Automated assessment of echocardiographic image quality using deep convolutional neural networks

Automated Assessment of Echocardiographic Image Quality Using Deep Convolutional Neural Networks

Robert B. Labs

A Thesis submitted in partial fulfilment of the requirements of
The University of West London for the degree of Doctor of Philosophy

Supervisors:

Professor Massoud Zolgharni,
Professor of Computer Vision, School of Computing and Engineering

Professor Jonathan Loo,
Chair Professor of Computing and Communication Engineering,
School of Computing and Engineering

January 2022

Acknowledgements

It is with immense gratitude that I acknowledge the support and leadership of my supervisors, Prof. Massoud Zolgharni and Prof. Jonathan Loo, for their insightful guidance and inspiring knowledge. I would also like to thank the entire members of IntSaV research group at UWL for their profitable feedback on my research projects.

To my family and friends for their persistent encouragement and love. Special thanks to the Duchess of Erith, Edith Labs, and to the Erinmoje's royals 'Bolurin, and 'Tinuke who gave me limitless support and latitude to achieve this academic landmark. Also, to the Erinmoje's Princess, Prof. Folasade Labs-Kolade, for her inspirational inputs and to my Uncle Engr. John Alade, who in time past provided parental guidance.

I will not forget my mentors, John T. Akinshiku, Dr. Isaiah D. Lawon, Dr. William F. Kumuyi, Prof. John O. Adeoti, Prof. Tunde Adegbola, and Dr. Timothy O. Nwasike, whose words of wisdom and lifestyles had instilled moral discipline and academic excellence, stirring up plausible outcomes in the spheres of my endeavour. I profoundly acknowledge.

Finally, I would like to thank Prof. Darrel Francis at Imperial College London for providing a valuable cardiac dataset and Prof. Zolgharni who provided the opportunity for clinical experience. I am indebted to friends who has made significant contributions in finance and moral support to the successful realisation of this important milestone. Apologies to those I could not mention by name.

This thesis is dedicated to the remembrance of my parents and grandparents of the Erinmojes Imperial Majesty and to those who by nature, could improve life-saving technology for the benefit of mankind.

Abstract

Myocardial ischemia tops the list of causes of death around the globe, but its diagnosis and early detection thrives on clinical echocardiography. Although echocardiography presents a huge advantage of a non-intrusive, low-cost point of care diagnosis, its image quality is inherently subjective with strong dependence on operators' experience level and acquisition skill. In some countries, echo specialists are mandated to supplementary years of training to achieve 'gold standard' free-hand acquisition skill without which exacerbates the reliability of echocardiogram and increases possibility for misdiagnosis. These drawbacks pose significant challenges to adopting echocardiography as authoritative modalities for cardiac diagnosis. However, the prevailing and currently adopted solution is to manually carry out quality evaluation where an echocardiography specialist visually inspects several acquired images to make clinical decisions of its perceived quality and prognosis. This is a lengthening process and laced with variability of opinion consequently affection diagnostic responses. The goal of the research is to provide a multi-discipline, state-of-the-art solution that allows objective quality assessment of echocardiogram and to guarantee the reliability of clinical quantification processes. Computer graphic processing unit simulations, medical imaging analysis and deep convolutional neural network models were employed to achieve this goal. From a finite pool of echocardiographic patient datasets, 1650 random samples of echocardiogram cine-loops from different patients with age ranges from 17 and 85 years, who had undergone echocardiography between 2010 and 2020 were evaluated. We defined a set of pathological and anatomical criteria of image quality by which apical-four and parasternal long axis frames can be evaluated with feasibility for real-time optimization. The selected samples were annotated for multivariate model developments and validation of predicted quality score per frame. The outcome presents a robust artificial intelligence algorithm that indicate frames' quality rating, real-time visualisation of element of quality and updates quality optimization in real-time. A prediction errors of **0.052, 0.062, 0.069, 0.056** for visibility, clarity, depth-gain, and foreshortening attributes were achieved, respectively. The model achieved a combined error rate of 3.6% with average prediction speed of 4.24 ms per frame. The novel method established a superior approach to two-dimensional image quality estimation, assessment, and clinical adequacy on acquisition of echocardiogram prior to quantification and diagnosis of myocardial infarction.

Table of Contents

List of Figures

List of Tables

Acronyms

2D	Two Dimensional
3D	Three Dimensional
A2C	Apical Two-Chamber
A4C	Apical Four-Chamber
AI	Artificial Intelligence
ASE	American Society of Echocardiography
AVC	Aortic Valve Closure
BHF	British Heart Foundation
BM	Block Matching Cardiac
CT	Cardiac Computed Tomography
CNN	Convolutional Neural Network
DARTS	Differentiable Architecture Search
EACVI	European Association of Cardiovascular Imaging
ECG	Electrocardiogram
ED	End-Diastolic
EF	Ejection Fraction
ENAS	Efficient Neural Architecture Search
ES	End-Systolic
FC	Fully Connected
FCN	Fully Convolution Network

GLS	Global Longitudinal Strain
GT	Ground-Truth
HD	Harsdorf Distance
KF	Kalman Filter
LSTM	Long short-term memory
LV	Left Ventricle
MRI	Magnetic Resonance Imaging
MVC	Mitral Valve Closure
NAS	Neural Architecture Search
PACS	Picture Archiving and Communication Systems
ReLU	Rectified Linear Unit RL Reinforcement Learning
RNN	Recurrent Neural Network
RV	Right Ventricle
SemiNAS	Semi-automated Neural Architecture Search
SGD	Stochastic Gradient Descent
TD	Transitions Down
TDI	Tissue Doppler Imaging
TEE	Transoesophageal Echocardiogram
TTE	Transthoracic Echocardiography

Chapter 1

Introduction

In the diagnosis of cardiovascular diseases around the world, transthoracic echocardiography (TTE) examination has become the most prevalent method of imaging tool for the assessment of myocardium and clinical quantifications (Mitchell *et al.*, 2019). This is due to its low-cost (Wang *et al.*, 2018), non-ionizing, nonintrusive and in-vivo examination properties. Hence, echocardiography is essential to modern cardiologists as the first-choice imaging tool for point-of-care diagnosis. Ultrasound imaging is portable, poses little or very minimal risk to health and provides clinical capacity for both dynamic or static visualisation but laden with varied human interpretations and sub-optimal image quality as its current obvious limits. Consequently, it is necessary to clinically assess the quality of images before clinical quantification is performed.

In a TTE workflow illustrated in Figure 1.1, the evaluation of echocardiographic image's quality is a standard clinical practice that underpins the reliability of diagnosis of cardiovascular diseases (Sassaroli *et al.*, 2019). It's considered as a non-trivial process, which is carried out manually all in a bid to reduce quantification error, misdiagnosis, and wrong classification of patient's needs. Nevertheless, the success of clinical quantification depends on the quality of the image obtained during manual ultrasound scans which in turn depends on the operators' experience (Labs, Zolgharni and Loo, 2021). Currently, the method of image quality assessment is a highly subjective process, where an echocardiography specialist visually inspects the images and rates an image based on certain features, such as wall definition and clarity of anatomical details in the image. Detecting complex heart abnormalities and interpreting hearts' anatomical and pathological features, requires the consideration of many cardiac images acquired in a clinical workflow. Although, the dynamic nature of echocardiography presents technical challenges in quality assessment beyond those of static images from X-ray, computed

tomography, magnetic resonance, and radioisotope imaging (Yoon, Kim and Chang, 2021) nevertheless, the adoption of manual evaluation process did not spare further challenges that elicit opinion variability and reduce diagnostic reliability of echocardiograms TTE procedures (Nagata *et al.*, 2018). Consequently, a more responsive method that can be integrated in a unified workflow, where real-time evaluation is done objectively and automatically is thus required.

In this chapter, an overview of echocardiography's quality image acquisition and the problem statement for automating such procedures are presented. Furthermore, the motivation, main aim and objectives, and the contributions of this research are enumerated. Finally, the thesis outline and research consortium are introduced.

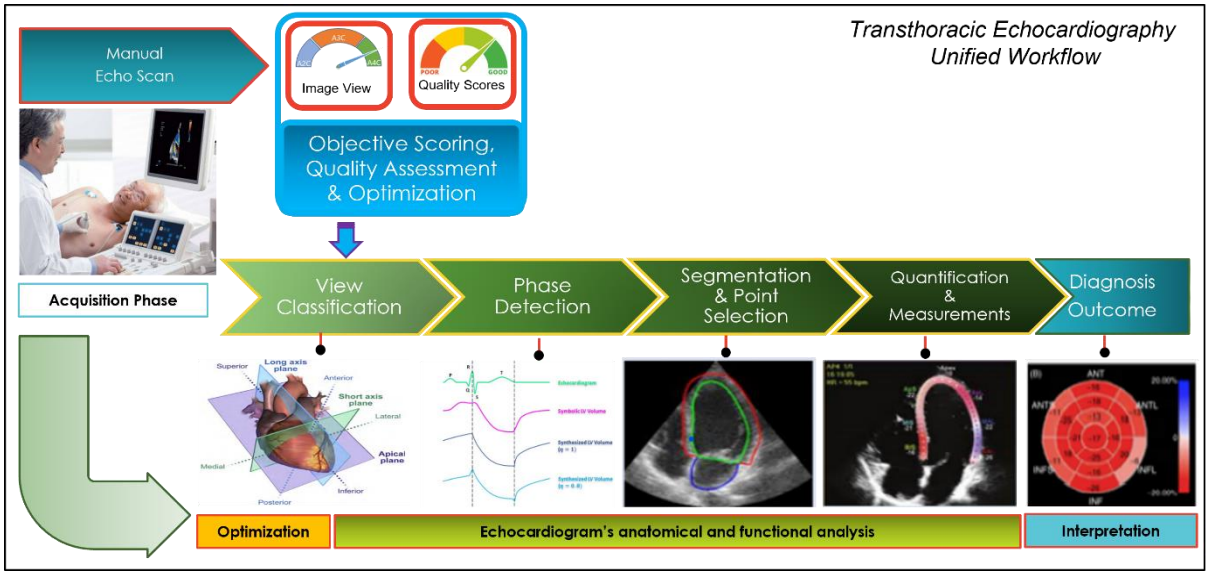


Figure 1.1: Automated unified workflow process (echo scan to diagnosis) is entirely manual and involves different layers of clinical experts. The new unified workflow can be automated to bridge the gap between acquisition and analysis which include automated quality scoring system, real-time assessment, and operators' optimization feedback. This is known as real-time optimization of 2D cardiac specimen quality assessment.

