visualizations display the basic groupings and separations, providing a straightforward view of the main patterns. On the other hand, the 3D visualization improves the process by allowing exploration of additional patterns and relationships among the clusters, since it provides a space view of how the data points

relate to each other. These visualizations served as a comprehensive tool in order to understand the underlying structures of the clusters within the data.

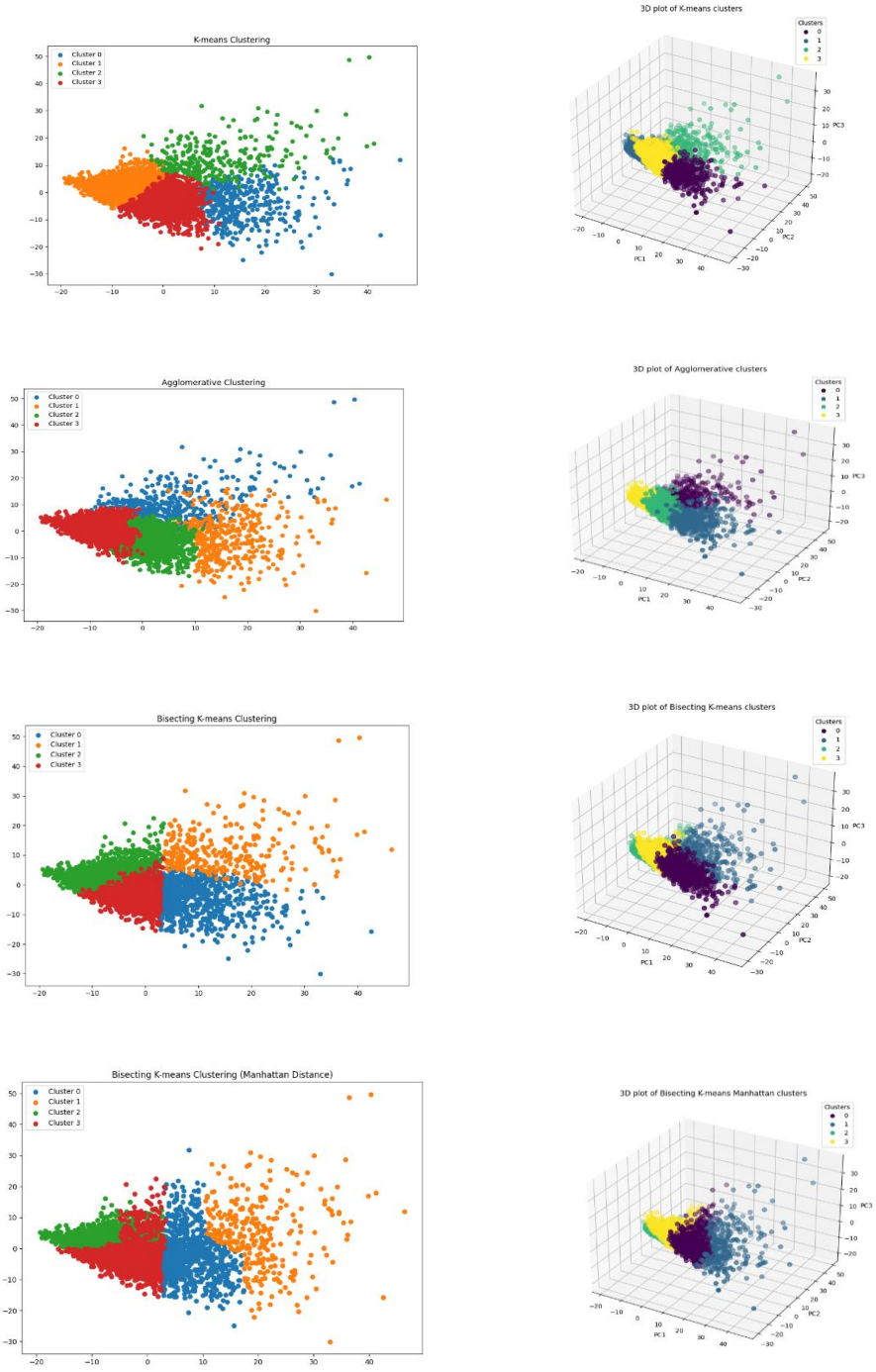


Figure 33: Visualization of the Clusters in both 2D and 3D plots

From both the 2D and 3D visualizations it is evident that the k-means clustering created distinct, non-overlapping clusters. This means that the features used for clustering have enough variance between the groups to create clear boundaries when segmenting the dataset into clusters.

3. Cluster Size Analysis

An examination into the size and data density of each cluster was conducted as a preliminary approach to the following analysis through the clusters created by each method.

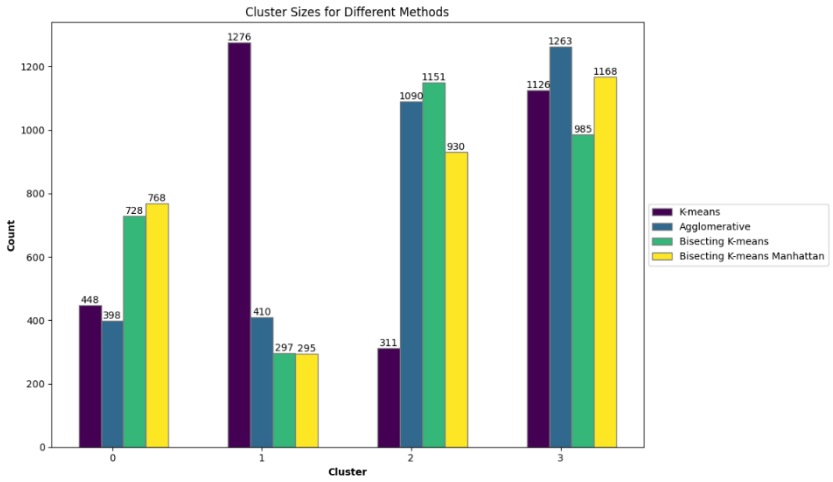


Figure 34: Video Distribution in the clusters per method

IV. Insights from the clusters

With the clusters validated and visualized, the objective transitioned towards deriving meaningful insights from the clustered data while understanding the patterns between the data points in each cluster and the differences among the different clusters. Below there are samples and insights for each cluster.

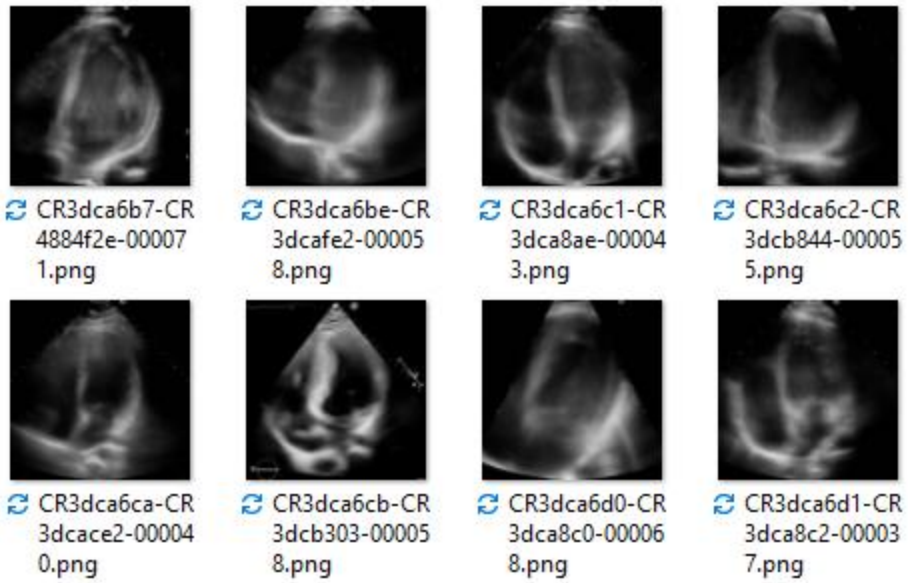


Figure 35: First Cluster Sample

The first cluster consists of echocardiographic data with a clear view of all the four cardiac chambers.

Figure 36: Second Cluster Sample

The second cluster has echocardiographic data that the left ventricle is vertically depicted.

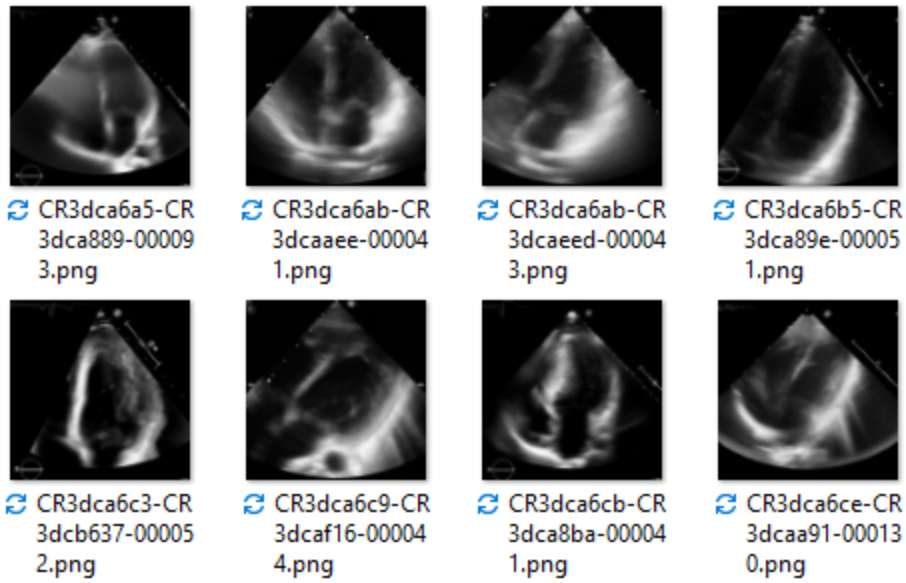


Figure 37: Third Cluster Sample

The third cluster consists of mainly noisy echocardiographic data.

Figure 38: Fourth Cluster Sample

The fourth and last cluster contains data that is focused in mostly the left ventricle while the other chambers are partially seen.

Conclusion - Left Ventricle segmentation via OpenCV

Even though the k-means algorithm and the intricate journey through feature extraction, dimensionality reduction, cluster analysis and the meaningful partition into four clusters. Even though OpenCV provides a myriad of functionalities and initially seemed promising in refining processing for a particular frame, it failed dramatically to generalize across this diverse dataset.

After multiple attempts to make the most of OpenCV and employing rather complicated techniques such as Hough Lines and Hough Ellipsis **(scikit-image. n.d)** the task of segmenting the Left Ventricle using traditional techniques was an unsuccessful journey that opened the path for researching alternative and more advanced methodologies. However, it was through OpeCV that this complex dataset was possible to visualize and get familiar with the complexities and intricacies of the echocardiographic data.

Recognizing the limitations of the initial approach was crucial, especially when it comes to traditional methods like those offered by OpenCV. The journey led to exploring deep learning methodologies, which offer potential solutions to handle the complex and high-dimensional echocardiographic data that posed significant challenges to usage of OpenCV. This shift in methodology was not a change in direction, but rather a development exploring deep learning in more depth and continue the exploration into the echocardiographic data with a fresh perspective and a fine-tuned approach.

Deep Learning Part 1 – Left Ventricle Segmentation

In this part of the study, considering the difficulties that arose with traditional methods, the aim was to automate the segmentation of the left ventricle in the EchoNet-Pediatric dataset utilizing deep learning. The methodology involved a systematic approach starting from data pre-processing, model design, and training, through to model evaluation and application to new data.

Preparing the Dataset

The dataset for this part consists of the Apical 4-Chamber (A4C) view videos and their corresponding volume tracings stored in the VolumeTracings.csv file. Each tracing is essentially a set of coordinates that defines a polygon, which in turn represents the region of interest in a given frame of a video.

Extracting Polygons for Videos

For each video, the goal was to extract a pair of polygons from which one depicts the left ventricular systole while the other the diastole. This is achieved by filtering the tracings data based on the filename of the video and subsequently grouping the data by frame number. For each frame, a polygon is then created using the coordinate values from the tracings data. This process resulted in a dictionary where each key is a frame number, and the corresponding value is the polygon (set of coordinates) for that frame.

Generating Masks from Polygons

Once we have the polygons for a video, the next step was to convert these polygons into binary masks. A mask is essentially a 2D array with dimensions matching the video frame, where pixels inside the polygon are set to 255 (white) and pixels outside the polygon are set to 0 (black). The mask is generated using the OpenCV library, which fills the area defined by the polygon coordinates and will serve as the ground truth during the model training phase.

Extracting Frames and Corresponding Masks

Each video in the dataset was read frame by frame using the OpenCV library. Each frame was converted to grayscale not only to reduce the computational load but also to emphasize the structural details in the echocardiographic frames, which are crucial for the segmentation task. Grayscale images retain the structural information while discarding the chromatic details. Continuing, for each frame for which we have a corresponding polygon in the tracings data, a mask is generated using the previously described method. These frames, along with their corresponding masks, are stored to be the input of the training model. It is important to note that the preprocessing procedure extracts and processes only those frames for which there is a corresponding polygon and frames without associated polygons are not used in the subsequent steps.

Division of Data into Training and Test Subsets

The dataset, consisting of frames and their corresponding masks, is divided into training and test subsets ensuring both subsets are representative of the overall distribution of data. Specifically, 80% of the data is utilized for training the model, allowing it to learn the underlying patterns, while

the remaining 20% is reserved for testing, enabling an unbiased evaluation of the model's predictive performance on unseen data.

U-Net Model Architecture

The U-Net model, developed by Olaf **Ronneberger et al. (2015)** for biomedical image segmentation, employs a fully convolutional neural network structure renowned for its efficiency in tasks where the input and output dimensions are equal. The U-Net model used in this study starts with 32 filters and progressively doubles the number of filters in the encoding path (downward slope) until it reaches 512. Then, it symmetrically reduces the number of filters in the decoding path (upward slope) back to 32.