1.1 Clinical Context and Problem Statement

The application of clinical echocardiography in modern healthcare has become more prominent because it presents rich anatomical details of the myocardium, also for its many pathological and low cost of ownership advantages. During transthoracic examination, echocardiograms must present an accurate representation of the myocardium, consequently, several shots are

required and obtained to enable cardiologists in building an encompassing picture for subsequent analysis and diagnosis. Nevertheless, images produced through scattering centres do not come with crisp edges (Labs *et al.*, 2020), yielding high possibility for poor quality images irrespective of ultrasound equipment in use. Since, interpreting cardiac functions is done through whatever quality of echocardiograms obtained, poor images have continually hampered cardiac measurement and cardiovascular diagnosis. The absence of integrated tools and coherent industrial standardisation to prevent sub-optimal acquisition (Benacerraf et al., 2018; Zhou et al., 2018) consequently necessitated the adoption of individually approved standard protocol which is prevalent in different clinical laboratories. Acquisition of echocardiographic images therefore requires significant experts' skill which vary with operators' experience and patients' pathological profiles. Nevertheless, the impact of sub-optimum image quality, admitted by (Nagata, Y., et al 2018) remains critical and precipitates measurement variability that leads to cardiovascular misdiagnosis. In a bid to enforce reliability standards, cardiac image quality assessment is considered paramount, but is currently assessed manually around the globe. Unfortunately, manual quality assessment introduces yet another significant drawback (Nagata *et al.*, 2018); (Liao *et al.*, 2019) that catalyse opinion variability, additional years of training for cardiologists, increase in demand for highly trained operators, and low trigger response to patient care. Since echocardiography view is significant to cardiologists but cumbersome, how can automated image identification accelerate cardiologists' workflow? And how can objective assessment sufficiently guarantee reliability in cardiac imaging when applied to transthoracic examination protocol? Although, these core questions are apparent but further raise the need to provide contrasting understanding of the constituents of an image quality with respect to two-dimensional echocardiogram and clinical quantifications. It is hoped that the solution will encompass the aspect of a two-dimensional image quality assessment and real-time optimization of echocardiographic image quality to mention but few. The impact of such solution will be significant during medical emergencies and routine echo exams in the health sector.

Consequently, the focus of this academic dissertation is based on 2D echocardiogram's image quality assessment rather than 3D or other imaging modalities. Below is the succinct overview of several steps involved in the implementation of automated assessment of echocardiographic image quality using deep convolutional neural networks.

1.1.1 Ultrasound and Scan Process

Each cardiac investigative process requires an operator taking multiple echocardiogram images of a patient which are sent to a cardiologist for careful evaluation. The equipment used in TTE is referred to as simply 'Ultrasound' which comes in various sizes depending on its intended application are employed for myocardial assessment as illustrated in Figure 1.2. During TTE exam, several cardiac images from different relevant angles are obtained for analysis and interpretation. This becomes part of the cardiologists' analytical and clinical workflow, illustrated in Figure 1.1. Consequently, the acquired images come in mixture of several differently apical standard planes (Figure 1.3), and views (e.g., A2C, A4C, PLAX, etc.) illustrated in Figure 1.4, to build a complete clinical opinion for diagnosis (Mitchell *et al.*, 2019); (Thomas *et al.*, 2005). Nevertheless, quantification of cardiac functions, especially myocardial ischemia, strains detection and the likes, the apical view and parasternal long axis (PLAX) are considered the most paramount to cardiologists' workflow (Lang *et al.*, 2015). These preferred views are unanimously recommended by American Society of Echocardiography (ASE), because each of the preferred views presents a combined view of the ventricles, valves, and wall cavities essential for quantification and clinical measurements.

Figure 1.2 Overview of transthoracic echocardiography assessment process under clinical protocol.

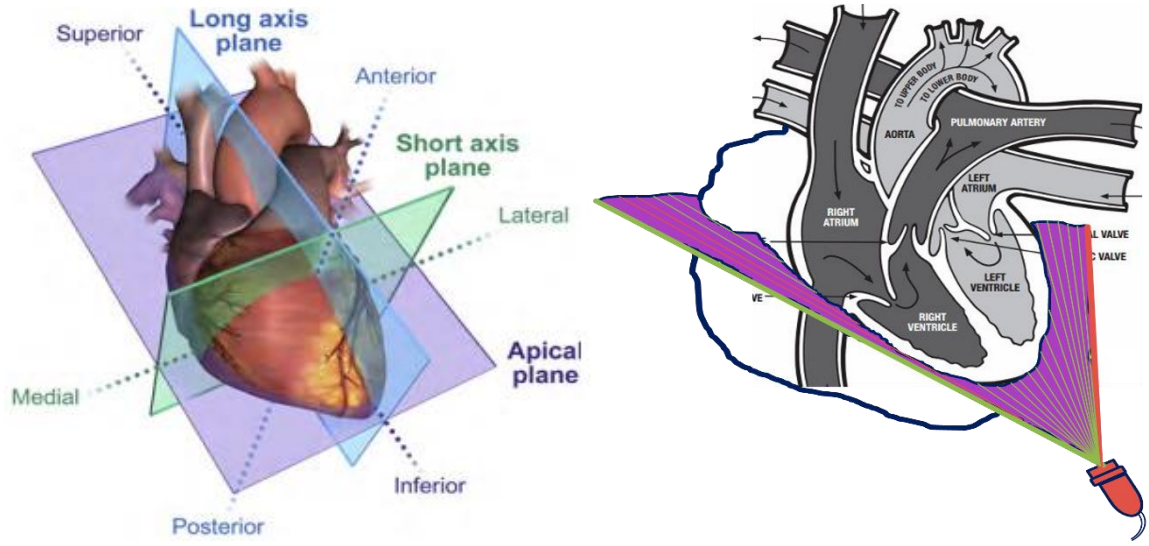


***Figure 1.3**: Illustration of Apical standard planes commonly associated with cardiovascular assessment and pseudo-echocardiogram image showing four apical chambers with respective valves.*

Nevertheless, the process of scanning and image interpretation require manual human intervention, experience, and skilful art to avoid aperture obstructions, echogenicity and suboptimal image acquisition and perform critical optimization of image quality. However, the ultrasound manual scanning process is slow and can be unreliable during emergency call outs, or with multiple patients, thereby inadvertently reducing the possibility for early detection and acknowledgement of cardiovascular abnormalities and priority for patient care. Group of expert cardiologists manually shifting through hundreds or thousands of cardiac specimens per patient, to build prognosis which sometimes are inconclusive, or a re-examination further requested, represents significant overhead that can be mitigated by implementing automatic assessment and sample characterization pipeline. Therefore, the development of automated view classification, objective characterization of cardiac specimens, automated assessment of 2D image quality and real-time acquisition optimization guidance are highly desirable and discussed in Chapters 4, 5, 6, and 7.

1.1.2 Cardiac Image View Classification

In clinical practice, echocardiography examinations follow standard protocols requiring various views of the heart's anatomy. Acquisition of different parts of the myocardium requires sampling images from multiple windows (Lang *et al.*, 2015) as illustrated in Figure 1.3. Each window allows specific anatomical features to be obtained by the transducer position. This

would be discussed in detail in Chapter 4. Nevertheless, the analysis and interpretation of echocardiographic images begins with view identifications, where the acquired image specimen is matched to its corresponding cardiac views. Assessment of cardiac samples is currently done by manual process and raises the question of reliability and consistency when a human agent sweeps through hundreds of cardiac samples.

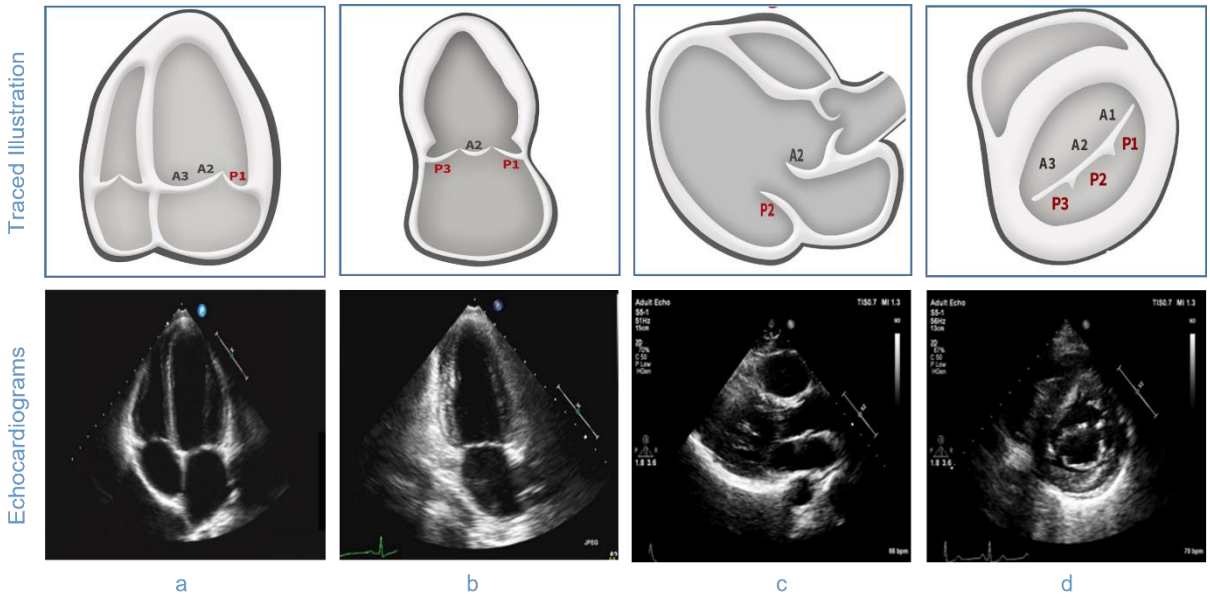


***Figure 1.4**: Examples of cardiac views in transthoracic echocardiography showing corresponding traced samples with respective echocardiogram specimen: a: apical four-chamber left ventricle focused (A4C-LV), b: apical two-chamber (A2C), c: Full parasternal long-axis (PLAX-Full), and d: Parasternal short-axis with left ventricle focused (PSAX-LV). Figure created from (Lynch and Jaffe, 2006)*

Usually, the appearance of acquired cardiac specimens are heterogeneous and different anatomical properties as illustrated in Figure 1.3. Firstly, echocardiogram image views would depend on patients' physical characteristics and pathological factors. Also, patients' supine or upright position might introduce subtle variations in echogenicity. Since there is no specific marked area to place the transducer on the patient's body, the appearance-based method cannot be applied for the view classification issue as asserted by (G. N. Balaji, Subashini and Chidambaram, 2015). Therefore, automatic classification of heart's views potentially would streamline cardiac workflow by aiding clinicians in reducing the inter-user discrepancy, improving the accuracy for high throughput of echocardiogram data, and clinical diagnosis. The task of automating cardiac views was recently applied to software packages, such as EchoPAC and QLAB (Philips), however, they still require some level of human involvement in detecting relevant views. The involvement of human interaction makes it near impossible to detect

structural features especially in a fast-moving pixelating cardiac frame, and discerning high levels of background noise which is associated with echocardiographic frames.