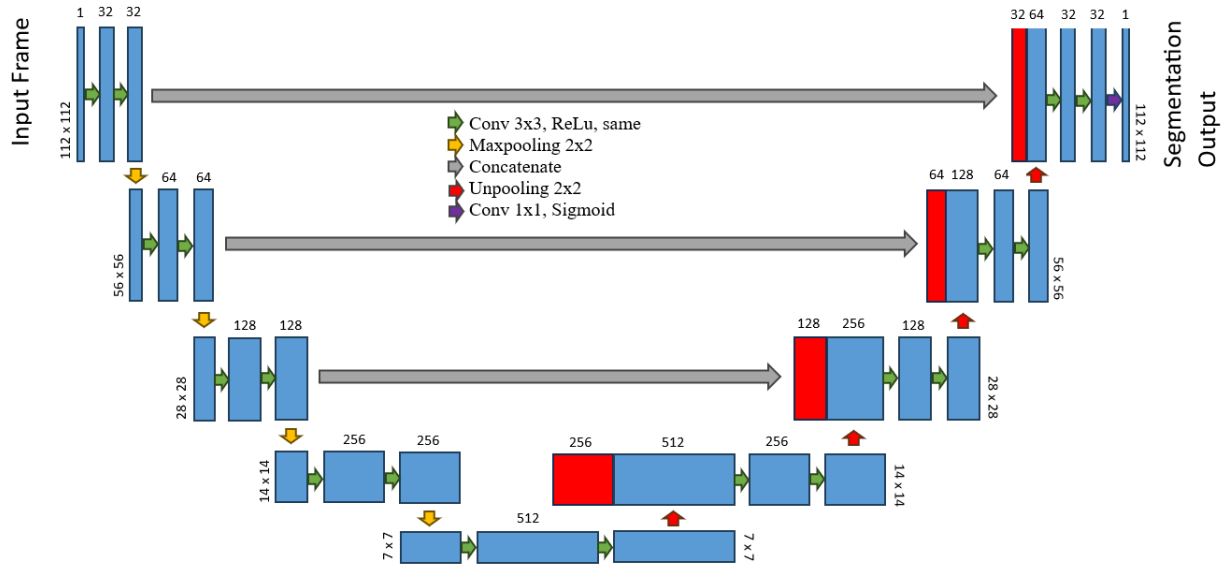


Figure 39: Left Ventricle Segmentation U-Net Model

Architecture Overview

The U-Net architecture is symmetric and consists of two paths:

Input: Two primary inputs – Initiating the downward slope

The U-Net model takes two primary inputs: the original image frame and its corresponding mask. The mask in this context is a binary image that highlights the regions of interest – the left ventricle.

The combination of original images and masks enables the model to learn relevant features and ignore irrelevant areas.

Encoder: Extracting Image Features

Layer 1: The model starts with 32 filters. It applies two consecutive convolutional layers, each followed by a ReLU activation function, to capture the underlying patterns in the image, such as edges, textures, and more complex patterns.

Layer 2 to 4: The number of filters doubles with each subsequent layer in the encoder (64, 128, 256, and finally 512). The image's spatial dimensions reduce (due to max-pooling), but the feature dimensions increase, capturing more complex patterns while reducing computational complexity.

Bottleneck: Most Compressed Information - Initiating the upward slope

Layer 5: In the middle of the network (the bottom of the "U"), the spatial dimensions are smallest, and the feature dimensions are at their maximum (512). This layer focuses on encoding the most critical features of the input images.

Decoder: Reconstructing and Segmenting the Image

Layer 6 to 9: The decoder expands the feature information back into the original image size. It gradually increases the spatial dimensions and decreases the feature dimensions (from 512 back down to 32). During this phase, the model also leverages the feature information from the encoder using concatenation to ensure precise localization of the segmented areas.

Final Layer: Creating the Segmented Output

Layer 10: The final layer uses a 1x1 convolution to map the 32-channel feature maps to the desired output size. It outputs a single channel, representing the segmented areas of interest (the left ventricle) within each image.

The implementation maintains the general U-Net structure, ensuring a symmetry between the encoder and decoder paths. The encoder is constructed using repeated applications of Conv2D and MaxPooling2D layers, while the decoder utilizes UpSampling2D layers to recover the spatial dimensions of the input. The final layer applies a 1x1 convolution to map the decoder's output to the desired segmentation mask.

Loss Function

To train the U-Net model for segmentation, a combined loss function was utilized that encompasses:

Dice coefficient

A popular metric for evaluating the performance of image segmentation algorithms, the dice coefficient is a measure of how similar two objects are, in this case, the predicted segmentation and the ground truth (masks), by calculating the size of overlap.

$$Dice\ Coefficient = \frac{2 \times |A \cap B|}{|A| + |B|}$$

It is worth noting that the Dice score is not only a measure of how many positives you find, but it also penalizes for the false positives that the method finds, similar to precision, thus ensuring that the model is penalized for spatial inconsistencies.

Dice Loss

The Dice loss is computed as one minus the Dice coefficient, transforming the similarity metric into a loss function to minimize.

$$Dice\ Loss = 1 - Dice\ Coefficient$$

Binary Cross-Entropy (BCE) Loss: A standard loss for binary classification problems, penalizing pixel-wise classification errors.

Compound Loss

Taghanaki et al. **(Taghaniki et al.,2019)** who researched multi-organ segmentation in medical imaging and addressed the issue of imbalanced data where the region of interest occupies a significantly smaller portion of the image compared to the background, especially in the context of this study at the systolic cardiac phase. They suggest that the combo loss function, that combines the Dice similarity coefficient with cross-entropy to penalize false positives and negatives performs exceptionally well at medical image segmentation.

Training Procedure

The model was compiled and trained using the Adam optimizer, a widely used optimizer in deep learning applications due to its adaptive learning rate properties. This ensures that while the model learns to minimize the loss function, it adapts its updates to each parameter based on the historical gradients, thus enabling the model to learn more steadily and possibly faster.

The model was trained using a batch size of 16 and for 10 epochs. During each epoch, the model processes batches of 16 images and their corresponding masks, updating its weights to minimize the loss on these batches across the training dataset. The validation data (X_test, y_test) are utilized to evaluate the model on unseen data during training, providing a snapshot of its performance and generalization capability.

```
Epoch 1/10
322/322 [==============================] - 814s 3s/step - loss: 0.4292 - accuracy: 0.9486 - val_loss: 0.2052 - val_accuracy: 0.9735
Epoch 2/10
322/322 [==============================] - 806s 3s/step - loss: 0.2031 - accuracy: 0.9738 - val_loss: 0.2015 - val_accuracy: 0.9748
Epoch 3/10
322/322 [==============================] - 807s 3s/step - loss: 0.1878 - accuracy: 0.9757 - val_loss: 0.1792 - val_accuracy: 0.9760
Epoch 4/10
322/322 [==============================] - 820s 3s/step - loss: 0.1751 - accuracy: 0.9773 - val_loss: 0.1754 - val_accuracy: 0.9770
Epoch 5/10
322/322 [==============================] - 806s 3s/step - loss: 0.1677 - accuracy: 0.9782 - val_loss: 0.1661 - val_accuracy: 0.9780
Epoch 6/10
322/322 [==============================] - 814s 3s/step - loss: 0.1596 - accuracy: 0.9793 - val_loss: 0.1803 - val_accuracy: 0.9765
Epoch 7/10
322/322 [==============================] - 807s 3s/step - loss: 0.1554 - accuracy: 0.9798 - val_loss: 0.1582 - val_accuracy: 0.9793
Epoch 8/10
322/322 [==============================] - 805s 2s/step - loss: 0.1498 - accuracy: 0.9805 - val_loss: 0.1693 - val_accuracy: 0.9777
Epoch 9/10
322/322 [==============================] - 804s 2s/step - loss: 0.1474 - accuracy: 0.9809 - val_loss: 0.1609 - val_accuracy: 0.9791
Epoch 10/10
322/322 [==============================] - 804s 2s/step - loss: 0.1398 - accuracy: 0.9818 - val_loss: 0.1620 - val_accuracy: 0.9789
```

Figure 40: Training Process with loss and accuracy of U-Net model

Model Training Results

After training the model, the results from the final epoch yielded notable outcomes, which are indicative of the model's performance. Specifically:

- Loss: 0.14
- Accuracy: 0.98
- Validation Loss: 0.16
- Validation Accuracy: 0.98

Loss represents how well the model performed during training. A lower loss indicates better performance, with 0 being the ideal scenario.

Accuracy is a measure of how well the model makes predictions. In this context, the accuracy is 0.98, meaning the model correctly predicted the training data 98% of the time.

Validation loss and accuracy are metrics that show how the model performs on new, unseen data. In this case, the model has a validation loss of 0.16 and validation accuracy of 0.98, suggesting that it is also performing well on the validation data.

These results suggest that the model has learned to identify the patterns in the training data effectively and is also able to generalize its predictions to new, unseen data quite well. The high accuracy and relatively low loss in both training and validation phases indicate a well-fitted model.

Below are some of the results compared to the initial image and the original mask as denoted by the experts:

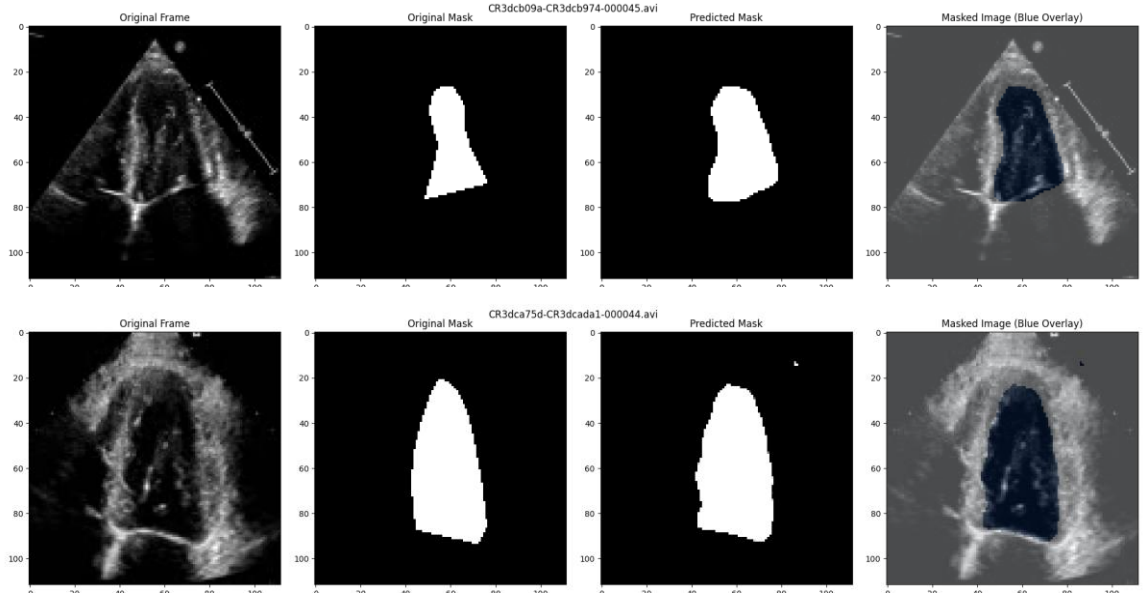


Figure 41: Examples of the Results of the Left Ventricle Segmentation Model compared to the original frame (left), the original mask (center left) the predicted mask (center right) and the final masked frame (right)

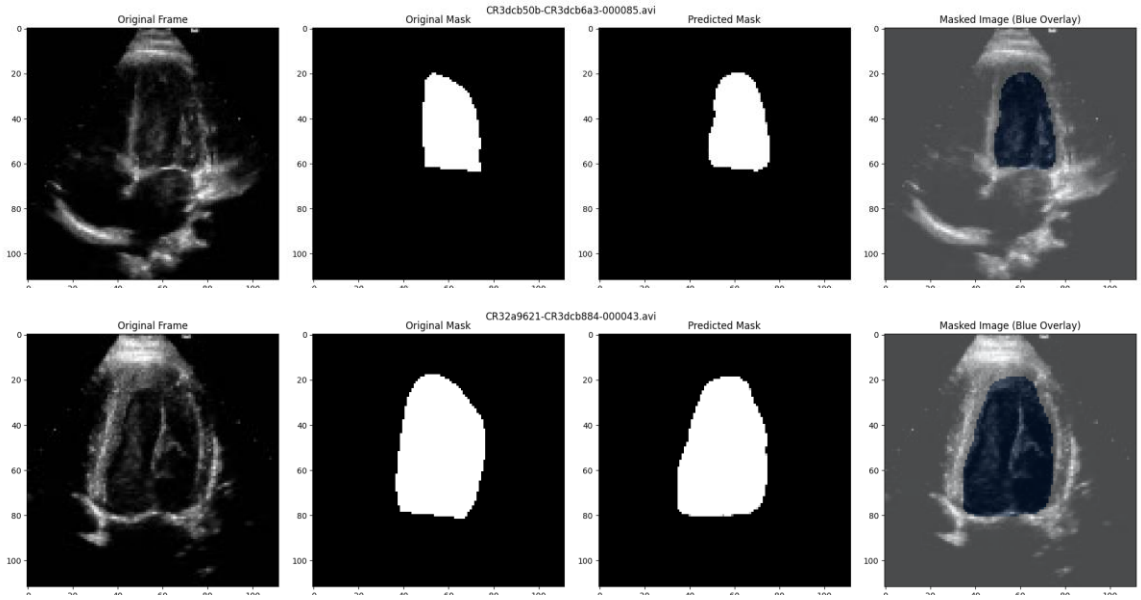


Figure 42: Examples of the Results of the Left Ventricle Segmentation Model compared to the original frame (left), the original mask (center left) the predicted mask (center right) and the final masked frame (right)

Conclusion

Through a thorough methodology surrounding by data preprocessing, model development, training, and evaluation while being in correlation with state-to-art studies regarding medical image segmentation, the study demonstrated a promising capacity for employing deep learning, specifically the U-Net model, in left ventricle segmentation in pediatric echocardiograms. Achieving an accuracy of 0.98 on both training and validation data, the model exhibited a profound ability to decipher patterns and generalize learning to unseen data. The visual results further support the numerical metrics, showcasing accurate segmentation in most cases.

Deep Learning Part 2 – Prediction of Left Ventricle Ejection Fraction

Introduction

This study was born out of a clear but challenging goal: to predict the left ventricle ejection fraction using the EchoNet-Pediatric dataset. The journey toward this objective was far from straightforward, involving numerous trials and errors and diving deeply into the complex task of left ventricle segmentation by exploring various approaches and methodologies to achieve it.

The quest to develop an automated segmentation solution through a deep learning approach was merely dependent on methodologies that have proven to be effective, however the prediction of

ejection fraction in pediatric echocardiographic video is not extensively explored due to the implications that come with the pediatric hearts and the lack of specific to this population dataset availability.

Addressing the complexity of video data, which demands substantial computational resources, and training a model that processes an input of multiple frames, posed its own set of substantial challenges.

This part of the study will walk you through not just the technical journey, but also the experiences and challenges encountered along the way. From the initial concept to the final solution, this is the journey:

Preliminary Steps – The road to Ithaca

In order to delve deeper into creating the ejection fraction model it was pivotal to determine the dataset to work on. The Apical 4-Chamber (A4C) view and their corresponding demographics, as well as the segmentation of the left ventricle model derived from using the Apical 4-Chamber (A4C) view along with the end systole and end diastole frames were utilized throughout this part of the study.

Building Upon research derived from adult's echocardiographic videos it was decided to use a 3D Two Stream Convolutional Network with input the first 16 frames of each video and their correlated Optical Flows which is the estimation of the motion of pixels between consecutive frames in a sequence of images. The below image depicts the result of the first try where the red line is the perfect fit line where all the data points should lie in the optimal scenario.

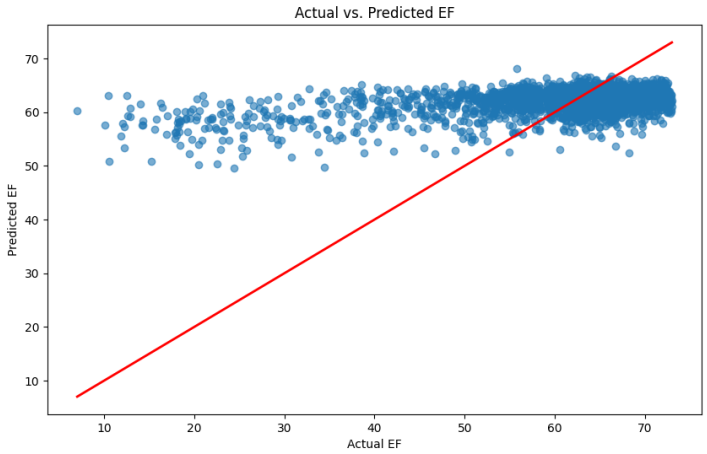


Figure 43: Predicted vs Actual EF (1st Attempt)

This first attempt had proven unsuccessful with validation loss (in terms of mean square error) reaching 150 steering the goal to alternative solutions.

The subsequent tactic was to integrate the demographics as an additional input to the model aiming to turn the data points towards the desired red line.

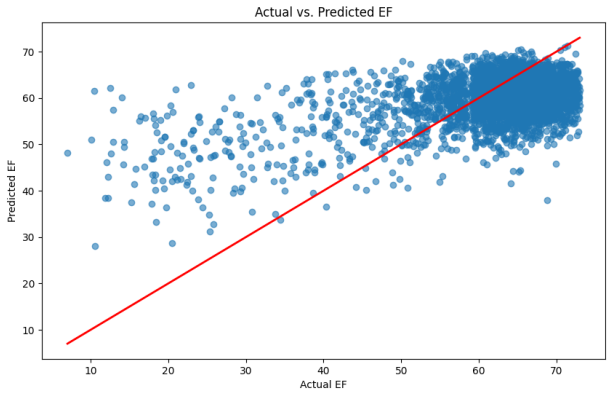


Figure 44: Predicted vs Actual EF (2nd Attempt)

Even though the preliminary analysis showed absence of apparent correlations between Ejection Fraction and demographic parameters such as age, weight, and height (the scatter matrix), this strategy provided some success, reducing the validation loss to 120, yet far ahead from the goal.

Integrating Left Ventricular Areas changes was the next addition, as an attempt to introduce the element of time into the model, which remarkably pushed the validation loss down to 90 and pushed the data points closer to the desired perfect fit line.

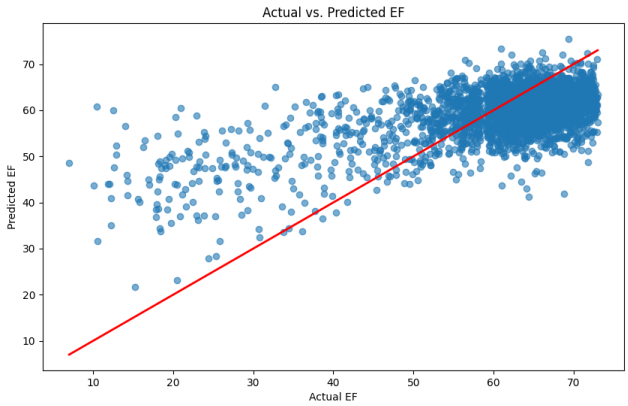


Figure 45: Predicted vs Actual EF (3rd Attempt)

Then hyperparameter tuning allowed the choice of appropriate hyperparameters that enabled reaching a validation loss down to 70.

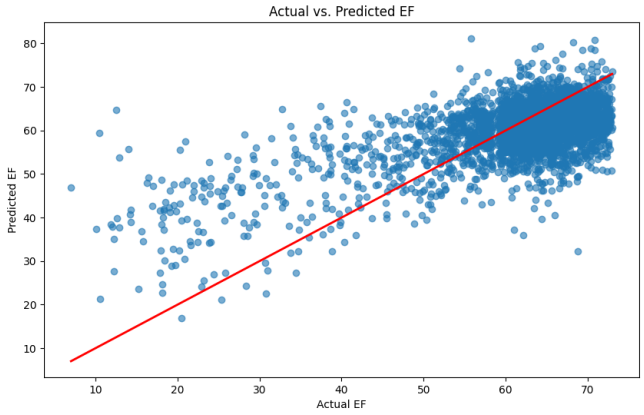


Figure 46: Predicted vs Actual EF (4th Attempt)

It is worthy to note at this point that given the inherent human error in echocardiography, a benchmark validation loss of 25 as the goal to surpass or match.

Subsequent explorations into data augmentation, such as video rotation, and incorporating masks and volume tracings into the model, regrettably, weakened the results.

Working and refining the above steps the journey towards crafting the optimal model gradually passed the most obstacles, such as Odysseus navigating through his mythical challenges, and the task at hand threw a bow to its final course for the destination. The methodology of the final model and its preprocessing tasks are delineated below.

Further preparing the final Dataset

When beginning this project, the first major task was figuring out the best way to prepare the dataset. The initial idea was to simply use the first 16 frames from each video. However, considering the goal was to predict the ejection fraction – a measure that is specifically concerned with the change in the heart left ventricles volume between beats – it became clear that using just any frames wouldn't work. Instead, it was crucial to select specific frames that captured the left ventricle from the end-diastole to the subsequent end-systole.

This decision was made while trying to identify a way to calculate the ejection fraction using a function that can be used dynamically to each video. Thus, on the one hand the 16 frames of the cardiac cycle would be available and on the other a feature that is supposed to be near the ejection fraction. The foundational belief was that embedding these meticulously extracted and computed features along with the preliminary approaches would successfully pave the way towards a model capable of discriminating the patterns in the data, thus making accurate predictions of ejection fractions.

There exist multiple methods of calculating the left ventricle ejection fraction widely adopted by medical professionals around the globe. Two of the most significant methods are the Simpson's Method and the Quinones (Teichholz) Method. Here's a more in-depth look at both:

1. Simpson's Method

It is based on the geometric assumption that the left ventricle can be approximated as a series of cylindrical disks along its long axis. The areas of these disks are summed to calculate the volume.

Here is the general formula to calculate ventricular volume using Simpson's Method:

$$V = \frac{Area_1 + Area_2 + \cdots + Area_n}{Area_n} \times L$$

Where $V$is the volume, $Area_i$ are the areas of the disks, and L is the length of the ventricle.

2. Quinones Method (Teichholz Formula)

It is based on a geometric model that approximates the left ventricle as a prolate ellipse of revolution.

Here is the formula to calculate ventricular volume:

$$V = 7.0 \times \frac{D^3}{2{,}4 + D}$$

Where D is the internal diameter of the ventricle at end-diastole or end-systole.

No matter which of the above method used to calculate the cardiac volume, the ejection fraction is compute by the formula:

$$EF = \frac{(EDV - ESV)}{EDV} \times 100$$

Where EDV is the end diastolic volume, and ESV is the end systolic volume.

While the Simpson's method has a wider acceptance in echocardiographic machines due to its precision, its practical application solely on videos presents great difficulties. This limitation enabled the decision to use the Quinones method, which offers a more feasible avenue for video-based calculations.

This rather complicated preprocessing task was handled as a separate step due to its demand for computational resources resulting in creating a h5py dataset **(h5py, n.d.)** to be handled in the training model algorithm.

Step 1: Filtering Outliers

The meticulous observation showed that the left ventricle segmentation included frames that were not accurate enough. Before any further processing, it was crucial to ensure the reliability of the data being used. Outliers, which are extreme points, can skew the results and lead to untrustworthy predictions. In the case of this study, the outliers refer to masks that deviate significantly from the median mask of a given dataset. By calculating the Mean Squared Error (MSE) between each mask and the median mask, those that were outside a determined threshold (defined as three times the

standard deviation above the mean MSE) were excluded. This step ensured that only the most representative and consistent masks are used for further analysis.

Step 2: Visualizing Cardiac Cycles - Analysis of Local Maxima and Minima

One of the standout features of the preprocessing was the visualization of the cardiac cycles. By plotting the smoothed areas of the heart, it provided a clear graphical representation of the cardiac activity over time. These visualizations helped in identifying the peaks (local maxima) and valleys (local minima) in the data, representing the end-diastole and end-systole, respectively.

Cardiac cycles, in essence, consist of the heart contracting and relaxing. In the context of the data, these events are represented as the local maxima (diastole) and local minima (systole) on the smoothed area plot (red and green respectively). Through advanced conditional checks, the code identifies these critical points. For instance, if the first frame has a higher smoothed area than the first detected local maxima, it is added as a local maxima to consider the potential of capturing the entire cardiac cycle.

Below there is a representative visualization sample of this approach. The end-diastole is denoted as purple, corresponds to the biggest area and was the chosen first frame for the subsequent analysis and the following systole denoted as yellow was chosen as the end-systole, thus the last frame.

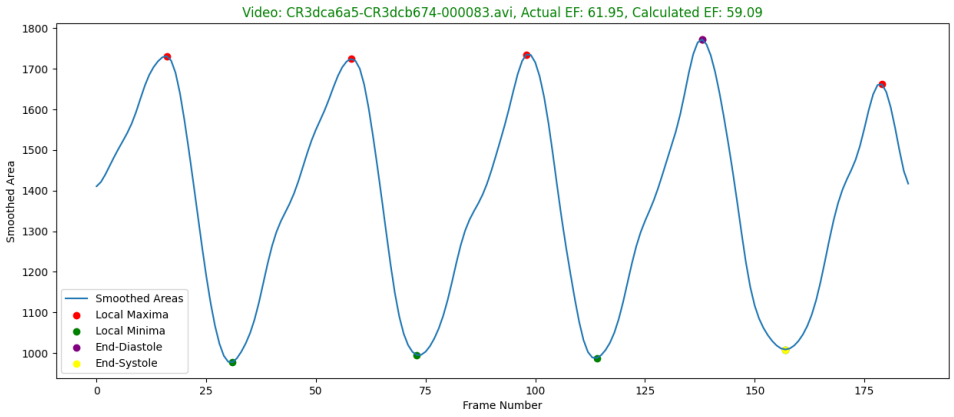


Figure 47: Cardiac cycles sample based on the area of the left ventricle, performing local optimization while depicting the chosen end diastole and end systole.

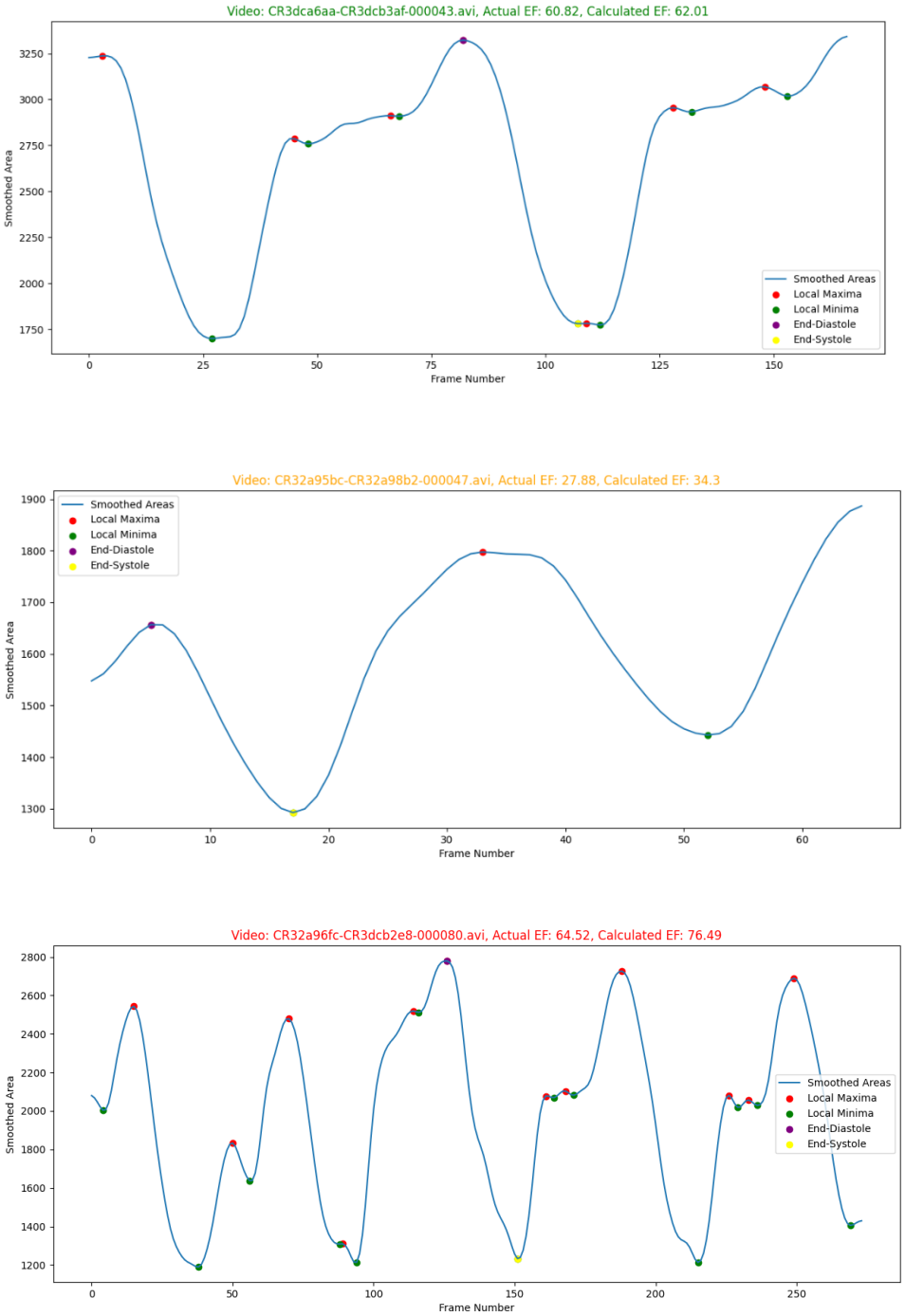


Figure 48: Cardiac cycles samples based on the area of the left ventricle, performing local optimization while depicting the chosen end diastole and end systole.