Therefore, an objective automation of view classification would serve a crucial aspect of analysis and be beneficial for pseudo-labelling a large database of unclassified image samples (Khamis, Zurakhov et al., 2017). Automation of cardiac view detection was carried out in our research group, (IntSav)[3], and my contribution was acknowledged as a co-author. Automated frame classification would be further discussed in Chapter 4.

1.1.3 Characterization of Cardiac Specimens

Objective characterization of two-dimensional echocardiograms is one of the potential approaches to define global characteristics of domain-specific elements of cardiac specimens in terms of anatomical and pathological features (Labs *et al.*, 2020). Anatomy of cardiovascular specimens presents enormous complexity to objective functions in terms of dynamic features identification and clinical quantifications. This is due to clinical protocol where cardiologists not only rely on still images but a fast-moving echocardiogram frame in real-time. Since several images are required to build a complete picture of patients' pathology and summary of diagnosis (Mitchell *et al.*, 2019); therefore, several specimens for multiple patients would indicate the enormous task of manually assessing several hundreds of specimens prior to diagnosis. But an automated system that can objectively allow cardiologist filter-search (e.g., image element-wise search) and group specific specimen down to pathological and anatomical relevance (described in chapter **5**) is certainly beyond the scope of apical view classification. This system could allow rapid assessment of functional, anatomic, and pathological features present in each specimen. Although the characterization focus is on apical-four specimens, the method can be applied to other standard planes and views.

Currently, there exist no automated solution that offers a comprehensive detailing of pathological and anatomical feature detection using twenty-one (21) different domain-specific criteria per cardiac specimen. These are capable of aiding cardiologists' rapid assessment, sorting and element-wise search of quality class of echocardiographic images before final diagnosis. However, plausible solution using deep learning model have been widely demonstrated for cardiac view classification (Abdi, Luong, Tsang, Allan, *et al.*, 2017), quality assessment and now it's been applied to achieve comprehensive image characterization and

pseudo-labelling of pathological semi-supervised deep learning model (Chen *et al.*, 2020); (Ouali *et al.*, 2020). Automation of cardiac view detection will be discussed in Chapter 5.

1.1.4 Automated Assessment of 2D Image Quality

A two-dimensional echocardiographic image quality is a domain knowledge paramount to myocardial reliable diagnosis. Images produced though scattering centres do not come with crisp edges unlike natural images. Therefore, acquisition of echocardiographic images requires significant experts' skills which vary with patients' pathological profiles. In clinical transthoracic examination workflow, strong indication for quantification of systolic function in apical four (A4C) and parasternal long axis (PLAX) view is a recommended standard practice (Rudski *et al.*, 2010). According to (Lang *et al.*, 2015), these and A2C are recommended quantification standards in clinical practice because their spatial orientations are congruent in nature, thus offering complementary advantages on heart functional measurement. Nonetheless, the clinical process of image acquisition comes with inherent challenges of operator skill/experience which cascade many successive analytical problems including diagnostic reliability output. These drawbacks remain significant, consequently inhibiting the adoption of echocardiograms as a reliable imaging modality for cardiac diagnosis despites its many advantages. Currently, echocardiogram's inherent poor image quality is exacerbated by operators' acquisition skills. Unlike photographic imaging, the benchmark standard for objective constituents of quality in echocardiography remains largely undefined. Consequently, a subjective process where image quality is manually assessed to determine its clinical and pathological relevance becomes the de facto in clinical workflow. Although, this process is time consuming, laboriously expensive, and unfortunately introduces major drawbacks that precipitate high demand for expert's and highly trained operators. Furthermore, the impacts of subjective standard and manual assessment in echocardiograms have been well documented in research literatures, a recent work by (Yoon, Kim and Chang, 2021) aggregated the impact of subjective quality assessment to clinical misdiagnosis and poor response trigger to patient care. Consequently, an automated assessment is thus required for consistency in image quality, reliability, and provide efficient assessment of real-time quality optimization. Moreover, an objective assessment system would:

1. provide reproducible research, quantitative and consistent methods for optimum image acquisition, reliable clinical quantification, and better patient care.

2. provide quantitative information on the adequacy of the images obtained, a valuable real-time tool for researchers, clinical practitioners and
3. find its significant usefulness as an independent arbitration, either in retrospect or real-time for global standardisation and benchmarking in echocardiographic imaging.

It can also provide an independent measure of the reliability a very low image quality index would indicate inadequate images, which may be due to poor acoustic windows or patients with severe abnormalities

An objective and quantitative method for image quality assessment is a useful component for an operator guidance system, as well as a valuable tool for research and clinical practice. As part of an operator guidance system, it can provide quantitative information on the adequacy of the images obtained. Therefore, the development of a fully automated, reliable, and reproducible pipeline is highly desirable. Automation of image quality assessment will be discussed in Chapter 5.

1.1.5 Acquisition and Optimization Guidance

Echocardiograms' image quality is paramount to accurate quantification and linear measurements of the left ventricle (LV) systolic functions. Accurate linear measurements of the LV systolic functions provide major clues to a healthy heart which triggers a corresponding diagnostic response for patient cares (Lang *et al.*, 2015). While sub-optimal quality image is common in echocardiogram workflow, its impact usually cascades successive issues in measurement variability, quantification reliability, and subsequent misdiagnosis (Nagata *et al.*, 2018); (Kurt *et al.*, 2009). Consequently, operator experience plays a significant role in the acquisition of high-quality images prior to quantification and lack of probe skilful manoeuvring could exacerbate the possibility for optimal image quality. Since no two cases are the same in TTE, varying anatomical pathologies and dissimilar image acquisition skills have added significant complexities to image quality (Aschkenasy *et al.*, 2006). This is a well-known challenge that currently limits cardiac automation workflow to a manual or at best, semi-automated domain. Moreover, the possibility of hardware's limitation in performance throughput, and acquisition environment conditions are essential factors that implicitly impair acquisition of optimum quality images.

To address the some of the problems relating to echocardiographic image quality, optimal acquisition, reliability, and operator's competency, are several research efforts geared towards automated assessment of image quality. Almost all the existing implementations of such a system were based on weighted average scoring (Liao *et al.*, 2019); (Abdi, Luong, Tsang, Jue, *et al.*, 2017a) where operators are fed with a continuous value representing objective quality score. However, the problem with such a system of weighted average score index with a single value score (even though compared relatively to experts' ground truth score) is grossly incapable of guiding operators to which aspect of quality is lacking in the overall assessment. Hence, practical deployability of such a system of assessment is only limited to experimental demonstration instead of clinical advantages.

Furthermore, some countries including the United Kingdom, mandates operators and clinicians to take additional years of training courses on echocardiography image acquisition ('SCoR & BMUS Guidelines 2020', 2020). This is envisaged to provide a level of expertise and clinical confidence in the diagnostic art of ultrasonography. Currently, a subjective, manual quality control process under the control of a sonographer appears to be the prevailing solution. This is performed during the acquisition phase, unfortunately introduces major drawbacks which impair diagnosis and patient's care. Consequently, the acquisition of optimum image quality can only be enforced using objective measures rather than human subjective process. A system with multiple objectives scoring that exhibits experts' opinion in the domain-specific areas would make an ideal assessment method for real-time optimization, thus providing specific feedback, and quality grading score for image adequacy and optimization. Moreover, it would allow prospective clinician to learn on the job rather than spending many more years to acquire gold standard skill in image acquisition and optimization.

A significantly valuable tool for researchers, medical health practitioners in the independent measure for global standardisation and benchmarking in echocardiography. Real-time operator's optimization guidance will be discussed Chapter 7.

1.2 Motivation

Cardiovascular diagnosis, interpretation and intervention procedures using computer-aided systems is a fast-growing field of research with infinite potentials to assist clinicians in obtaining accurate measurements and identify anomalies with persistent reliability, precision, and greater speed. Besides echocardiography as cardiac image assessment tool, many aspects of medical imaging and diagnosis have moved to the automatic domain and benefiting from automatic quality assessment, but echo images quantification, prognosis and diagnosis has remained largely with cardiologist's subjectivity and varying interpretations. The issues have largely been affected by lack of consistency of standards, lack of performance evaluation, repeatability and large user variability during image acquisition and post processing phases.

Fortunately, there are now some promising scientific tools which are based on objective metrics; more recent objective quality assessment (IQA) using deep convolutional neural network methodology (DCNN), (Abdi, Luong, Tsang, Allan, *et al.*, 2017) have received plausible recommendations. The advent of deep learning architecture models provides multifaceted opportunity and motivation to employ artificial intelligence algorithms in solving the most prominent causes of death of our time (*World Health Organization: 2019*).

As already a common knowledge that DCNN model relies on effective but large dataset, obtaining a pool of large cardiac dataset, which was highly personalised and prohibited dues to UK's data protection act, could now be obtained under ethical approval from Health Regulatory Agency, paving way for exploratory and translatory research in clinical applications and cardiac diagnosis. Also, the research latches on the availability of high-performance computing (HPC) hardware and consortium of clinical experts to obtain detailed hemodynamic and echocardiographic information along with the domain-specific reviews to the extent of gaining full understanding of the clinical manifestations of normal or abnormal cardiac functions.

1.3 Aims and Objectives

This PhD project forms part of a larger project which aims at developing a reliable and robust pipeline to automate acquisition, analysis, and interpretation of cardiac unified workflow. At micro level, the automation of cardiac standard view identification, and image quality assessment over acquisition with real-time optimisation is the core objective of this research work. A unified system, illustrated in Figure 1.1, thus represents a novel pipeline that automates

the process of optimization of manual cardiac acquisition, analysis, clinical measurement, and interpretation for the purpose of achieving reliable cardiac diagnosis.

This thesis has focused on creating domain-specific assessment and standard under the supervision and review of several cardiologists and clinicians, namely view classification, characterization of cardiac specimen, objective quality attributes, assessment methods, and operators' guidance for specific optimization of cardiac image quality in real-time. The pipeline and model should be a lightweight model since integration could be on mobile devices or limited memory space hardware, while processing at preferably high speed to allow system on chip (SoC) embedding and deployment in real-time applications or in retrospective process. Therefore, the main objectives of this study are listed as below:

- Develop and define coherent clinical and domain-specific attributes of 2D image quality required for objective assessment standards.
- Evaluate and develop a reproducible semi-supervised learning algorithm to extract anatomic features for specimen characterization and pseudo-labelling utility for global characteristics of cardiac specimen.
- Develop a quality assessment method that encapsulates multiple quality score indexes per cardiac frame. Solution to enable clinician to assess specific image quality in real-time acquisition or in post-acquisition phase.
- Develop a robust algorithm or pipeline modality to allow operators feedback for specific quality attributes, showing what aspect of image quality that needs to be improved and showing the optimised result in real-time before measurement and quantifications is performed.