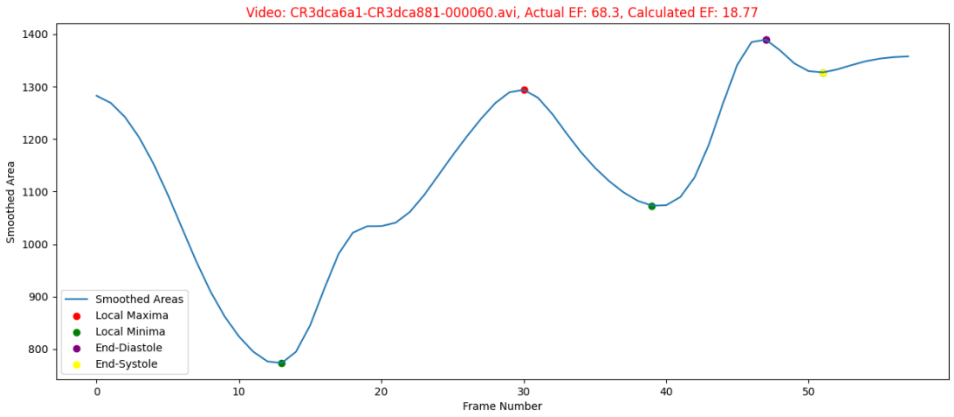


Figure 49: Cardiac cycles samples based on the area of the left ventricle, performing local optimization while depicting the chosen end diastole and end systole.

The title was highlighted to visually display the absolute difference between the Actual EF as calculated by the experts and the calculated Quinones EF.

Step 3: Computing the Quinones EF

One more notable aspect of this research is the computation of the Quinones Ejection Fraction (EF), a formula derived from ventricular dimensions. After identifying the end-diastole and end-systole frames, the areas under these points are used to compute the LVIDs (Linear Dimension at systole) and LVIDd (Linear Dimension at diastole). Those linear dimension calculations assume that these areas can fit roughly in a circular shape thus able to compute its diameter. These dimensions are then used in the Quinones formula to estimate the EF, providing an approach to determining the ejection fraction without direct measurement. This approach is commonly used in echocardiography to estimate linear dimensions from area measurements, especially when a more precise geometric model of the left ventricle cannot be easily derived from the available data. The computed EF then was incorporated into the demographics dataset.

Step 4: Adjusting Frame Numbers:

After calculating the number of frames between end diastole and end systole there were variations, as expected by the preliminary analysis. It was imperative for the model input that each video is represented by a consistent number of 16 frames. This step involved either interpolation (for videos with fewer frames) or taking the median (for videos with more frames) to ensure each video data is standardized to 16 frames.

Step 5: Introduce intensity variations.

Intensity variations across the cardiac cycle in the echocardiographic images were incorporated since the variations in pixel intensity may correlate with the movement and changes in the heart throughout the cycle that might hint pathological conditions of the left ventricle, thus affecting the Ejection Fraction (EF).

Step 6: Integration of Left Ventricle Area Changes

The alteration in the LV area throughout the cardiac cycle is a direct reflection of the heart's ability to pump blood effectively. The difference in the LV area during the entire cardiac cycle could give a representation of the heart’s activity throughout different phases, which is linked to the Ejection Fraction (EF).

Step 7: Data Augmentation on the Minority Class (EF<50) and Noise Addition

In the initial preprocessing analysis of this thesis, it was profound that the distribution of the not normal Ejection Fraction (EF) was significantly less than the one of the normal. Thus, paved the way to introduce data augmentation to the data where the Ejection Fraction (EF) was less than 50. The minority class data underwent rotations of 7 and 14 degrees to create new synthetic samples. In addition, Gaussian noise was added to these rotated images to introduce further variance and ensure a solid approach potentially enhancing the model's ability to generalize.

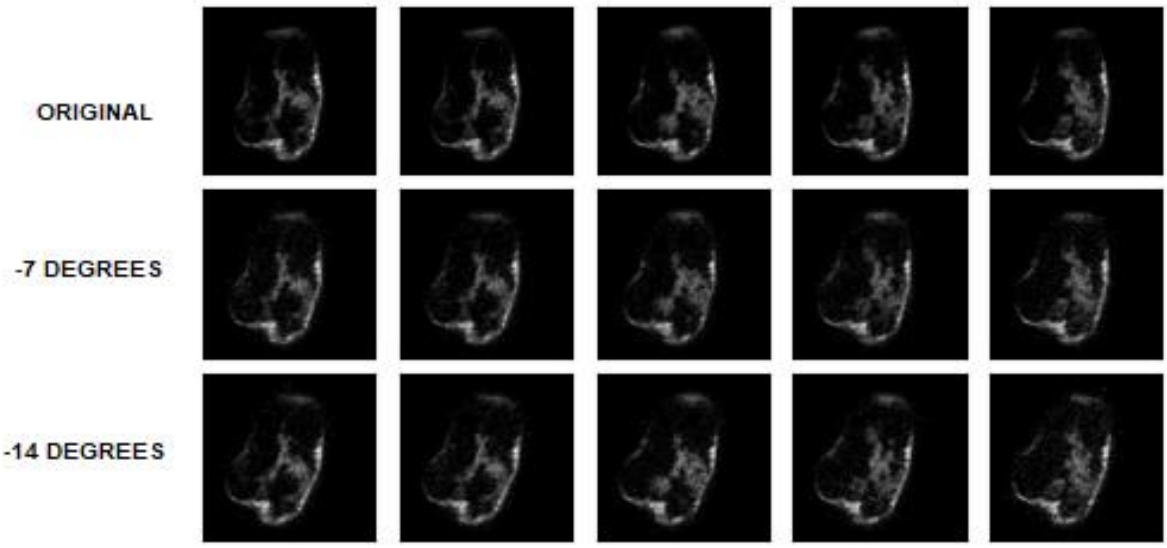


Figure 50: A sample of the original and noised rotated frames.

Tripling Minority Class Data: For non-image data (demographic data, lv_area_changes, and intensity_variations), the minority class samples were tripled to bring their shape on par with the shape of the image data. This was a useful strategy in dealing with class imbalance.

Step 8: Compiling the Final Dataset:

After all the meticulous steps of data preparation, from outlier filtering to the computation of the Quinones EF, the final dataset is compiled. This dataset, which consists of adjusted frames, their corresponding masks, demographic data, calculated ejection fractions, intensity variations, left ventricular area changes and the actual ejection fractions were now ready to be incorporated in the model, transforming it into a Multi Input.

Step 9: Data Normalization

The normalization of the data was an essential step to prepare the dataset for modeling. Data normalization was performed to scale the demographic numerical data apart from the Quinones Ejection Fraction (EF), ensuring that each parameter has a similar data distribution. This is particularly important for models that use gradient descent as an optimization strategy, as it helps to ensure that all features are treated equally when training the model, apart from the calculated Quinones Ejection Fraction (EF) that was deliberately left out to dominate this dataset.

Step 10: Satisfying the Training and Test Sets

Attempting to implement data stratification for the training and test sets, the data were divided into three age categories: 0-5, 6-12, and 13-18 years. Within each age group, Ejection Fractions (EFs) were classified as either normal or not normal based on whether they were above or below 50. The proportion of EFs less than 50% was calculated for each age category to understand their distribution. Using these proportions, a selection of samples was made from each age group to assemble a representative test set, maintaining the original EF distribution. The remaining data formed the training set, ensuring both sets were reflective of the overall data characteristics.

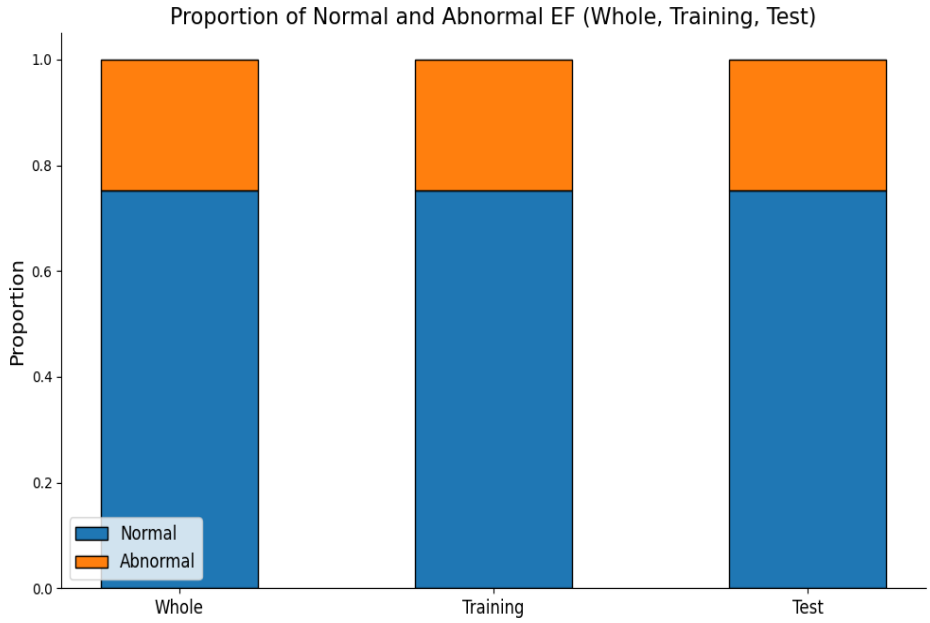


Figure 51: Proportion of normal and not normal EF of the three sub-datasets

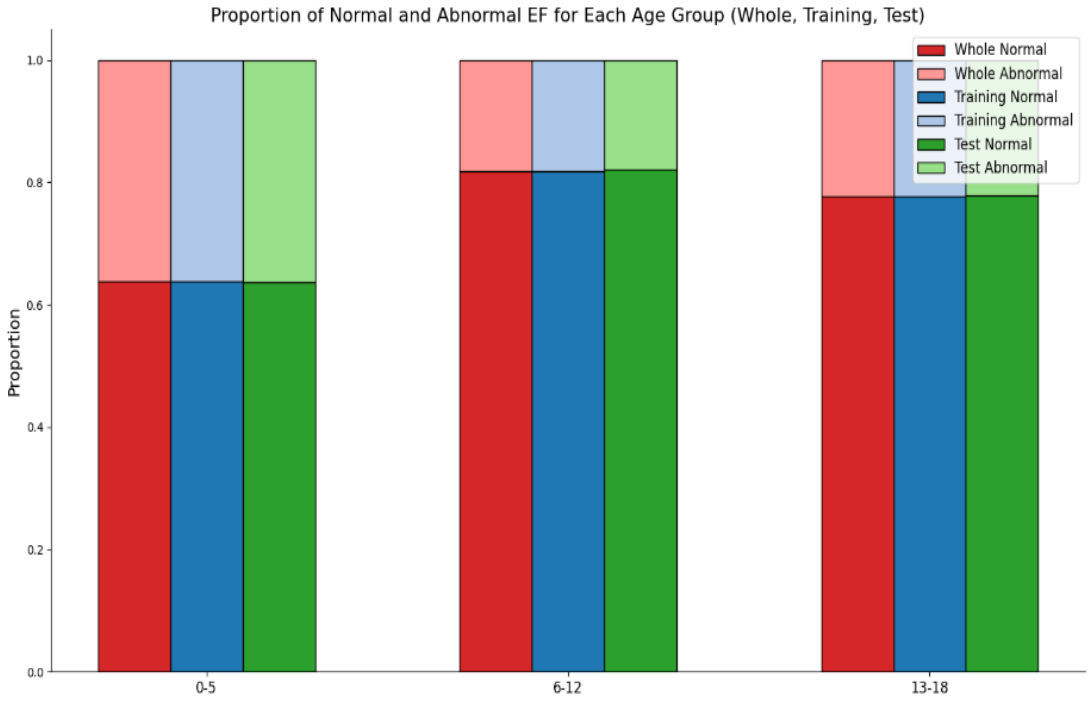


Figure 52: Proportion of normal and not normal EF per age group per sub-dataset

Model Architecture

The model, once a two stream, was transformed to a multi-stream architecture strategically designed to process different types of inputs in parallel using TensorFlow and Keras before merging them for the final predictive step.

The first stream processes the grayscale video frames (spatial stream) derived from the segmentation model using 3D convolutional layers, while the second stream addresses the optical flow between frames (temporal stream) also via 3D convolutional layers. Both streams utilize MaxPooling and BatchNormalization to manage the computational complexity and enhance the training stability.

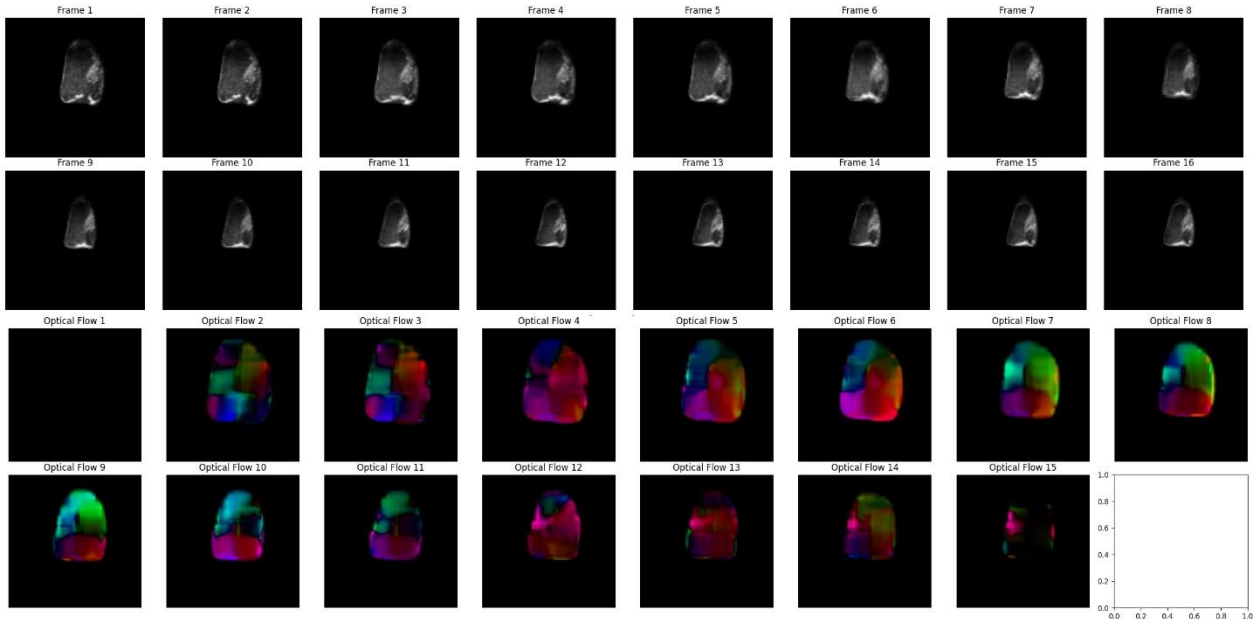


Figure 53: A sample of the 16 input frames from End Diastole to End Systole allong with their Optical Flow

In addition, three separate input streams for demographic data, left ventricle (LV) area changes, and intensity variations are integrated, each processed through a dense layer with 'ReLu' activation followed by Batch Normalization and Dropout for regularization.

The outputs of all these independent streams are concatenated merging their insights into a fully connected dense layer with 'ReLu' activation, finally passing through dropout and a final dense layer to produce the final continuous output, predicting the Ejection Fraction (EF).

The following plot illustrates the architecture of the final model, showcasing the multiple inputs and how the five inputs contribute to the prediction.