1.4 Overview of Contributions

Considering the novel elements of the research undertaken, the main contributions of this thesis involve theoretical and practical element and can be summarised as follows:

(i) **View/Frame Detection**

- Application of the state-of-the-art neural network search (NAS) technique to design lightweight DCNN architectures.
- Large Dataset (PACS) with 14 different cardiac specimens of echocardiographic views; consisting of 41,321 images, currently larger than any previous studies.
- Analysis of model accuracy, computational accuracy, and performance of the developed models.
- Analysis of spatial impact of the input image resolution and size of training data on the model's performance.
- Analysis of the correlation between the image quality and accuracy of the model for view detection.

(ii) **Characterization of Cardiac Specimen**

- Analysis of independent dataset (EchoLAB - 27, 230) for domain-specific features characteristics and ground truth annotations
- Propose a comprehensive definition of clinically relevant (21) assessment criteria of A4C standard views and seven (7) objective characterization of A4C cardiac specimens.
- Development of DCNN semi-supervised pipeline for pseudo-label annotation process of unlabelled A4C cardiac specimens.
- Repository release of the complete annotated and characterization (A4C) patient dataset to allow future studies and external validation of the new approach or methods.

(iii) **Automated Assessment of 2D Image Quality**

- Proposed novel four (4) objective quality attributes for cardiac specimens in A4C and PLAX quality assessment.

- Demonstrate the feasibility and applicability of four quality attributes framework which can be adapted for benchmarking, reference standard of evaluation and objective quality scoring of 2D echocardiographic cine loop.
- Preparation and annotation of large (PACS-1) 33,784 independent echocardiography patient dataset showing four attributes of objective quality standard namely: ***anatomical visibility, chamber clarity, depth-gain, and fore-shortening*** for A4C, PLAX's echocardiograms optimum quantifications.
- Repository release of experts annotated patient dataset containing A4C, and PLAX to allow future studies and external validation of the new approach or methods.
- Detailed implementation of multi-stream deep learning architecture pipeline to process and allow access to specific image attributes in A4C and PLAX view of echo cine loop.
- Propose quality scores representation using symbolic scores to depict high quality, average quality, and high-quality score values.

(iv) **Acquisition and Optimization Guidance**

- Proposed a fully optimised deep learning (C-LSTM) architecture pipeline that simultaneously predicts four independent quality scores and view classification from echo sequences.
- Development of robust real-time quality assessment and optimization pipeline for 4 specific quality attributes to aid optimum image acquisition and reliable clinical quantifications.
- Provide evaluation of real-time application pipeline suitable for operator feedback for data acquisition, and quality optimization in A2C, A4C, and PLAX cardiac standard views.

During this PhD study, part of the outcomes of the research have been published in one journal, while other aspects were accepted for publication at several relevant national and international conferences with an international award. In addition, three journal paper is under peer-review process with Journal of Intelligent Medicine, Journal of Medical Image Analysis, Journal of Computers in Biology and Medicine respectively. Publications are listed in Appendix A.

1.5 Thesis Structure

This thesis comprises of eight chapters and two appendices. In the first section of every chapter, is an overview on the subject in focus, followed by the main body:

Chapter 2 describes the clinical background of the echocardiogram specimen, physical properties of ultrasound imaging and significance of echocardiographic acquisition, different types of ultrasound transducers, different imaging modalities and apical views standard. Furthermore, the concept and constituents of clinical image quality and method of its assessment will be discussed.

Chapter 3 presents the technical overview of common neural networks classification models and the general overview of the state-of-the-art deep convolutional neural network commonly applied for image and video regression problems. Moreover, the technical literature review of NAS, and **semi-automatic** search herein known as 'semi-autoNAS' solutions will be explained.

Chapter 4 investigates the echocardiogram frame classification using different deep learning methodologies and a semi-autoNAS derived lightweight neural network architecture have been proposed suitable for view/frame classification of 14 standard echocardiographic image specimen using a large private patient dataset of echo images. This chapter also details the performance of three chosen state-of-the-art and semi-autoNAS derived ultra-lightweight models that compared with the NAS derived Differentiable Architecture Search (2Cell - DARTS) model used for view classification.

Chapter 5 presents details on echocardiographic image quality definition using a public dataset CAMUS and large private dataset PACS. It provides justification why single annotation of image quality on CAMUS is insufficient for clinical and unified workflow and explores the domain-specific quality definition of cardiac specimen, structures, and pathological attributes to define objective criteria, a multi-layered annotations dataset providing 4 different attributes per image. The chapter also features related work on automated assessment of 2D image quality using a proposed multi stream C-LSTM model. Furthermore, it discusses the advantages of Spatio-temporal attributes associated with cardiac cine loop and suitable model to extract Spatio-temporal features for high-end model performance. Finally compares the results of the proposed model with other existing models in the 2D echocardiography.

Chapter 6 explores the use case for global framework and quality attributes in echocardiographic specimen using semi-supervised model for EchoLAB dataset specimen characterization and pseudo-labelling of 2D ultrasound cardiac images. A robust neural network architecture has been proposed using ensemble learning architecture which combines an ultra-lightweight architecture (semi-autoNAS derived model) and three state-of-the-art in the classification of apical four images and provide automatic annotation of a large unlabelled ultrasound image samples. The literature review on the semi-supervised application to echocardiographic specimens and performance of the implemented pipeline were discussed.

Chapter 7 provides a complete pipeline model for operator's feedback, suitable for real-time image optimization experiments using PACS-2 Dataset which consist of A2C, A4C and PLAX standard views as recommended by ASE. Finally, the chapter discusses the architecture and real-time optimization techniques available on the pipeline model.

Finally, **Chapter 8** summarises the work done in this thesis, provides general conclusion, and presents future directions in the field of medical imaging. Finally, a list of appendices, publications, and the list of references are presented.

1.6 The Research Consortium

This study forms part of a larger 3-year British Heart Foundation (BHF)-funded collaborative project, focused on developing automated echocardiographic image analysis pipelines, involving academic and clinical partners at University of Lincoln, Imperial College London, University of West London, and St Mary's Hospital, London. This research is devoted to the development of deep learning models for image quality assessment which involves three constituent steps within the unified echocardiographic examination workflow (i.e. development of objective quality standard, development of real-time assessment methods for multi criteria image quality, and development of operator optimization guidance system), while the work by other researchers involved developing, view classification, LV segmentation, strain measurements, phase detection, doppler image analysis, and electrocardiogram signal analysis.

Chapter 2

Clinical Background

Cardiac functional analysis is an established approach being used to diagnose most cardiovascular diseases and presents clinical clues to interventions. Over the past decade, the application of medical imaging techniques has facilitated state-of-the-art image-based analysis of cardiac functions. For any medical examination of heart's health, cardiologists rely on imaging and analytical technology to measure, quantify myocardium, and signal activities within the purkinje network (Gaudet *et al.,* 2016). In clinical cardiac diagnosis, existing technology ranges from Electrocardiogram (ECG), measuring the electrical activity of heart's rhythm; Magnetic Resonance Imaging (MRI) to produce details of heart's anatomy among others; Cardiac Computed Tomography (Cardiac CT) for producing detailed structure of internal organs, tissues, and skeletal structure; and Echocardiogram (Ultrasound). Some levels of biological risk are involved in deployment of all except the echocardiogram, proving superior in terms of non-ionizing in vivo exposure, lower cost of ownership, high temporal resolution, and excellent portability. However, the performance of ultrasound imaging technique suffers from the limitation of the spatial resolution constrained by the acoustic diffraction and limitation imposed by the operating ultrasound frequency (Chen *et al.,* 2021) in contrast to the rest.

In this chapter, a detailed account of the physics of ultrasound imaging, various imaging modalities and significance are discussed. Then, the echocardiogram specimen, fundamental challenges of objective characterization and method of quality assessment are explained. Finally, a brief overview of the different datasets used in this thesis are provided.

2.1 Overview of Cardiac Physiology

Human heart is a life-supporting four-chambered organ responsible for blood and hormonal circulation. The heart cavity consists of four chambers: the right and left atrium, and right and left ventricles with separating walls to distinguish the right and left blood chambers, this is called intraventricular septum (Hinton and Yutzey, 2011). The heart also has one-way valves that separate the chambers and the major arteries, which prevent blood backflow into the chamber. The aortic valve, separating the aorta from the left ventricle, while the pulmonic valve, separating the pulmonary artery from the right ventricle, are known as semilunar valves. The two atrioventricular (AV) valves are the tricuspid and mitral valves. The tricuspid valve marks the separation between the right atrium and right ventricle while the mitral valve separates the left atrium from the left ventricle. Figure 2.1 depicts an Apical Four-Chamber (A4C) view of the heart is one view that presents a combined visualisation of all four chambers of the heart.

The histological decomposition of heart as cardiomyocyte cell (Talman and Kivelä, 2018) with three functional layers of connective tissue, the muscles of the heart, and the inner lining of the heart which protects the valves and chambers that are called the epicardium, myocardium, and endocardium respectively. Further protection is a double-layer, fluid-filled sac known as the (parietal and visceral) pericardium, surrounding the heart. The epicardium is joined to the myocardium on one side and to the pericardium in several layers, with pericardium fluids. While coronary arteries, veins, vessels, and nerves run below the epicardium, the endocardium is trabeculated and composed of connective tissue layers – papillary muscles, external purkinje network (Labs *et al.*, 2020). The heart pumps out blood in a systolic process with the LV contracting to generate the right pressure and squeezes blood out of the heart (Maznyczka *et al.*, 2021). This process is followed by a diastole cycle, while the heart relaxes its right ventricle chamber is immediately filled with a new volume of blood. These sequences are known as a cardiac cycle, which forms a single heartbeat (Figure 2.2). An obvious fact in the length of cardiac cycles is influenced by physical and psychological different factors such as exercise, emotions, fever, diseases, and some medication heart contract at different rates (Hanft, Korte and McDonald, 2007). The normal heart typically beats at around 75 beats per minute, so the length of each cardiac cycle is usually less than one second (Hernández-Vicente *et al.*, 2021). But during this short time, a lot of pressure changes take place in the heart (Klabunde, 2011).

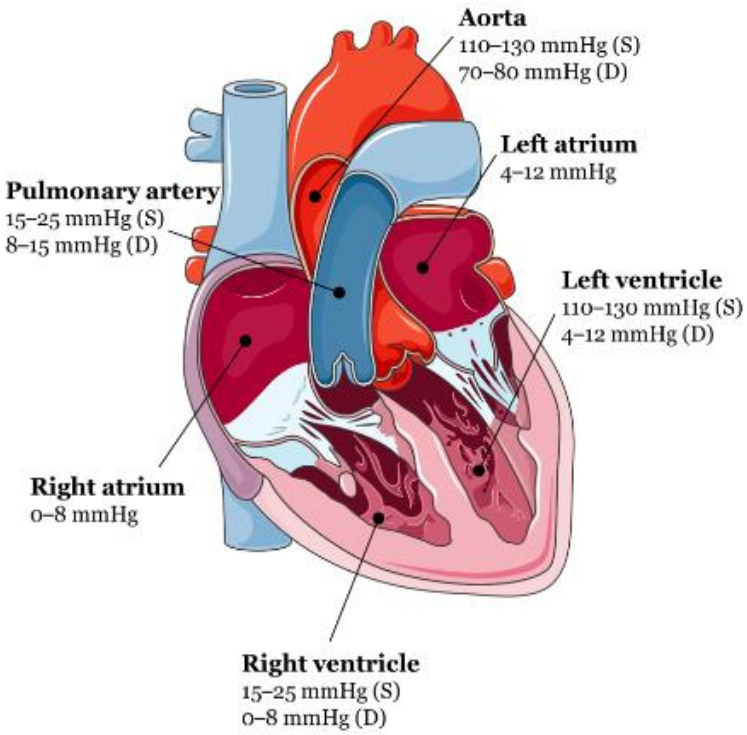


Figure 2.1: Overview Schematic representation of apical four chamber of the heart with corresponding pressure level during diastole (D) and systole (S) cycles. Adapted from (Maksuti, E, 2016) and (Mariana Ruiz Villarreal, 2006).

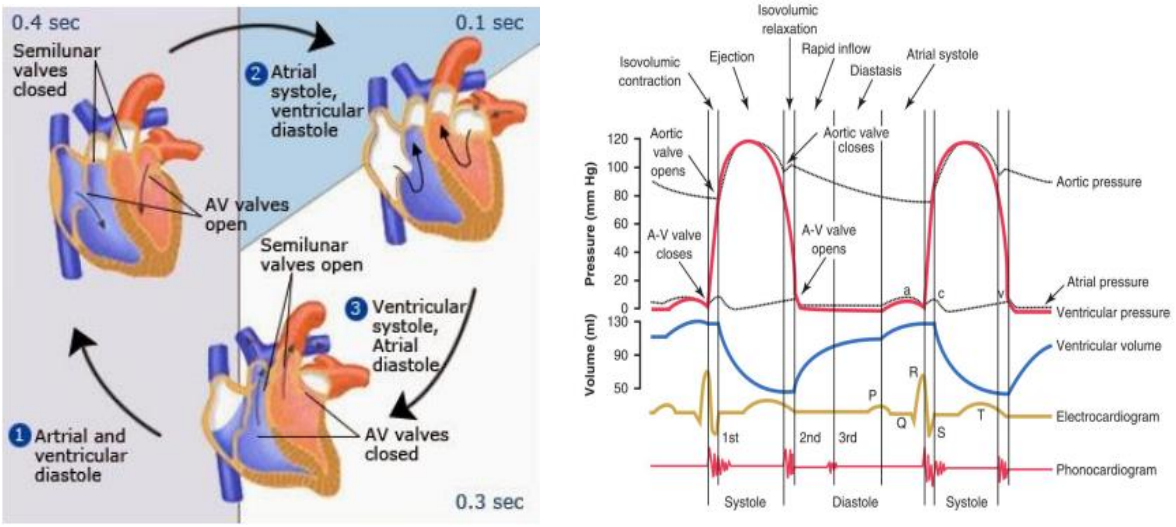


Figure 2.2: Illustration showing the phases of cardiac cycles for atrial and ventricular events with corresponding electrocardiogram (ECG/EKG) Wiggers diagram, showing the components of the cardiac cycle. Total length of cardiac cycle usually 0.8 seconds can be elevated during physical or psychological activity. Adapted from (Hernández-Vicente et al., 2021) and (Klabunde, 2011).

2.2 Echocardiogram Acquisition and Assessment

An echocardiogram is a non-invasive examination that uses sound waves to look at the size, shape, motion, performance of heart and its valves, pumping capacity, and the location and extent of any tissue damage. This procedure is also known as echocardiography (cardiac echo) or diagnostic cardiac ultrasound and provides exclusive ability for real-time images of the beating heart (Horton, 2010). Echocardiography can allow extracting other measures such as measuring the EF, cardiac output, and diastolic function (i.e., how well the heart relaxes) (Cleve and McCulloch, 2018). In a patient with a suspected cardiac disorder, echocardiography is essential in assessing the motion of wall cavities and to detect some cardiomyopathies such as hypertrophic cardiomyopathy, dilated cardiomyopathy, and left ventricular systolic dysfunctions (LVSD). Echocardiogram is helpful in the early diagnosis of myocardial infarction evaluating the regional wall cavities and motion abnormalities, treatment, and follow-up in patients with heart failure or in clinical emergencies. Echocardiography exams are conducted by trained clinicians, cardiologists or sonographers trained in echocardiography (Modin, Andersen and Biering-Sørensen, 2018). Classic echocardiography examination procedures, usually between 15 to 45 minutes, starts by applying acoustic gel on patients' left hand chest region and placing a transducer probe in the relevant intercostal window as illustrated in Figure 2.3. Usually for cardiac assessment, a phased array transducer can be placed in different 'window' locations of the chest, and at different angles, to capture the relevant view required for diagnosis. The most common echo windows are the parasternal, apical, subcostal, and suprasternal as illustrated in Figure 2.4. From each window, an operator can manipulate the transducer either by rotating and/or tilting the transducer without moving it to a new window (Figure 2.5). These lateral or medial translations can provide significant improvement on the image quality obtained or clinical measurement for further functional analysis.

The equipment then displays the corresponding image assessed by the operator as apical-two chambers (A2C), apical-four chambers (A4C), parasternal long axis image (PLAX), etc. Hence the quality of imaged obtained depends on operators' skill, experience, and patients' pathological or physiological profiles.

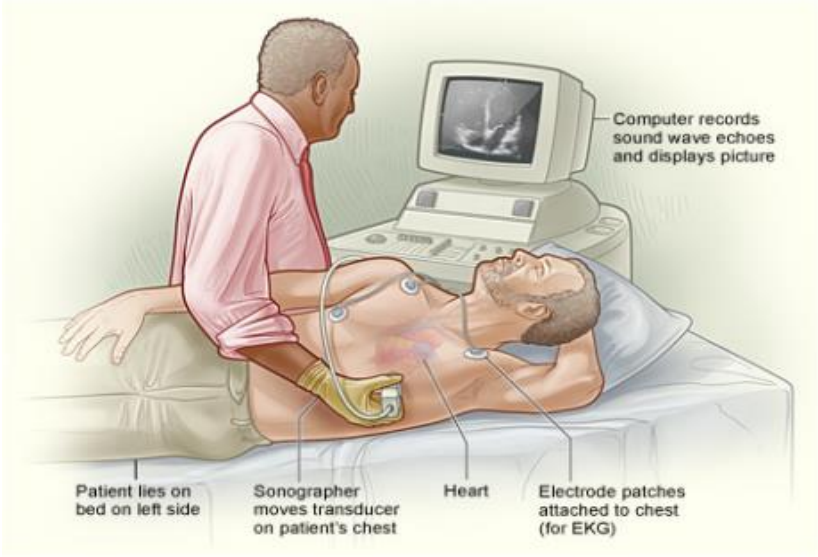


Figure 2.3: Illustration of Transthoracic Echocardiography (TTE) clinical examination process in the lab.

The adaptation of transthoracic echocardiography (TTE) is the focus of study, although there exist other types of echocardiograms significantly useful depending on the potential heart problem doctors wish to investigate. Nevertheless, TTE has been famously and frequently adapted for clinical echocardiography. The reason could be for in vivo easy access that patients need not worry about the procedure unlike TEE where a probe is inserted for clarity. I enumerate other types of echocardiograms and their significance in clinical assessment of myocardium.

(i) **Transthoracic Echocardiogram (TTE**): During this examination, a trained operator spreads acoustic gel onto the chest and presses a device called a transducer (probe) against the skin. A high-frequency sound waves is propagated into the areas of interest which is then reflected or bounce off the walls and valves of the heart. The reflected sound waves known as echoes, is detected by the transducers and displays a moving image of the heart's chambers, walls, and valves on a monitor while the scan is accomplished. This procedure is done in vivo and commonly used for the assessment of cardiac functions in children and adults. The general application is wide in clinical laboratory and during clinical emergencies. Therefore, it is logically sensible to evaluate echocardiogram produced under TTE exams.

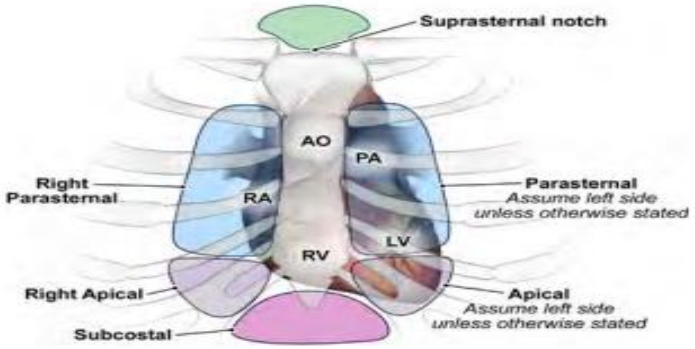


Figure 2.4 The standard recommended transducer positions in transthoracic echocardiography windows to obtain images. Imaging windows showing Parasternal, Apical, Subcostal (SC), and Suprasternal notch (SSN). (Mitchell, C., et al, 2018).

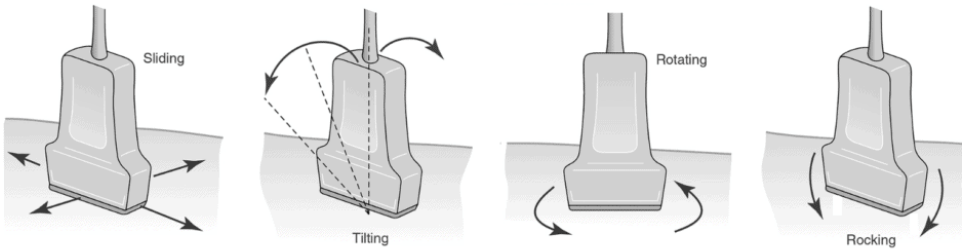


Figure 2.5: Transducer manoeuvring / manipulation for optimum quality image acquisition: sliding in 4 directions, tilting sideways, rotating on probe axis, and rocking for optimum contact pressure are the possible translations for probes fine-tuning that affects imaging outcomes and quality.

(ii) **Stress Echocardiogram:** Stress echocardiogram examination known as stress echo is to find out if the patient has decreased blood flow to the heart muscle (i.e., coronary artery disease). The stress echo uses ultrasound imaging of the heart to evaluate the wall motion in response to physical stress. This examination increases the heart rate and blood pressure. During this examination, two sets of images will be taken including one at rest, and another after working out on a treadmill or stationary bike. If the patient's health condition limits physical activity, a medication will be injected to simulate the effect of exercise (Prisant, Watkins and Carr, 1984).

(iii) **Transoesophageal Echocardiogram (TEE):** For TEE examination, the transducer instead of being moved over the outside of the chest wall is passed down the oesophagus. TEE would allow producing clearer pictures of the heart because the transducer is located closer to the heart and the lungs and bones of the chest wall do

not block the sound waves generated by the transducer (O'Rourke and Mendenhall, 2019; Blinn, Margulis and Joshi, 2019).