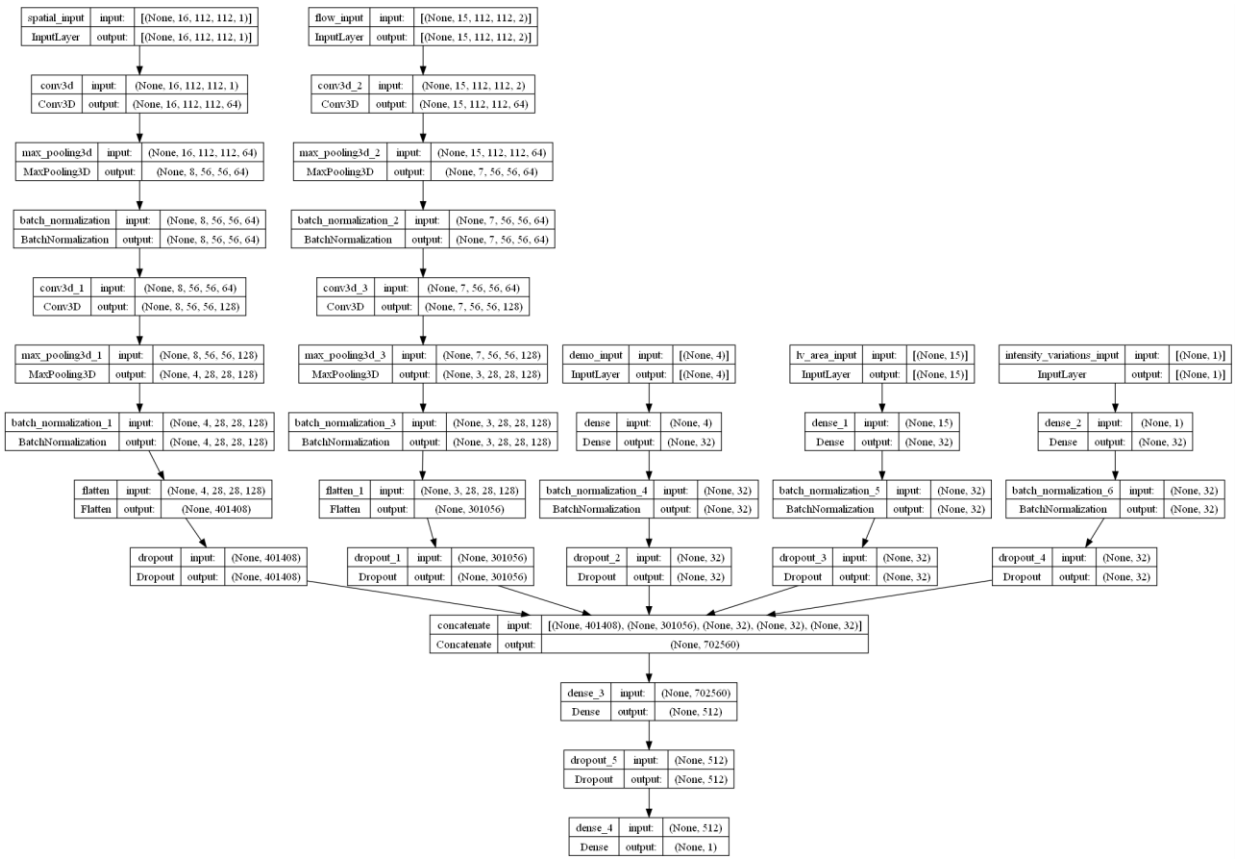


Figure 54: Final Multi Input Model Visualization

Activation function

The activation function ReLu **(towardsdatascience.com, n.d.),** even though it looks like a linear function, introduces non-linearity to the model that allows it to solve complex problems and understand intricate patterns.

Loss Function

The Mean Squared Error (MSE) **(Science Direct, n.d.)** was selected as the loss function for training the model. Given the model's task of predicting a continuous variable (the Ejection Fraction), the MSE calculates the average of the squares of the errors between the predicted and actual EF values. This choice is typical for regression problems, as it effectively penalizes larger errors in prediction and is differentiable, which is beneficial during backpropagation.

Optimization Strategy

The optimization of the model parameters was performed using the Adam optimizer with a learning rate of 0.00001. Adam **(keras, n.d.)**, a widely-used optimization algorithm, computes adaptive learning rates for each parameter by leveraging the moving averages of the parameters. This choice provides efficient, low-computational-cost optimization while also being suitable for problems with noisy or sparse gradients, which is often the case in deep learning.

Training Procedure

The training phase was performed over 20 epochs with a batch size of 16, utilizing several callbacks to enhance the training process. Early Stopping was employed to monitor the validation loss, which stopped training if no improvement was observed for five consecutive epochs while reverting the model weights back to those of the best-performing epoch. Model Checkpointing saved the model weights at the epoch with the lowest validation loss, ensuring that the best-performing model was retained for further tasks. Additionally, the Reduce Learning Rate on Plateau strategy was used, reducing the learning rate when the validation loss ceased to improve, providing the model with finer control as it approached a minimum. These strategies aimed to safeguard the model against overfitting and assist the training process by guiding it through the different possibilities.

Below there is a representation of some epochs of the training process showcasing these strategies.

```
Epoch 1/20

195/195 [==============================] - ETA: 0s - loss: 192.3020

Epoch 1: val_loss improved from inf to 632.82275, saving model to best_model_weights.h5

195/195 [==============================] - 2210s 11s/step - loss: 192.3020 - val_loss: 632.8228 - lr: 1.0000e-05

Epoch 2/20

195/195 [==============================] - ETA: 0s - loss: 96.7948

Epoch 2: val_loss improved from 632.82275 to 117.63184, saving model to best_model_weights.h5

195/195 [==============================] - 2202s 11s/step - loss: 96.7948 - val_loss: 117.6318 - lr: 1.0000e-05

…

Epoch 4/20

195/195 [==============================] - ETA: 0s - loss: 69.6917
```

Epoch 4: val_loss did not improve from 76.50826

195/195 [==============================] - 2195s 11s/step - loss: 69.6917 - val_loss: 89.6854 - lr: 1.0000e-05

..

Epoch 9/20

195/195 [==============================] - ETA: 0s - loss: 54.0595

Epoch 9: val_loss did not improve from 56.05175

Epoch 9: ReduceLROnPlateau reducing learning rate to 1.9999999494757505e-06.

195/195 [==============================] - 2198s 11s/step - loss: 54.0595 - val_loss: 90.8225 - lr: 1.0000e-05

…

Epoch 11: val_loss did not improve from 56.05175

195/195 [==============================] - 2198s 11s/step - loss: 42.0400 - val_loss: 56.1998 - lr: 2.0000e-06

Epoch 11: early stopping

The following plot depicts the Training and Validation Loss across all the epochs:

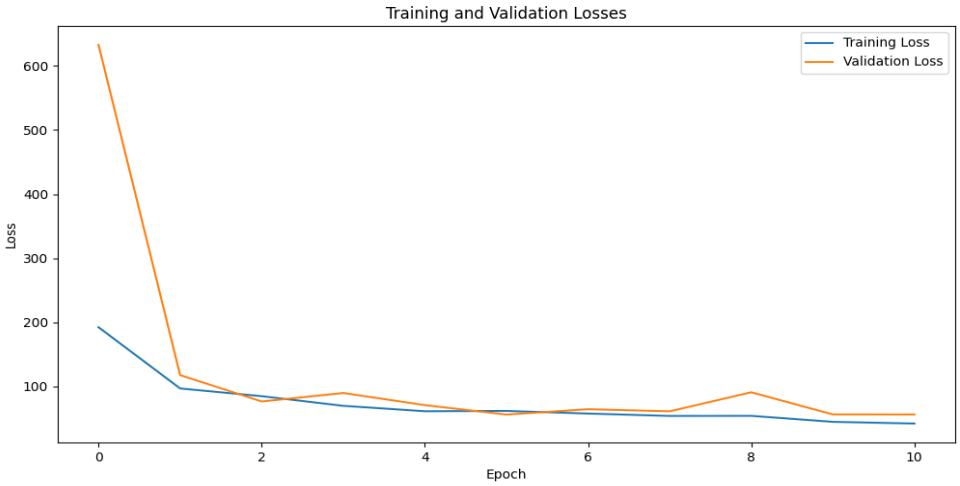


Figure 55: Train and Validation Loss per epoch plot

Model Training Results

The model which contained augmented data returned the following results:

- Mean Absolute Error (MAE): 5.36

- Mean Squared Error (MSE): 47.55
- Root Mean Squared Error (RMSE): 6.89
- R-squared (R^2): 0.77

The plot below shows that the data points as well as the trend line in green (first polynomial) lie very close to the perfect fit line compared to the initial attempts.

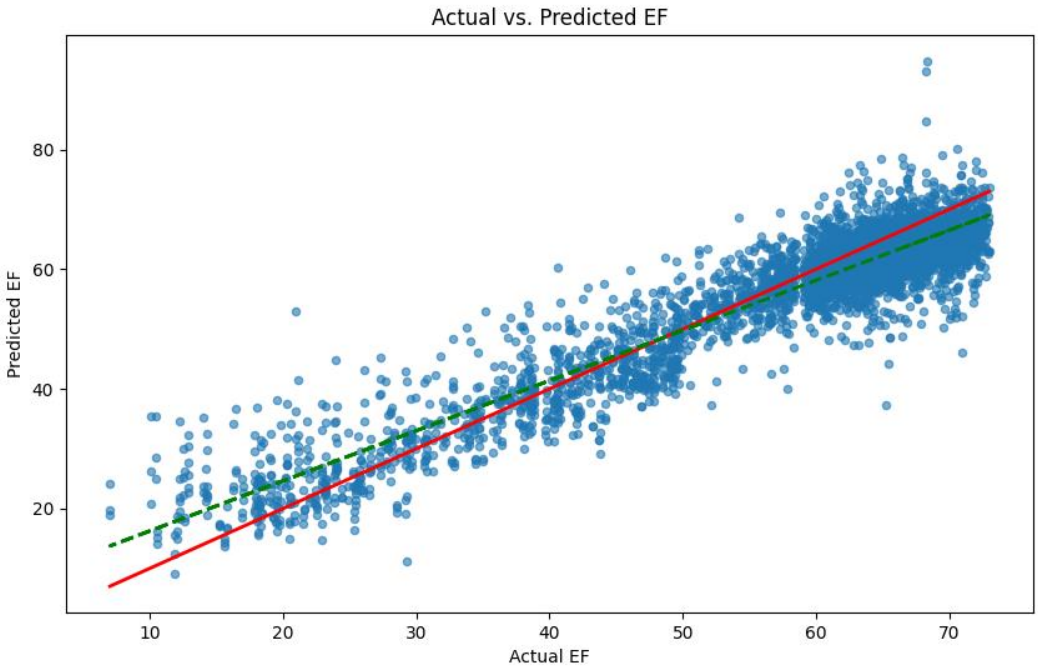


Figure 56: Actual vs Predicted EF Plot of the Final Model

Residual Analysis

Of equal importance was the residuals analysis. Residuals represent the difference between the actual values and the values predicted by the model. For a given data point, the residual is calculated as:

$$Residual = Actual\,Value - Predicted\,Value$$

In the context of a regression model, residuals help understanding how far off the predictions are from the actual values. If a model fits the data perfectly, the residuals would be zero for all data points.

In regression analysis, one of the key assumptions is that the residuals (or errors) are normally distributed. If this assumption proves true, it means that the model is well specified and that the error in the regression model does not have any patterns that the model is not capturing. In simpler terms, it indicates that the model is making errors that are random and not systematic.

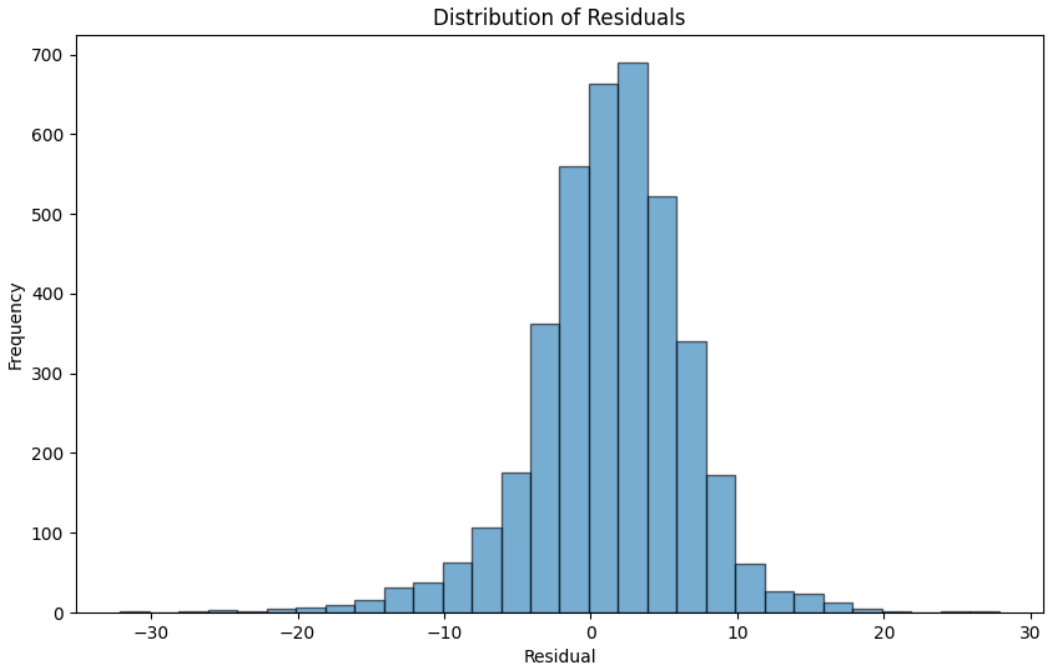


Figure 57: Distribution of Residuals of the Final Predicted values for residual analysis

The residuals, as can be seen in the above graph, are normally distributed and centered around zero. This shows that there was no obvious bias in the model predictions. If the residuals were skewed to the left or right, it would have indicated that the model consistently overpredicts or underpredicts the target variable. The bell shape indicates that most of the model's predictions are close to the actual values, with fewer predictions deviating significantly.

Chapter 4

Results

Introduction

In the pursuit of knowledge, presenting the research results serves as a defining moment. These findings not only validate the methods employed but also set the stage for conclusions and future work. The primary goal of this study was to predict the Left Ventricle Ejection Fraction (EF) with a high degree of accuracy. This chapter presents the quantitative outcomes derived from the final multi-input regression model trained on the EchoNet-Pediatric dataset and compares the results with those of recent related studies.

Quantitative Analysis

Final Results

After time intensive training, the model was saved as "h5.py" and tested on the initial dataset. The results were not only remarkable but also surpassed our goal. The model achieved the following metrics:

- Mean Absolute Error (MAE): 2.69
- Mean Squared Error (MSE): 15.96
- Root Mean Squared Error (RMSE): 3.99
- R-squared (R^2): 0.84

These metrics provide a complete understanding of the model's performance. A MAE of 2.69 indicates that the model's predictions are, on average, approximately 2.69 units away from the true values, which is significantly less than the human error.