(iv) **Three-Dimensional Echocardiogram (3D Echo):** A Three-Dimensional (3D) echocardiogram uses either transoesophageal or transthoracic echocardiography to generate a 3D image of the heart. This examination includes multiple images from various angles. It's used for recognising problems with heart valves, before heart valve surgery for replacement heart valves, or diagnosing heart problems in children (Lang *et al.*, 2015). Currently, 3D echocardiography suffers from a considerable reduction in frame rate and image quality, and this has hindered its adoption into routine practice. When such issues are resolved, automatic analysis of the 3D images could also be explored. Meanwhile, 2D echocardiography remains unrivalled and clinically relevant, particularly when high frame rates are needed.

2.3 Imaging Modalities and Significance

Echocardiographic imaging has evolved from single line pulse-echo where amplitude and depth are processed to yield A-Mode echocardiography. In the application of cardiac ultrasound imaging, diverse types of mode can be controlled by the operator, each of them conveying a specific type of information to the clinician (Bom, N., et al., 2004). There are three basic modes used to image the heart which is briefly explained in the following section:

- **A-Mode:** In the A-mode presentation of ultrasound images, echoes returning from the body are displayed on oscilloscope which presents a graph of voltage representing echo amplitude on the ordinate (y-axis) or as a function of time on abscissa (x-axis), hence the term "A-mode". With the assumption of a constant speed of sound, time on the x-axis can be presented as distance from the ultrasound transducers. Hence, the A-mode reveals the location of echo-producing structures only in the direction of the ultrasound beam. It has been used in the past to localise echo-producing interfaces such as midline structures in the brain (echoencephalography) and structures to be imaged in B-mode. The concept of A-mode is, however, useful in explaining how pixels are obtained from scan lines in the B-mode imaging system.

- **M-Mode (Motion Mode)**: M-mode has a great spatial resolution (Sarvazyan, Urban and Greenleaf, 2013), which is useful for measuring ventricular dimensions in systole and diastole. In this mode, motion of the internal organ can be reflected for systolic assessment. Its application is also used in cardiac timing and the measurement of dimensions. M-Mode displays a one-dimensional image that measures the distance of the object from the single transducer at a given moment. The ultrasound shows this information as a 2D image which is depth and time. M-mode images have a very high sampling rate, which results in a high time resolution. Therefore, very rapid motions can be recorded, displayed, and measured. However, in these types of images, the ultrasound line is fixed to the tip of the ultrasound sector. It may therefore be difficult to align the M-mode perpendicular to the structures which are displayed (i.e., the septum), thus leading to false measurements (Loizou, Pattichis and D'hooge, 2018).

- **B-Mode (Brightness Mode):** This mode is more commonly known as 2D that allows a plane of tissue (both depth and width) to be imaged and displays the ultrasound reflection as an 8-bit grayscale image that composed of bright dots representing the ultrasound echoes (Sarvazyan, Urban and Greenleaf, 2013). The brightness of each dot is determined by the amplitude of the returned echo signal. This allows for visualisation and quantification of anatomical structures, as well as for the visualisation of diagnostic and therapeutic procedures. The anatomic relationship between various structures is easier to recognise than M-mode echocardiographic images. The formation of a B-mode image depends on the pulse-echo principle; assuming the speed of sound remains constant, the position of a target of interest may be inferred by the time taken from emission to its return to the transducer. The limitless number of imaging planes through the heart is possible, however, the standard view will be used to assess the intra and extra cardiac structure (Prada et al., 2015). This thesis focuses on the principle of B-Mode imaging in echocardiogram.

- **TDI (Tissue Doppler Imaging):** Tissue Doppler Imaging (TDI) is a modality that allows measuring myocardial velocities to evaluate global and regional myocardial systolic and diastolic function. It can also be employed to quantify right ventricular and left atrial function (Atzeni *et al.*, 2017). TDI is useful as a diagnostic as well as prognostic tool in different cardiac conditions such as coronary artery disease, heart failure (both systolic and diastolic), valvular heart disease, cardiomyopathies, and constrictive pericarditis. Also, TDI measurements are helpful to recognise patients who will benefit from cardiac resynchronisation therapy. Although TDI is reproducible and quite easy to acquire, it is

underutilised in routine clinical practice. (Kadappu and L. Thomas, 2015). Figure 2.6 displays some examples of different modalities used in transthoracic echocardiography.

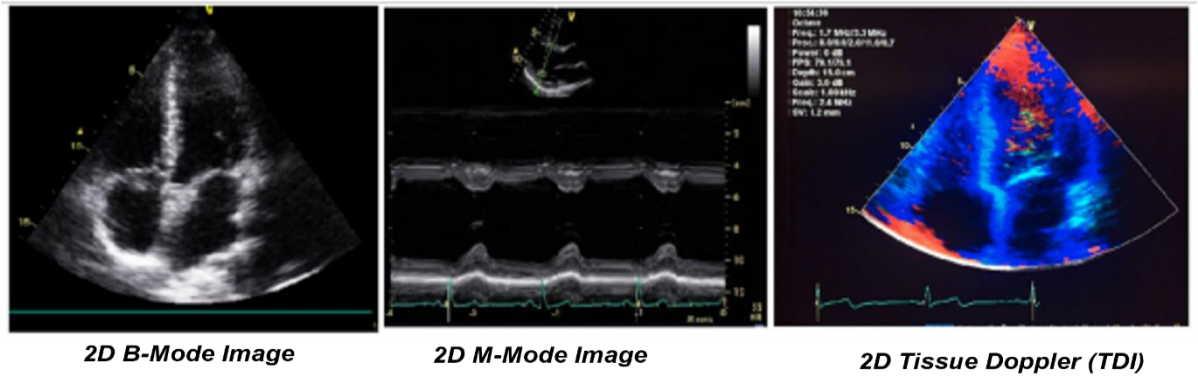


Figure 2.6: Examples of different modalities used in transthoracic echocardiography, showing B-Mode, M-Mode, and TDI echocardiograms.

2.3 Characterization of Cardiac Specimen

Anatomy of cardiovascular specimens presents enormous complexity to objective functions in terms of dynamic features identification and clinical quantifications (Sarvazyan, Urban and Greenleaf, 2013). This is due to clinical protocol where cardiologists not only rely on still images but a fast-moving echocardiogram frame in real-time. However, a plausible solution (Potter and Marwick, 2018); using deep learning models have been widely demonstrated for cardiac view classification (G.N. Balaji, Subashini and Chidambaram, 2015), quality assessment (Labs et al. - 2020) and now it's been applied to achieve pseudo-labelling and characterization of cardiac samples in a semi-supervised model (Cozman, Cohen and Cirelo, 2003); (Shiming Xiang, Feiping Nie, and Changshui Zhang, 2010); (Chen *et al.*, 2020). Objective characterization of two-dimensional echocardiograms is one of the potential approaches to define the critical features of domain-specific elements of cardiac specimens in terms of anatomical and pathological features.

During transthoracic exam, several images are by default, required to build a complete picture of patients' pathology and summary of diagnosis; however, several specimens for multiple patients indicate the enormous task of manually assessing several hundreds of specimens prior to diagnosis. An automated system that can objectively allow cardiologist filter-search and

group specific specimen down to pathological and anatomical relevance (as described in table I & II) is certainly beyond the scope of apical view classification. This system could allow rapid assessment of functional, anatomic, and pathological features present in each specimen.

Although this study focused on apical-four, the impact encapsulates apical-four chamber images (A4C) and parasternal long axis (PLAX) images which are the recommended views for critical quantification and clinical measurement the principles can be applied to other apical standard views. Consequently, objective characterization of these two views are significant prerequisites to globalisation standard in objective assessment, implementation of real-time optimization, and anatomical function sorting which can speed up clinical process.

2.4 Definition of Objective Quality Attributes

Due to its ubiquitousness and nonionizing advantages, echocardiograms have found its significance in antenatal, obstetric, and general diagnosis of myocardial infarction. Although, echo image does provide rich information about myocardium, it does not present crisp edges of a well-defined resolution when compared to photographic images. Since quality assessment is currently done by observer visual assessment, there exists huge variability as to what constitutes the quality element of 2D echocardiographic image at most clinical practice.

While there exists no coherent quality standard of objective image quality by which cardiac specimens can be judged, it has been affirmed that accurate diagnosis of the left ventricle functions is hugely dependent on the quality of echo images. Consequently, there is a need for a coherent global standard of domain-specific criteria for 2D image quality. Mitigation by observer visual assessment which is highly subjective thus became the status-quo and requires varying incoherent definitions under clinical pathologies. The inherent variations in echocardiographic image quality standards indicates the complexity faced among clinicians and provides apparent evidence for incoherent assessment under clinical trials, especially with less experienced cardiologists. Several researcher including (Luong *et al.*, 2021) admitted there exist no reference standard for the evaluation of 2D echocardiographic image quality hence, the scale of criteria for 2D image quality, proposed in many publications does not represent expert visual assessment, consensus on 2D echocardiographic image quality and failed to meet translatory clinical relevance. These are only paramount to the quantification of the left

ventricle (LV) functions, which serves the ultimate purpose of cardiac diagnosis or myocardial assessment.

This research defined and presented the analysis of domain-specific quality attributes of 2D echocardiogram specimen under: (i) On-Axis, (ii) Chamber Clarity, (iii) DepthGain and (iv) foreshortening attributes that can gain translatory advantage in clinical practises (Labs, Zolgharni and Loo, 2021). Objective quality attributes employ the use of a multi-stream deep convolutional neural network to extract image's quality features as defined by a set of clinically relevant criteria and computes objective quality scores which can be used to arbitrate the quality of cardiac specimens in A4C.

2.5 Assessment Methods of 2D Image Quality

The assessment of cardiac image quality and the clinical measurements of cardiac functions are domain knowledge which is performed using high quality images. In clinical practice, clinicians have established that poor quality images have dire consequences on cardiac quantification and clinical measurements. During the cardiac cycle, when the LV contracts, the endocardium muscle shortens in the longitudinal and circumferential dimensions which can produce a negative strain that shows a constrained displacement defined by heart's cavity and wall structure. In this phase, the muscle will thicken or lengthen in the radial direction to produce a positive strain (Thomas H Marwick, 2006) while the apical chambers show alternating and varying dimensions and volumes. The preferred apical orientation, among other vital quality attributes is observed in real-time but can present foreshortening artefacts, hyperechoic or hypoechoic among other undesirable acquisition anomalies.

Currently, the method of image quality assessment is a subjective process, where an echocardiography specialist visually inspects the images and decides on what anatomical features present in the image to be pathologically relevant. This process is laced with a spread spectrum of opinion and decision variability (Nagata *et al.*, 2018) even when an image is reassessed by the same operator. These variabilities and uncertainties (Liao *et al.*, 2019) are found to impair quantification accuracy of cardiac functions, diagnosis, and the overall quality of patient care. Unfortunately, a subjective assessment method is incapable of detecting temporal changes in a 2D fast moving pixel data. This is because the amplitude of reflected

ultrasound detected by the imaging system can vary considerably over several logarithmic units of signal strength, and its well beyond the capacity of human visual perception. (Mitchell *et al.*, 2019). Therefore, these represent a layer of complexity in the currently adopted assessment method in 2D echocardiogram.