The MSE and RMSE, on the other hand, penalize larger errors more than smaller ones, making them especially useful when large errors are particularly undesirable. An RMSE of 3.99 suggests that the model's predictions differ on average from the actual values by roughly 3.99 which again is significantly less than the human error.

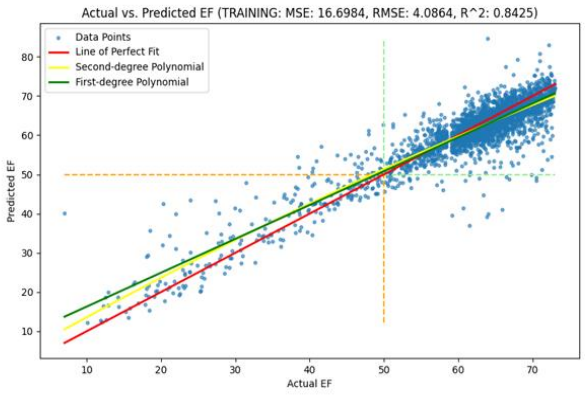
Lastly, the R^2 value shows how much of the variation in Ejection Fraction the model can explain using the given data as a proportion. The R^2 value of 0.84 suggests that 84% of the variability in the Ejection Fraction can be explained by the model, which is an impressive achievement for a regression task.

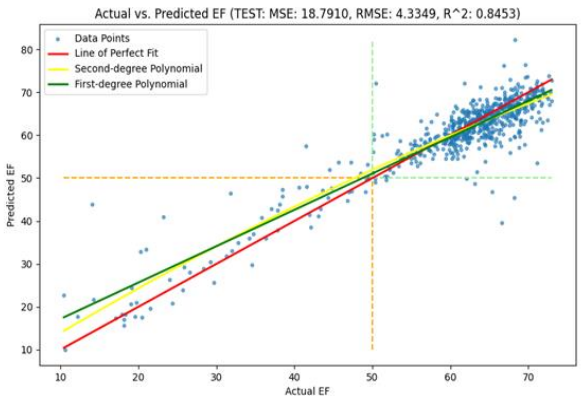
Visualization and Interpretation

The three following plots provide a visual representation of the relationship between the actual and predicted Ejection Fraction (EF) values for the training set, the test set, and the entire dataset.

The line of perfect fit, depicted in red, represents the ideal scenario where every prediction matches the actual value. The first-degree and second-degree polynomial trend lines, shown in green and yellow respectively, indicate the general direction and bend of the data points.

The orange and light green dashed lines create four quadrants, where data points that fall between different colors indicate instances where the EF was not accurately classified as either normal or abnormal.





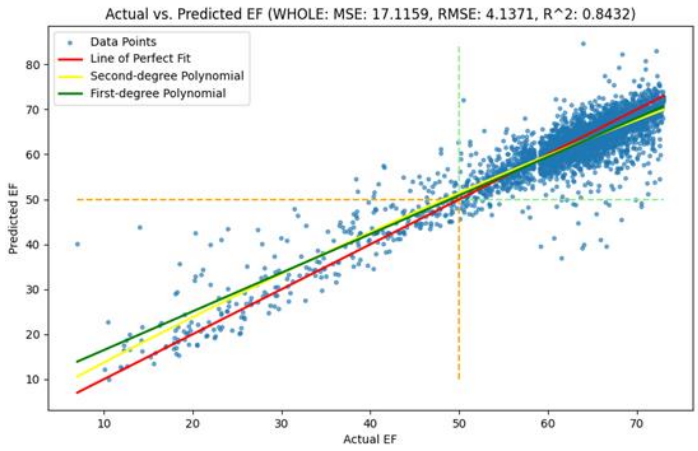


Figure 58: Actual vs Predicted EF of the Final Model for the train, test and complete datasets.

Residual Distribution

The residual distribution plot displays that residuals are normally distributed around zero, showcasing no bias in the model's predictions. The bell shape suggests most predictions are close to the actual values. What is more, there is a narrower spread across the range compared to the augmented data, that indicates the model's consistent prediction capabilities.

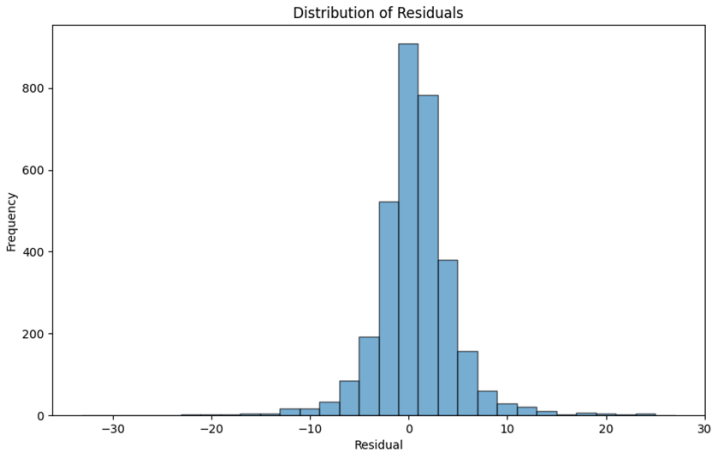


Figure 59: Residual Distribution Analysis on the original dataset.

Comparative Analysis

In evaluating the performance of the model, it's essential to compare its outcomes with existing methodologies. The results are compared with two notable studies in the domain of video-based deep learning models for predicting Ejection Fraction (EF) both conducted in Stanford University.

Video-Based Deep Learning Model for Automated Assessment of Ejection Fraction in Pediatric Patients. **(Reddy *et al.*, 2022)**

Dataset: Same as this study - EchoNet-Pediatric.

Key Results:

- The model predicted an R^2 of 0.73 for the A4C view and 0.74 for the PSAX view. When combined, the overall R^2 was 0.78.
- Mean Absolute Error (MAE) for EF estimation was 3.66%.

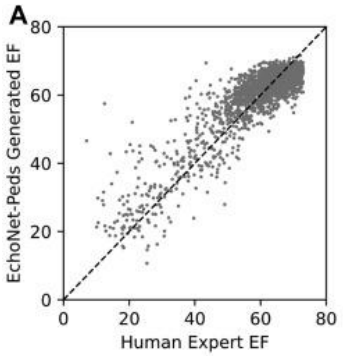


*Figure 60: EchoNet-Pediatric Results **(©Reddy et al., 2022)***

The model in this study, with an R^2 of 0.84, demonstrates superior performance in predicting EF compared to the EchoNet-Peds model. Additionally, the MAE of 2.69% is lower, indicating better precision in predictions.

Video-based AI for beat-to-beat assessment of cardiac function **(Ouyang *et al.*, 2020)**.

Dataset: A large dataset of 10,030 annotated echocardiogram videos.

Key Results:

- Mean Absolute Error (MAE): 4.1%
- Root Mean Squared Error (RMSE): 5.3%
- R^2: 0.81

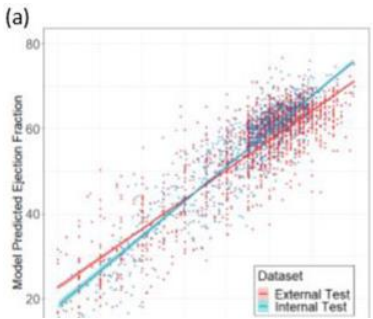


*Figure 61: EchoNet-Dynamic Results **(©Ouyang et al., 2020)***

Compared to EchoNet-Dynamic, the model in this study exhibits a slightly better R^2 value but with a notably reduced MAE and RMSE.

The table below summarizes the comparison of the three studies:

	This Study	***EchoNet-Pediatric (Reddy et al., 2022)***	***EchoNet-Dynamic (Ouyang et al., 2020)***
Dataset	EchoNet-Pediatric	EchoNet-Pediatric	10,030 annotated echocardiogram videos
MAE	2.69	3.66	4.1
RMSE	3.96	-	5.3
R^2	0.86	0.73 (A4C view), 0.74 (PSAX view), 0.78 (combined)	0.81

Table 1: Comparison of this study to Stanford's EchoNet studies

In conclusion, the model exhibits performance superior to EchoNet-Pediatric and competitive to EchoNet-Dynamic in predicting the Left Ventricle Ejection Fraction (LVEF) when benchmarked against contemporary research.

Chapter 5

Discussion

Computer Vision and Deep Learning have been steadily reshaping the medical imaging diagnostics field, including echocardiography, with breakthroughs that have the potential to alter these subfields while offering new crossroads where traditional medicine expertise meets rapidly evolving cutting-edge technologies.

Interpretation of Results

The results from this study emphasize the potential of incorporating Artificial Intelligence (AI), particularly in the domain of pediatric echocardiography.

The accuracy and efficiency of the two deep learning models developed, one for automated left ventricular segmentation and one for automated ejection fraction prediction not only validate the capabilities of joining the forces of Computer Vision and Deep learning on pediatric echocardiographic videos but also could contribute to the ongoing evolution of pediatric echocardiography.

Furthermore, addressing the limitations of pediatric echocardiography presented in this study and its significant role in diagnosing and treating congenital and acquired cardiovascular diseases in children compared to other imaging techniques such as MRI and CT, the results of the study show that integrating computer vision and deep learning to echocardiography can help minimize the challenges posed by inter-observer variability as well as minimize the dependence on specialized experts. In addition, a great advantage is the minimum data requirement. Only a cardiac cycle is needed, which in children lasts less than a second, and this can be beneficial for data acquisition, minimizing the time needed to take a sample video, addressing the patient movement especially for children in early ages.

The outcomes of this research, as presented in the results sections, offer valuable insights supporting these hypotheses, highlighting the potential of AI-driven solutions that offer a promising path forward, ensuring that pediatric patients receive the most accurate and comprehensive care possible.

Limitations and Challenges

Undertaking this research was a journey not without its challenges. Starting from the Data Acquisition, researching the intricacies of employing cutting edge technologies in echocardiographic, but not limited to, imaging, dealing with a dataset of the delegate population of children, trying to exceed the results of research conducted by interdisciplinary teams and the relation with the author's heart struggles this journey was far from technical but deeply linked with emotions, passion, and personal experiences.

The primary objective was to utilize a dataset not widely explored, aiming to create an original thesis that had the potential to incorporate a variety of the material covered during the Master's in Data Science program. Given the fact that Data Science is at the forefront in today's world this was not an easy task. Initial efforts to acquire a health-related tumor centric dataset from another Institution were not fruitful. However, after extensive research a dataset was identified that had a lot of potential that enabled the opportunity to incorporate a variety of the program's curriculum.

On the technical front, the complexity of deep learning requires a relatively big dataset, and the quality of predictions depends on the quality and diversity of the training set. Although the dataset was robust, and the biggest pediatric dataset as of today, one could argue that more varied data would have further enhanced the model's accuracy, especially concerning data points with a labeled Ejection Fraction less than 50%. However, the dataset showed multiple complexities not only purely because of the high dimensionality that comes with video data, their variety in length and number of frames, the fact that video and image data are not perceived by computers as by the human brain, the two views and their four relative xml files but also because of the inherent noise that comes with medical imaging.

This noise, along with the constant movement of the myocardium during the cardiac cycle, the overlapping cardiac structures due to multiple views in different videos, the varying image quality,

and at times, additional visuals added to the videos by the sonograms, made it impossible for traditional machine vision techniques with OpenCV to effectively segment the left ventricle. Even though OpenCV has proven effective in various applications and offers a plethora of tools, countless hours spent trying multiple methods and mixed methodologies returned insufficient results. Specifically, beyond the typical functionalities and morphological operations, methods such as background subtraction, Hough lines and Hough ellipsoids, trend lines based on pixel intensity, patterning the sonographic angle, and several others demonstrated that this task was exceptionally challenging to address.

As for the machine learning part, due to the unique nature of echocardiographic images, multiple pretrained feature extraction methodologies such as VGG: Developed by the Visual Geometry Group at Oxford **(Simonyan & Zisserman, 2014)** and ResNet (Residual Networks): Introduced by Microsoft Research **(He *et al.*, 2016)** were ineffective, necessitating further research for finding methodologies to extract meaningful features for the machine learning (ML) algorithms. In addition, it is important to note there are not available libraries in python that offer a bisecting K-means clustering algorithm. This gap required the development of a custom solution.

Moreover, while both deep learning models performed admirably, they have the potential for occasional errors. It is important to acknowledge that no model, no matter how sophisticated, can guarantee absolute accuracy. This emphasizes the need for a thorough validation process before integrating such models into the echocardiographic ultrasound machines. In addition, it is important to add that other approaches like Time Distributed Two Stream Convolutional Networks (CNNs), 2-Dimensional (2D) Spatial Temporal Convolutional Networks (CNNs) and simpler approached were tried put none proven sufficient enough to produce effective results.

Finally, this research “was close to the heart” of the researcher due to personal struggles with myocarditis. Revisiting the intricacies of the cardiac functionalities and recalling numerous imagine tests through the healing process brought back strong memories and emotions. These feelings strengthened the determination to produce significant results.

Future Work

This study highlights the great opportunities presented by integrating machine learning into pediatric echocardiography. Looking ahead, despite the admirable results, there are several key areas emerging that if addressed, could on the one hand produce even better results and on the other could make it the beginning of new endeavors.

A natural progression following the development of the robust models, is to make their insights more accessible and interpretable to end-users via the Graphical User Interface.

The human heart is a complicated organ with intricate structural and dynamic functionalities. It is best understood from multiple perspectives, at least for humans. Thus, this could be the case with adding the Parasternal Short Axis (PSAX) view in the model while incorporating the whole dataset. However, the scope of this thesis was to explore predicting the Ejection Fraction, solely on the left ventricle Apical 4-Chamber (A4C) view, since Auto EF functionality in the existing ultrasounds is based on this view.

The intersection of medical and data science offers great opportunities. Yet, to reach its full potential, multidisciplinary cooperation is essential. Institutions and researchers from different fields should unite embracing the nuances of artificial intelligence not with skepticism but with an open mind set. This will not just benefit the healthcare systems, but most importantly patients worldwide. This is the case with this thesis as well. A next step could be partnering with a health institution, either a medical school or a pediatric cardiac research center and representatives from ultrasound equipment manufacturers. Such collaboration will ensure that the developed models are not just technically robust, but also medically reliable and feasible to be integrated into ultrasound systems, leading to clinical trials.

The efficiency of a data-driven approach is often related to the extension of its underlying dataset. Future efforts to expand the current dataset, incorporating echocardiographic videos from a diverse range of patients, different age groups, health conditions, geographical backgrounds and manufacturers can enhance the model's generalizability, making it a complete model for any equipment and patient.