Furthermore, earlier works on quality assessment methods all indicate using a weighted average of image overall quality score. Unfortunately, a single score value of objective quality cannot provide specific feedback to the aspect of image quality that needs to be optimised. Therefore, an assessment method that would be domain-specific showing score values for different attributes of image quality criteria would provide a succinct and precise feedback to the aspect of image quality needing optimization. Using this feedback can be significant to achieving optimum image quality, reproducibility, accurate quantifications and to provide possibility for automated diagnosis in clinical practice.

Rather than obtaining an overall quality score for the image from a weighted average of these quality measures, we instead used four defined attributes separately and proposing novel approach using four (4) quality attributes and method of assessment in real-time during acquisition phase

The application of objective quality assessment method is one of the objectives that this thesis investigated, as it is paramount to image acquisition, quality optimization and overall diagnosis of cardiac specimens.

2.6 Overview of Dataset Used

Representative multi-centre patient datasets are essential for ensuring that any developed models would scale up considerably well to other sites and environments. Therefore, this study employed several private and public datasets, originating from different clinical sites and acquired by different imaging equipment from various vendors, and representative of real-world patient population. Table 2.2 provides a summary of the datasets used. Chapter 4, 5 and 6 will provide the details of each dataset, when dealing with problem-specific data.

Table 2.1: Summary of echocardiographic patient datasets used for different applications/tasks including classification, segmentation, and Speckle Tracking in this project.

Dataset Name	CAMUS	PACS	PACS-1/2	EchoLab
Model Usage	Classification	Classification	Regression	Semi-Supervised Regression
Type	Public	Private	Private	Private
Source	University Hospital of St Etienne (France)	NHS Trust, Imperial College Healthcare	NHS Trust, Imperial College Healthcare	NHS Trust, Imperial College Healthcare
Acquisition	GE Vivid E95	GE and Philips	GE and Philips	Philips iE33
No of Patients	450	374	374	61
Specimen	A2C, A4C	14 Apical Standards	A2C, A4C, PLAX	A4C
Ground-Truth	1 annotation	1 annotation	2 Annotations, 1 Expert and 1 Accredited.	2 annotations by 2 Experts
No of Frames	**1,800**	**41,321**	**40,000**	**27,230**

2.7 Conclusion

This chapter provided the clinical background to cardiology, the significance of echocardiograms, image acquisition, different types of echocardiograms and ultrasound modalities with a focus on the Transthoracic Echocardiography (TTE) and B-Mode modality. Also detailed dataset sources with respective clinical properties. Finally discussed the concept of objective quality element in 2d echocardiogram specimen, and the method of assessment in real-time. Further experiments on these are detailed in Chapters 4, 5, 6 and 7.

Chapter 3

Technical Background

The application of deep convolutional neural networks is the bedrock for recent advances in artificial intelligent (AI) technology (LeCun, B., and Hinton, G. 2015; Shen, G. et al., 2017; Suzuki, 2017). Self-driving cars, virtual assistants like Alexa, Sri, and the healthcare industry, have recorded plausible feats which include advanced medical diagnosis and clinical echocardiography (Shrestha and Sengupta, 2018).

Deep convolutional neural network (DCNN) is a subset of Machine Learning (ML) that exhibits the ability to learn high-level features from a dataset. Although it requires significant computing power (GPU) and large labelled dataset, it can achieve state-of-the-art accuracy, sometimes exceeding human-level performance making it best candidate to extract hierarchical features (Dean, J., 2016) in complex structure of images like echocardiograms.

In this chapter, first, an overview of the physics of ultrasound imaging and artificial neural networks will be presented, including a brief introduction to DCNN and different approaches to neural network design. Then, the general DCNN architectures used to solve classification, regression and pseudo-labelling problems will be described. Lastly, an overview of neural network architecture search (NAS) and its counterparts will be illustrated.

3.1 Physics of Ultrasound Imaging

Acoustic sound wave above audible frequency range of 20 - 20,000Hz exhibits interesting characteristics of propagatable mechanical compression and rarefactions therefore become the basis for ultrasound imaging technology. Ultrasound differs from audible sound only in its frequency, and it is 500 to 1000 times higher than the sound we normally hear. Sound

frequencies used for diagnostic applications typically range from 2 to 15 MHz, although frequencies as high as 50 to 60 MHz are under investigation for certain specialised imaging applications. Diagnostic ultrasound is used to assess and evaluate patients' internal organs during a clinical procedure, sound waves are propagated in-vivo (Rayleigh 1975), while the returned pulse-echo is used to construct image structure. These interactions provide the information needed to generate high-resolution, grey-scale images of the body, as well as display information related to blood flow (Merritt, no date). Consequently, the physics of ultrasound imaging consist of the following principle:

(i) **Propagation of Acoustic Waves:** Sound is the result of mechanical energy travelling through matter as a wave producing alternating compression and rarefaction. Pressure waves are generated and propagated by limited physical displacement of the material through which the sound is being transmitted. Changes in pressure with time define the basic units of measurement for sound. The distance between corresponding points on the time-pressure curve is defined as the wavelength (λ), and the time (T) to complete a single cycle is called the period. The number of complete cycles in a unit of time is the frequency (*f*) of the sound. Frequency and period are inversely related. If the period (T) is expressed in seconds, $f = 1/T$, or $f = T \times \sec^{-1}$. Furthermore, frequency and wavelength have some relationship with propagation velocity of sound (*c*) which, largely determined by the resistance of the medium to compression, is a product of frequency (*f*) and wavelength (λ). Consequently, resistance is influenced by the density of the medium and its stiffness or elasticity. In Figure 3.1, propagation speed (velocity) can be increased when stiffness increases or by decreasing the density (Jensen, 1991).

Therefore, when the transmitted acoustic waves (ultrasound beams) transverse specific object with dissimilar cavities of acoustic impedance (Table 3.1), can be modelled using Snell's law, and are subjected to one or all the following phenomenon: (a) Acoustic Reflection (Rayleigh's 1975), (b) Refraction (c) Scattering, (d) Attenuation, and (e) Transmission as illustrated in Figure 3.2.

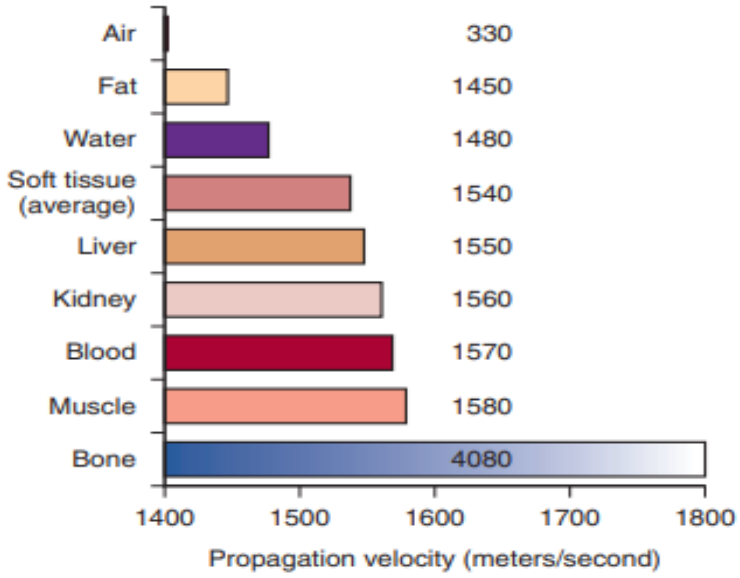


Figure 3.1: Propagation speed in tissues is determined by the physical properties of tissue which varies considerably. Clinical ultrasound probes assumed average propagation velocity of soft tissue of 1540 m/sec.

Reflection: The way ultrasound is reflected when it strikes an acoustic interface is determined by the size and surface features of the interface. If large and relatively smooth, the acoustic reflection will be like the reflection of light by a mirror. Therefore, the amount of energy reflected by an acoustic interface can be expressed as a fraction of the incident energy (*P*); this is termed the reflection coefficient (*R*). However, in cardiology, different tissue naturally offers specific resistance known as acoustic impedance and this includes tissues, bones, and fluids (Table 3.1). Acoustic impedance denoted by (Z) is determined by the product of the density (ρ) of the medium propagating the sound and the propagation velocity (c) of sound in that medium (Z = ρc). If a specular reflector is perpendicular to the incident sound beam, the amount of energy reflected can be illustrated by Figure 3.2. As with propagation velocity, acoustic impedance is determined by the properties of the tissues involved and is independent of frequency. Thus, the reflection coefficient *R* is the ratio of intensity of reflected echo beam versus intensity of the incident beam observed at the boundary ρ_1 , c_1 *and* ρ_2 , c_2. This represents the interface acoustic impedance *Z* for respective layers Z_1 and Z_2 given by the equations (3.1 and 3.2):

$$Reflection = \frac{(\rho_2\, c_2) - (\rho_1\, c_1)}{(\rho_2\, c_2) + (\rho_1\, c_1)} \qquad (3.1)$$

$$R = (\frac{Z_2 - Z_1}{Z_2 + Z_1})^2 \qquad (3.2)$$

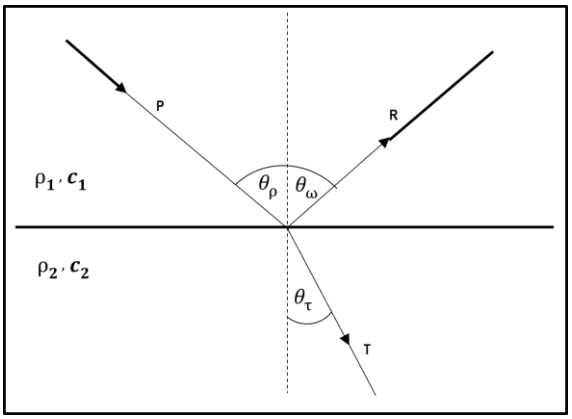


Figure 3.2: Illustrating the relationship between acoustic impedance and reflection coefficient

Refraction: The change in the direction of the incident propagation wave is governed by Snell's law (equation 3.3) and is known as refraction. Refraction occurs when acoustic waves transit through a dissimilar medium having dissimilar acoustic impedance properties. Refraction is governed by basic Snell's law in equation (3.3):

$$\frac{sin\ \theta_\rho}{sin\ \theta_t} = \frac{c_1}{c_2} \qquad (3.3)$$

As illustrated in Figure 3.2, θ_ρ , θ_τ represent the angle of incidence and refraction respectively. This angle is usually less than the critical angle where refraction causes no ultrasound to enter a medium is 90 degrees.