Medical science and Data Science by their very nature are evolving constantly, and as more data becomes available it is crucial for models and methodologies to adapt accordingly. Thus, periodic reviews, refinements, retraining, and updates are of great importance to future work, ensuring this research remains current and aligned with the latest advancements in both fields.

Chapter 6

Conclusion

The field of cardiac diagnostics has witnessed a remarkable transformation, especially in recent years. With the advent of advanced medical technologies and the rapid evolution of Data Science, new horizons have opened that offer innovative approaches. At the heart of this study lies the admirable human heart as portrayed by the EchoNet-Pediatric Dataset, an expert annotated, videographic and demographic dataset from children between 0 and 18 years of age who had undergone echocardiography.

Guided by literature from various subjects and material taught in the Master's program in Data Science at the American College of Greece, this thesis embarked on a journey that mirrors the program’s structure. From data analysis and data visualization methodologies to machine learning, from computer vision to deep learning models, it illustrates the inference of new knowledge with the potential to change the world as we know it.

The fusion of data science and echocardiography, as explored in this study, indicates a new era in pediatric echocardiography. While the results are promising, it's imperative to approach the adoption of such technologies with caution, ensuring that they undergo extensive clinical testing.

Every research process, no matter how comprehensive, has its struggles and limitations. Nevertheless, the accomplishments of this research were evidenced by the results and showcase the potential of incorporating computer vision and deep learning in echocardiography, not only for left ventricle segmentation and prediction of ejection fraction but in general.

On a personal note, the ultimate goal was to explore whether it is possible to provide better, faster, and more consistent diagnostic insights, ensuring that the young patients receive the best care possible.

The study's essence lied in the transition from pixels, the smallest units of digital imaging, to invaluable clinical insight that can shape the future.

References

1. University of Michigan. (2019). *Anatomy of a Human Heart*. Last Access: (2023, September 2). https://www.michiganmedicine.org/health-lab/anatomy-human-heart
2. University of California San Diego. (2017, February). *Cardiac Cycle*. Last Access: (2023, September 2). https://cvil.ucsd.edu/wp-content/uploads/2017/02/cardiac-cycle.pdf
3. Mayo Clinic. (n.d.). *Ejection fraction: What does it measure?* Last Access: (2023, September 3). https://www.mayoclinic.org/tests-procedures/ekg/expert-answers/ejection-fraction/faq-20058286
4. Josh Hopkins Medicine (n.d.). *Echocardiogram.* Last Access (2023, September 3). https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/echocardiogram
5. World Health Organization. (2021, June 11). *Cardiovascular diseases (CVDs)*. Last Access: (2023. September 3) https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
6. World Heart Federation. (2023, August 11). *CALL TO ACTION ON ADDRESSING THE GLOBAL BURDEN OF PEDIATRIC AND CONGENITAL HEART DISEASES.* Last Access: (2023. September 3) https://world-heart-federation.org/news/call-to-action-on-addressing-the-global-burden-of-pediatric-and-congenital-heart-diseases/
7. Lai, W. W., Geva, T., Shirali, G. S., Frommelt, P. C., Humes, R. A., Brook, M. M., Pignatelli. R.H., Rychik, J. (2006). *Guidelines and standards for performance of a pediatric echocardiogram: A report from the Task Force of the Pediatric Council of the American Society of Echocardiography.* Journal of the American Society of Echocardiography, 19(12), 1413-1430. https://doi.org/10.1016/j.echo.2006.09.001

8. Slostad, B., Karnik, A., Appadurai, V., Narang, A. (2023) *Applications of Artificial Intelligence in Echocardiography*. *Current Cardiovascular Risk Reports* 17(7):123–132. https://doi.org/10.1007/s12170-023-00721-6
9. Otto, C. M. (2019). *Textbook of Clinical Echocardiography*. (7th ed.). Elsevier Health Sciences.
10. Penny, D. J., Cetta, F., Mital, D., Feltes, T. F. (2021). *Moss & Adams' Heart Disease in infants, Children, and Adolescents Including the Fetus and Young Adult.* (10th ed.). Wolters Kluwer.
11. Opfer, E., Shah, S., (2018) *Advances in pediatric cardiovascular imaging.* Mo Med. 115(4):354-360. PMID: 30228767; PMCID: PMC6140247. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6140247/
12. Bhargava, R., Hahn, G., Hirsch, W., Kim, M.J., Mentzel, H.J., Olsen, O.E., Stokland, E., Triulzi, F., Vazquez, E. (2013). *Contrast-enhanced magnetic resonance imaging in pediatric patients: Review and recommendations for current practice.* Magn. Reason. Insights. 6:95–111 https://doi.org/10.4137/MRI.S12561
13. Chodick, G., Kim, K.P., Shwarz, M., Horev, G., Shalev, V., Ron E. (2009). *Radiation risks from pediatric computed tomography scanning.* Pediatr Endocrinol Rev. 7(2):29-36. PMID: 20118891; PMCID: PMC6298033.
14. Ash, J.A., Chowdhury, Y.S. (2023*) Pediatric Echocardiography Assessment, Protocols, and Interpretation.* StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan–. PMID: 34033337.
15. Kamel, H., Elsayegh, A.T., Nazmi, H. et al. (2022). *Assessment of left ventricular systolic function using two- and three-dimensional speckle tracking echocardiography among healthy preschool-age pediatric children.* Egypt Heart J 74, 21 https://doi.org/10.1186/s43044-022-00258-w
16. Benavidez, O.J., Gauvreau, K., Jenkins, K.J., Geva, T. (2008) *Diagnostic errors in pediatric echocardiography: development of taxonomy and identification of risk factors.* Circulation. 117(23):2995-3001. https://doi.org/10.1161/CIRCULATIONAHA.107.758532. PMID: 18519849; PMCID: PMC4237021.

17. Morbach, C., Gelbrich, G., Breunig, M., Tiffe, T., Wagner, M., Heuschmann, P. U., Störk, S. (2018) *Impact of acquisition and interpretation on total inter-observer variability in echocardiography: results from the quality assurance program of the STAAB cohort study*. The International Journal of Cardiovascular Imaging, 34(7):1057–1065. https://doi.org/10.1007/s10554-018-1315-3

18. Law, Y.M., Lal, A.K., Chen, S., Čiháková, D., Cooper, L.T. Jr, Deshpande, S., Godown, J., Grosse-Wortmann, L., Robinson, J.D., Towbin J.A. (2021). *Diagnosis and Management of Myocarditis in Children: A Scientific Statement.* The American Heart Association. Circulation. 144(6):e123-e135. https://doi.org/10.1161/CIR.0000000000001001. PMID: 34229446.

19. Madani, A., Arnaout, R., Mofrad, M., Arnaout, R. (2018). *Fast and accurate view classification of echocardiograms using deep learning.* npj Digital Med 1, 6 https://doi.org/10.1038/s41746-017-0013-1

20. Reddy, C.D., Lopez, L., Ouyang, Dz., Zou, J.Y., He B. (2023). *Video-Based Deep Learning for Automated Assessment of Left Ventricular Ejection Fraction in Pediatric Patients.* J Am Soc Echocardiogr. 36(5):482-489. https://doi.org/10.1016/j.echo.2023.01.015. PMID: 36754100. The EchoNet-Pediatric dataset https://echonet.github.io/pediatric/

21. Kligfield, P., Gettes, L.S., Bailey, J.J., Childers, R., Deal, B.J., Hancock, E.W., van Herpen, G., Kors, J.A., Macfarlane, P., Mirvis, DM., Pahlm, O., Rautaharju, P., Wagner, G.S. (2007) American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; American College of Cardiology Foundation; Heart Rhythm Society; Josephson, M., Mason, J.W., Okin, P., Surawicz, B., Wellens, H. *Recommendations for the standardization and interpretation of the electrocardiogram: part I: The electrocardiogram and its technology: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: endorsed by the International Society for Computerized Electrocardiology*. Circulation. 2007 Mar 13;115(10):1306-24. https://doi.org/10.1161/CIRCULATIONAHA.106.180200. PMID: 17322457.

22. Saeed, M., Van, T.A., Krug, R., Hetts, S.W., Wilson.,M.W. (2015). *Cardiac MR imaging: current status and future directions.* Cardiovasc Diagn Ther. 5(4): 290–310. https://doi.org/10.3978/j.issn.2223-3652.2015.06.07
23. Prempeh, A.B.A., Scherman, J., Swanevelder, J.L., (2020). *Transesophageal echocardiography in minimally invasive cardiac surgery.* Curr Opin Anaesthesiol. 33(1):83-91. https://doi.org/10.1097/ACO.0000000000000807. PMID: 31789893.
24. Edler, I., Lindström, K., *The history of echocardiography.* (2004). Ultrasound Med Biol. 30(12):1565-644. https://doi.org/10.1016/S0301-5629(99)00056-3. PMID: 15617829.
25. Keller. M., Magunia. H., Rosenberger. P., Koeppen. M., (2023). *Echocardiography as a Tool to Assess Cardiac Function in Critical Care-A Review.* Diagnostics (Basel). 13(5):839. https://doi.org/10.3390/diagnostics13050839. PMID: 36899983; PMCID: PMC10001271.
26. Asch, F.M., Poilvert, N., Abraham, T., Jankowski. M., Cleve. J., Adams, M., Romano, N., Hong, H., Mor-Avi, V., Martin, R.P., Lang, R.M. (2019). *Automated Echocardiographic Quantification of Left Ventricular Ejection Fraction Without Volume Measurements Using a Machine Learning Algorithm Mimicking a Human Expert.* Circ Cardiovasc Imaging. 12(9):e009303. https://doi.org/10.1161/CIRCIMAGING.119.009303. PMID: 31522550; PMCID: PMC7099856.
27. Vidal-Perez, R., Grapsa, J., Bouzas-Mosquera, A., Fontes-Carvalho, R., Vazquez-Rodriguez, J.M. (2023) *Current role and future perspectives of artificial intelligence in echocardiography.* World J Cardiol. 15(6):284-292. https://doi.org/10.4330/wjc.v15.i6.284. PMID: 37397831. PMCID: PMC10308270.
28. Karatzia, L., Aung, N., Aksentijevic, D. (2022). *Artificial intelligence in cardiology: Hope for the future and power for the present.* Front Cardiovasc Med. 9:945726. https://doi.org/10.3389/fcvm.2022.945726. PMID: 36312266. PMCID: PMC9608631.
29. Slomka, P.J., Dey, D., Sitek, A., Motwani, M., Berman, D.S., Germano, G. (2017). *Cardiac imaging: working towards fully-automated machine analysis & interpretation.* Expert Rev Med Devices. 14(3):197-212.

https://doi.org/10.1080/17434440.2017.1300057. PMID: 28277804. PMCID: PMC5450918.

30. Monaghan, T.F., Rahman, S.N., Agudelo, C.W., Wein, A.J., Lazar, J.M., Everaert, K., Dmochowski, R.R. (2021). *Foundational Statistical Principles in Medical Research: Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value.* Medicina (Kaunas). 57(5):503.
https://doi.org/10.3390/medicina57050503. PMID: 34065637. PMCID: PMC8156826.
31. Zhou, J., Du, M., Chang, S., Chen, Z. (2021). *Artificial intelligence in echocardiography: detection, functional evaluation, and disease diagnosis.* Cardiovasc Ultrasound. 19(1):29.
https://doi.org/10.1186/s12947-021-00261-2. PMID: 34416899. PMCID: PMC8379752.
32. Chicco, D., Warrens, M.J., Jurman, G. (2021) *The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation.* PeerJ Comput Sci. 7:e623.
https://doi.org/10.7717/peerj-cs.623. PMID: 34307865. PMCID: PMC8279135.
33. Vandenberk, B., Chew, D.S., Prasana, D., Gupta, S., Exner, D.V. (2023). *Successes and challenges of artificial intelligence in cardiology.* Front Digit Health.5:1201392.
https://doi.org/10.3389/fdgth.2023.1201392. PMID: 37448836. PMCID: PMC10336354.
34. Loncaric, F., Camara, O., Piella, G., Bijnens, B. (2021). *Integration of artificial intelligence into clinical patient management: focus on cardiac imaging.* Revista Española de Cardiología (English Edition) 74(1):72-8.
https://doi.org/10.1016/j.rec.2020.07.003
35. Ronneberger, O., Fischer, P. & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells & A. F. Frangi (eds.), *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* (pp. 234--241). Springer International Publishing. ISBN: 978-3-319-24574-4
https://doi.org/10.1007/978-3-319-24574-4_28
36. Krittanawong, C.; Omar, A.M.S.; Narula, S.; Sengupta, P.P.; Glicksberg, B.S.; Narula, J.; Argulian, E. (2023). *Deep Learning for Echocardiography: Introduction for Clinicians and Future Vision: State-of-the-Art Review.* Life.13(4):1029
https://doi.org/10.3390/life13041029

37. Hariton, E., Locascio, J.J. (2018) *Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials.* BJOG. 125(13):1716. https://doi.org/10.1111/1471-0528.15199. PMID: 29916205. PMCID: PMC6235704.
38. Kim, M., Yun, J., Cho, Y., Shin, K., Jang, R., Bae, H.J., Kim, N. (2019) *Deep Learning in Medical Imaging.* Neurospine. 16(4):657-668. https://doi.org/10.14245/ns.1938396.198. PMID: 31905454. PMCID: PMC6945006.
39. Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M., Kaus, M.R., Haker, S.J., Wells, W.M. 3rd, Jolesz, F.A., Kikinis, R. (2004) *Statistical validation of image segmentation quality based on a spatial overlap index.* Acad Radiol. 11(2):178-89. https://doi.org/10.1016/s1076-6332(03)00671-8. PMID: 14974593. PMCID: PMC1415224.
40. Lopez-Jimenez, F., Attia, Z., Arruda-Olson, A.M., Carter, R., Chareonthaitawee P., Jouni, H., Kapa, S., Lerman, A., Luong, C., Medina-Inojosa J.R., Noseworthy, P.A., Pellikka, P.A., Redfield, M.M., Roger, V.L., Sandhu, G.S., Senecal, C., Friedman, P.A., (2020). *Artificial Intelligence in Cardiology: Present and Future.* Mayo Clinic Proceedings. 95(5):1015-1039. https://doi.org/10.1016/j.mayocp.2020.01.038.
41. Basu, K., Sinha, R., Ong, A., Basu, T. (2020) *Artificial Intelligence: How is It Changing Medical Sciences and Its Future?* Indian J Dermatol. 65(5):365-370. https://doi.org/10.4103/ijd.IJD_421_20. PMID: 33165420. PMCID: PMC7640807.
42. Yang, X., Tridandapani, S., Beitler, J.J., Yu, D.S., Yoshida, E.J., Curran, W.J., Liu, T. (2012) *Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: an in vivo study of late toxicity*. Med Phys. 39(9):5732-9. https://doi.org/10.1118/1.4747526. PMID: 22957638. PMCID: PMC3443195.
43. Taghanaki, S.A., Zheng, Y., Zhou, S.K., Georgescu, B., Sharma, P., Xu, D., Comaniciu D., Hamarneh, G. (2019) *Combo loss: Handlinginput and output imbalance in multi-organ segmentation.* Com-puterized Medical Imaging and Graphics, vol. 75, pp. 24–33 https://doi.org/10.1016/j.compmedimag.2019.04.005
44. Ouyang, D., He, B., Ghorbani, A, Yuan, N., Ebinger, J., Langlotz, C.P., Heidenreich, P.A., Harrington, R.A., Liang D.H., Ashley E.A., Zou, J.Y. (2020) *Video-based AI for beat-to-beat assessment of cardiac function. Nature*. 580(7802):252-256.

https://doi.org/10.1038/s41586-020-2145-8. PMID: 32269341. PMCID: PMC8979576.

45. Seaborn. (n.d). *Scatter Matrix.* Last Access: (2023, July 15). https://seaborn.pydata.org/examples/scatterplot_matrix.html
46. Riverbank Computing. (n.d). *PyQt6*. Last Access: (2023, July 8). https://www.riverbankcomputing.com/static/Docs/PyQt6/
47. OpenCV. (n.d). *OpenCV documentation.* Last Access: (2023, September 29). https://docs.opencv.org/4.x/d9/df8/tutorial_root.html
48. scikit-image. (n.d). *GLCM Texture Features. Last Access & Line, Circular and Elliptical Hough Transforms* Last Access: (2023, August 12). https://scikit-image.org/docs/stable/auto_examples/features_detection/plot_glcm.html# https://scikit-image.org/docs/stable/auto_examples/edges/plot_circular_elliptical_hough_transform.html
49. scikit-learn. (n.d). *Principal Component Analysis & Standard Scaler & k-Means.* Last Access: (2023, August 20). https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html
50. Analytics Vidhya. (2021)., *Elbow Method for Finding the Optimal Number of Clusters in K-Means.* Last Access: (2023, August 19) https://www.analyticsvidhya.com/blog/2021/01/in-depth-intuition-of-k-means-clustering-algorithm-in-machine-learning/
51. SciPy. (n.d.). *Hierarchy Dendrogram.* Last Access: (2023, August 19) https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.dendrogram.html
52. towardsdatascience.com. (n.d.). *Breaking down the agglomerative clustering process & Activation Functions in Neural Networks.* Last Access: (2023, September 20) https://towardsdatascience.com/breaking-down-the-agglomerative-clustering-process-1c367f74c7c2 https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

53. medium.com. (2020). *Bisecting Kmeans Clustering.* Last Access: (2023, August 20) https://medium.com/@afrizalfir/bisecting-kmeans-clustering-5bc17603b8a2
54. Science Direct. (n.d.). Manhattan Distance & Mean Square Error. Last Access: (2023, August 20)
https://www.sciencedirect.com/topics/mathematics/manhattan-distance
https://www.sciencedirect.com/topics/engineering/mean-square-error
55. scikit-learn. (n.d). *Selecting the number of clusters with silhouette analysis on KMeans clustering*. Last Access: (2023, August 20).
https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html
56. h5py. (n.d.). *Datasets.* Last Access: (2023, August 22).
https://docs.h5py.org/en/stable/high/dataset.html
keras (n.d.). Adam. Last Access: (2023, October 1).
https://keras.io/api/optimizers/adam/
57. Simonyan, K., Zisserman, A. (2014) *Very Deep Convolutional Networks for Large-Scale Image Recognition.* Arxiv.org.
https://doi.org/10.48550/arXiv.1409.1556
58. He K., Zhang, X., Ren, S., Sun J. 2016. *"Deep Residual Learning for Image Recognition"*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778
https://doi.org/10.1109/CVPR.2016.90.

Appendices

Appendix A

STANFORD UNIVERSITY SCHOOL OF MEDICINE ECHONET-PEDIATRIC DATASET RESEARCH USE AGREEMENT

By registering for downloads from the EchoNet-Pediatric Dataset, you are agreeing to this Research Use Agreement, as well as to the Terms of Use of the Stanford University School of Medicine website as posted and updated periodically at http://www.stanford.edu/site/terms/.

1. Permission is granted to view and use the EchoNet-Pediatric Dataset without charge for personal, non-commercial research purposes only. Any commercial use, sale, or other monetization is prohibited.

2. Other than the rights granted herein, the Stanford University School of Medicine ("School of Medicine") retains all rights, title, and interest in the EchoNet-Pediatric Dataset.

3. You may make a verbatim copy of the EchoNet-Pediatric Dataset for personal, non-commercial research use as permitted in this Research Use Agreement. If another user within your organization wishes to use the EchoNet-Pediatric Dataset, they must register as an individual user and comply with all the terms of this Research Use Agreement.

4. YOU MAY NOT DISTRIBUTE, PUBLISH, OR REPRODUCE A COPY of any portion or all of the EchoNet-Pediatric Dataset to others without specific prior written permission from the School of Medicine.

5. YOU MAY NOT SHARE THE DOWNLOAD LINK to the EchoNet-Pediatric dataset to others. If another user within your organization wishes to use the EchoNet-Pediatric Dataset, they must register as an individual user and comply with all the terms of this Research Use Agreement.

6. You must not modify, reverse engineer, decompile, or create derivative works from the EchoNet-Pediatric Dataset. You must not remove or alter any copyright or other proprietary notices in the EchoNet-Pediatric Dataset.

7. The EchoNet-Pediatric Dataset has not been reviewed or approved by the Food and Drug Administration, and is for non-clinical, Research Use Only. In no event shall data or images generated through the use of the EchoNet-Pediatric Dataset be used or relied upon in the diagnosis or provision of patient care.

8. THE ECHONET-PEDIATRIC DATASET IS PROVIDED "AS IS," AND STANFORD UNIVERSITY AND ITS COLLABORATORS DO NOT MAKE ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, NOR DO THEY ASSUME ANY LIABILITY OR RESPONSIBILITY FOR THE USE OF THIS ECHONET-PEDIATRIC DATASET.

9. You will not make any attempt to re-identify any of the individual data subjects. Re-identification of individuals is strictly prohibited. Any re-identification of any individual data subject shall be immediately reported to the School of Medicine.

10. Any violation of this Research Use Agreement or other impermissible use shall be grounds for immediate termination of use of this EchoNet-Pediatric Dataset. In the event that the School of Medicine determines that the recipient has violated this Research Use Agreement or other impermissible use has been made, the School of Medicine may direct that the undersigned data recipient immediately return all copies of the EchoNet-Pediatric Dataset and retain no copies thereof even if you did not cause the violation or impermissible use.

In consideration for your agreement to the terms and conditions contained here, Stanford grants you permission to view and use the EchoNet-Pediatric Dataset for personal, non-commercial research. You may not otherwise copy, reproduce, retransmit, distribute, publish, commercially exploit or otherwise transfer any material.

Limitation of Use

You may use EchoNet-Pediatric Dataset for legal purposes only.

You agree to indemnify and hold Stanford harmless from any claims, losses or damages, including legal fees, arising out of or resulting from your use of the EchoNet-Pediatric Dataset or your violation or role in violation of these Terms. You agree to fully cooperate in Stanford's defense against any such claims. These Terms shall be governed by and interpreted in accordance with the laws of California.

Appendix B

Sample of FileList.csv

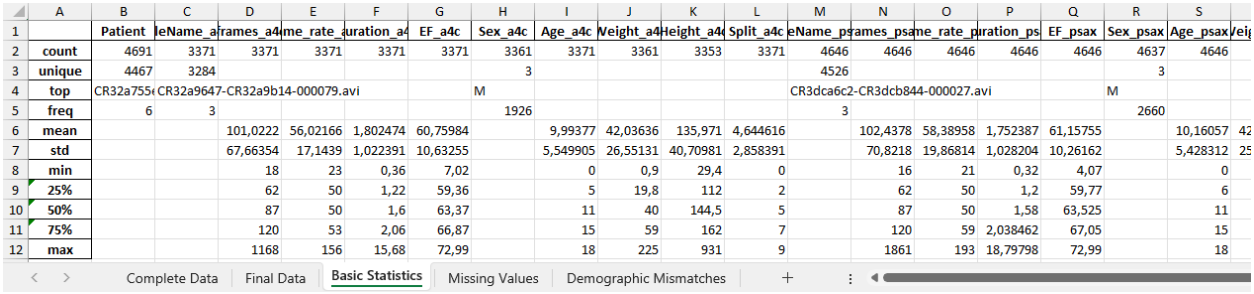
```
FileName,EF,Sex,Age,Weight,Height,Split
CR32a7555-CR32a7582-000039.avi,40.83,F,0,10.2,68.5,5
CR32a7555-CR32a97af-000033.avi,52.62,F,1,15.5,85.0,5
CR32a7555-CR32a97e1-000024.avi,24.85,F,0,4.0,56.0,5
CR32a7555-CR32a9850-000040.avi,50.96,F,4,18.0,99.0,5
CR32a7555-CR32a988d-000034.avi,56.76,F,0,13.2,75.0,5
```

Sample of VolumeTracings.csv

```
FileName,X,Y,Frame
CR32a7555-CR32a7582-000039.avi,58,58,39
CR32a7555-CR32a7582-000039.avi,53,51,39
CR32a7555-CR32a7582-000039.avi,48,42,39
CR32a7555-CR32a7582-000039.avi,49,41,39
CR32a7555-CR32a7582-000039.avi,49,41,39
CR32a7555-CR32a7582-000039.avi,49,40,39
```

Appendix C

Sample of statistics excel file:

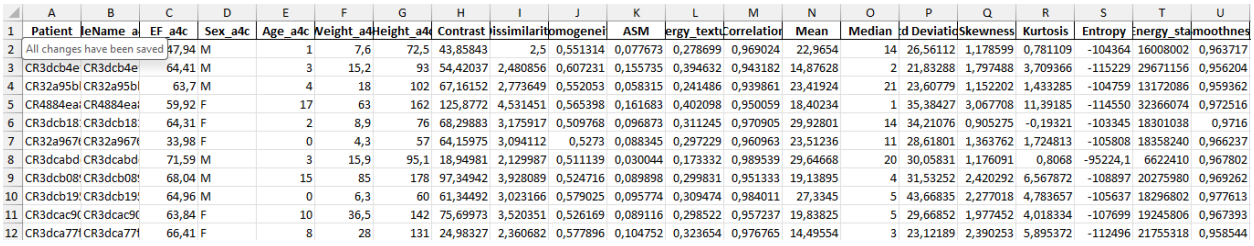


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		Patient	leName_a	rames_a4c	me_rate_a	uration_a4	EF_a4c	Sex_a4c	Age_a4c	Veight_a4c	Height_a4c	Split_a4c	eName_ps	ames_psax	me_rate_p	uration_ps	EF_psax	Sex_psax	Age_psax	Veig
2	count	4691	3371	3371	3371	3371	3371	3361	3371	3361	3353	3371	4646	4646	4646	4646	4646	4637	4646	
3	unique	4467	3284					3					4526					3		
4	top	CR32a755	CR32a9647-CR32a9b14-000079.avi					M					CR3dca6c2-CR3dcb844-000027.avi					M		
5	freq	6	3					1926					3					2660		
6	mean			101,0222	56,02166	1,802474	60,75984		9,99377	42,03636	135,971	4,644616		102,4378	58,38958	1,752387	61,15755		10,16057	42
7	std			67,66354	17,1439	1,022391	10,63255		5,549905	26,55131	40,70981	2,858391		70,8218	19,86814	1,028204	10,26162		5,428312	25
8	min			18	23	0,36	7,02		0	0,9	29,4	0		16	21	0,32	4,07		0	
9	25%			62	50	1,22	59,36		5	19,8	112	2		62	50	1,2	59,77		6	
10	50%			87	50	1,6	63,37		11	40	144,5	5		87	50	1,58	63,525		11	
11	75%			120	53	2,06	66,87		15	59	162	7		120	59	2,038462	67,05		15	
12	max			1168	156	15,68	72,99		18	225	931	9		1861	193	18,79798	72,99		18	

Complete Data | Final Data | Basic Statistics | Missing Values | Demographic Mismatches

Appendix D

Sample of the extracted featurs excel file:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Patient_	leName_a	EF_a4c	Sex_a4c	Age_a4c	Veight_a4	Height_a4	Contrast	issimilarit	omogenei	ASM	ergy_textu	Correlatior	Mean	Median	d Deviatic	Skewness	Kurtosis	Entropy	Energy_sta	moothnes
2	All changes have been saved		47,94	M	1	7,6	72,5	43,85843	2,5	0,551314	0,077673	0,278699	0,969024	22,9654	14	26,56112	1,178599	0,781109	-104364	16008002	0,963717
3	CR3dcb4e	CR3dcb4e	64,41	M	3	15,2	93	54,42037	2,480856	0,607231	0,155735	0,394632	0,943182	14,87628	2	21,83288	1,797488	3,709366	-115229	29671156	0,956204
4	CR32a95b	CR32a95b	63,7	M	4	18	102	67,16152	2,773649	0,552053	0,058315	0,241486	0,939861	23,41924	21	23,60779	1,152202	1,433285	-104759	13172086	0,959362
5	CR4884ea	CR4884ea	59,92	F	17	63	162	125,8772	4,531451	0,565398	0,161683	0,402098	0,950059	18,40234	1	35,38427	3,067708	11,39185	-114550	32366074	0,972516
6	CR3dcb18	CR3dcb18	64,31	F	2	8,9	76	68,29883	3,175917	0,509768	0,096873	0,311245	0,970905	29,92801	14	34,21076	0,905275	-0,19321	-103345	18301038	0,9716
7	CR32a967	CR32a967	33,98	F	0	4,3	57	64,15975	3,094112	0,5273	0,088345	0,297229	0,960963	23,51236	11	28,61801	1,363762	1,724813	-105808	18358240	0,966237
8	CR3dcabd	CR3dcabd	71,59	M	3	15,9	95,1	18,94981	2,129987	0,511139	0,030044	0,173332	0,989539	29,64668	20	30,05831	1,176091	0,8068	-95224,1	6622410	0,967802
9	CR3dcb08	CR3dcb08	68,04	M	15	85	178	97,34942	3,928089	0,524716	0,089898	0,299831	0,951333	19,13895	4	31,53252	2,420292	6,567872	-108897	20275980	0,969262
10	CR3dcb19	CR3dcb19	64,96	M	0	6,3	60	61,34492	3,023166	0,579025	0,095774	0,309474	0,984011	27,3345	5	43,66835	2,277018	4,783657	-105637	18296802	0,977613
11	CR3dcac9	CR3dcac9	63,84	F	10	36,5	142	75,69973	3,520351	0,526169	0,089116	0,298522	0,957237	19,83825	5	29,66852	1,977452	4,018334	-107699	19245806	0,967393
12	CR3dca77	CR3dca77	66,41	F	8	28	131	24,98327	2,360682	0,577896	0,104752	0,323654	0,976765	14,49554	3	23,12189	2,390253	5,895372	-112496	21755318	0,958544