Scattering: Echocardiograms are formed by scattering centres which occur when the reflected acoustic wave becomes equal or smaller than the wavelength of the incident beam interacts with a structure of a radically different impedance. Such echo originating from relatively small, weakly reflective, irregularly shaped objects, are less angle dependent exhibiting the properties of a 'diffuse reflector' with different acoustic impedance compared to the surrounding tissue. Hence, diffuse reflectors cause ultrasound waves to scatter in all directions thus resulting in multiple echoes propagating from the numerous tiny structures. Scattering produces echoes with smaller amplitudes while interacting with each other in

turn causes both constructive and destructive interference of the waves known as speckle which is seen as an irregularity in the grayscale of the image.

Attenuation: The intensity of acoustic amplitude diminishes as beams travel through a medium due to depth, scattering and absorption. Transmitted acoustic beams weaken with distance and are absorbed when energy is converted to another form of energy or when diffusely reflected in a direction other than original direction of propagation (Mézière *et al.*, 2014). This phenomenon accounts for the attenuation of the plain wave, given in equation (3.4) in Nepers or equation (3.5) in decibel.

$$\Delta A\left(\frac{Np}{m}\right) = A_{\omega} e^{-\alpha z} \qquad (3.4)$$

$$\Delta X(dB) = 10\, log \frac{X2}{X1} \qquad (3.5)$$

Where $A_{\omega}$is the unattenuated amplitude of the propagating beam at some point in the medium, A in Nepers per metre, is the attenuated amplitude assessed at distance z, and α is the attenuation coefficient (Figure 3.3) of the beam travelling in z direction, given in nepers/metre or in decibel/length (Napier's constant at 2.71828 or conversion to dB dividing by 0.1151). While delta *X* is expressed in decibel of some quantity, *X1* and *X2* are two different intensity values for amplitude of the transmitted and returned echo beams measured in decibel. Both equations are valid for acoustic beam attenuation properties of beam waves.

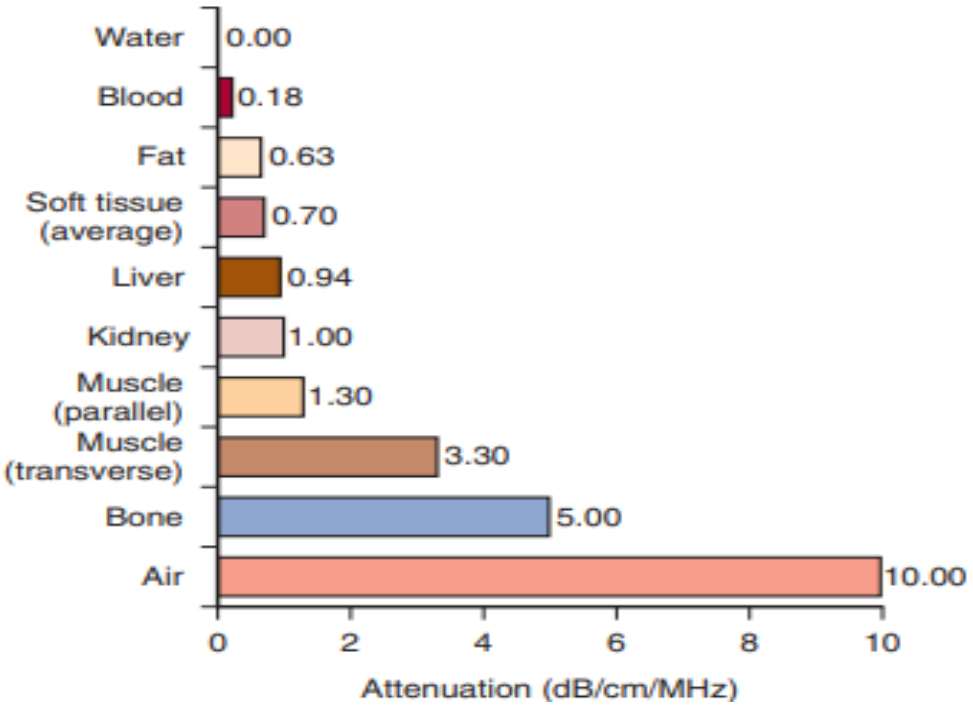


Figure 3.3: Illustrating the attenuation coefficient of different mediums. Attenuation occurs as an acoustic beam passes through tissue, it loses energy through the transfer of energy to tissue by heating, reflection, and scattering. Attenuation can be affected by chosen frequency and the medium's acoustic impedance. Attenuation also increases in proportion to insonating frequency, resulting in less penetration at higher frequencies.

(ii) **Image Reconstruction Process:** Image data are computed in the ultrasound imaging system (Figure 3.4) after listening to the reflected echo using pulse echo demodulation or reconstruction process, depicted in Figure 3.5. The demodulator computes the envelope and magnitude of the received signal, then initiates log compression and scan conversion processes. In the scan converter, image data is held in memory and continuously updated with new echo data arriving at the demodulator buffer. Phased array probes are essential in echocardiology and account for scan conversion from polar to cartesian coordinates with interpolating processes. At the same time, information is continuously read out to a video buffer to provide real-time visualisation of the scanned images. Consequently, echocardiogram images are displayed on the monitor where further onboard processing is carried out to meet specific clinical requirements, e.g., enhancement, quantifications, clinical measurements, or data storage. Figure 2.6 illustrates the hardware setup of trends in clinical ultrasound hardware.

Figure 3.4: Overview of hardware and trends in Ultrasound Imaging and diagnosis systems

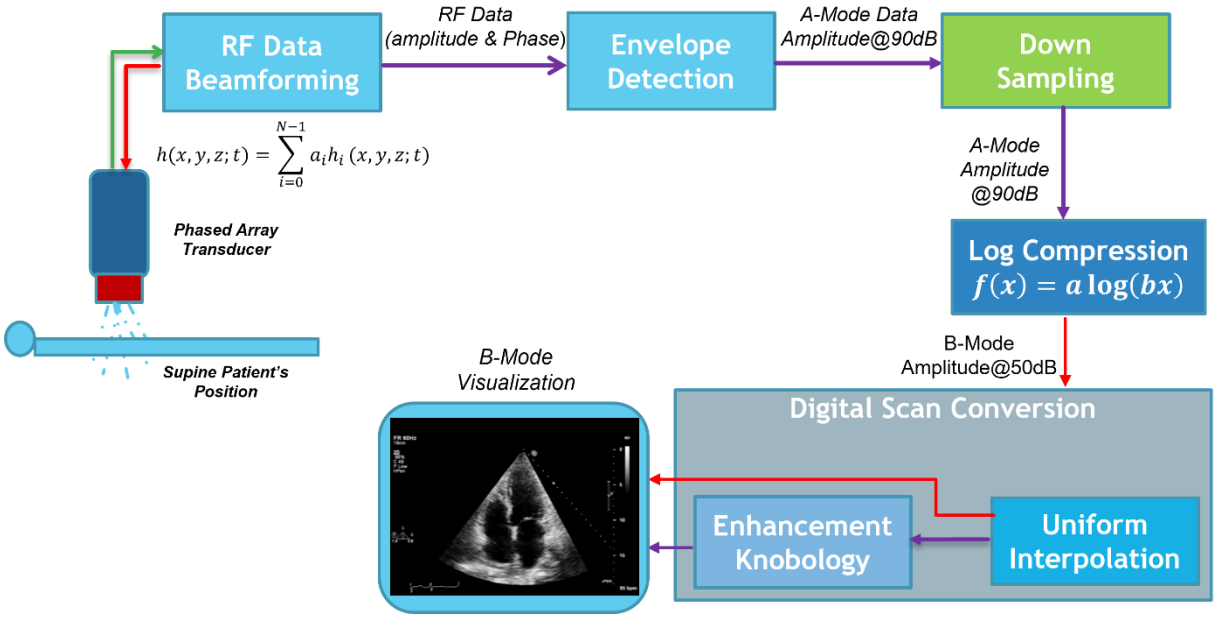


Figure 3.5: Echocardiogram Image generation and reconstruction process, showing the stages in pulse-echo amplification, envelope detection, down sampling stages and visualisation.

In the process of ultrasound imaging, human tissues offer some sort of resistance defined by acoustic impedance (z) which is the rate of resistance encountered by ultrasound beam. This is a physical property of tissues and bones which is dependent on densities measured in kg/m3, and speed of sound wave measured in m/s as summarised in Table 3.1. The increase in impedance varies with tissue densities and depth which is the reason for huge difference in patient pathological and anatomical profiles. The effect of acoustic impedance in medical ultrasound becomes noticeable at interfaces between different tissue types (Jensen, 1991). The ability of an ultrasound wave to transfer from one tissue type to another presents a subtle

acoustic impedance transition in the magnitude of the differential impedance of the two or more tissue elements. In medical imaging, pulsed waves are commonly used because they provide much improvement in axial resolution. Nevertheless, the continuous wave (CW) analysis also, is useful to determine the beam's diffraction pattern and effects of near and far-field even though it provides no information regarding the time domain (Sarvazyan, Urban and Greenleaf, 2013).

Table 3.1: Summary of material densities and acoustic impedance relating to echocardiograms image acquisition, generation, and reconstruction.

Material	Densities (kg/m^3)	Propagation Velocity (metres/second)	Impedance $kg/(m^2s)$
Air (@25deg)	1.16	330	0.0004×10^6
Pure Water (@22deg)	998	1482	1.48×10^6
Fat	928	1450	1.34×10^6
Liver	1050	1578	1.65×10^6
Blood	1060	1584	1.65×10^6
Muscle	1041	1580	1.71×10^6
Bone	1600	3360	7.8×10^6
Heart	1045	1570	1.64×10^6

3.2 Cardiac Ultrasound Probes

Ultrasound probes (transducers) are one of the important parts of ultrasound equipment, they contain essential electronic and piezoelectric components for the generation and propagation of acoustic wave and reception of returned pulse-echo to construct echocardiogram image illustrated in Figure 2.8. Since ultrasound transducers work as both transmitter and receiver (Figure 2.9), ultrasound probes are used as a listening device to detect pulse-echo reflection; they are technically evaluated at design level for performance efficiency (equation 3.6.).

$$P_c^2 = \frac{Electrical\ energy\ converted\ to\ mechanical\ energy}{Applied\ electrical\ energy} \qquad (3.6)$$

In practice, transducers are designed using man-made piezoelectric crystals such as barium titanate, lead metaniobate, and lead zirconate titanate (Hendee and Ritenour 2002) with acoustic baffles to minimises attenuation coefficient, which values varies with several vendors across

the globe. There are three main types of transducers: linear, curved and sector shown in Figure 3.6. Linear transducers are used primarily for small parts requiring high resolution and typically involve shallow depths. Nevertheless, to produce a broader view, sectorial transducers are used. These have small footprint and wide far-field view as illustrated in Figure (3.7). Also, available for general diagnosis are transducers with 64 piezo elements (arranged in convex or linear array), or 128 elements commonly used in sector phased array transducers are best suited for echocardiology clinical diagnosis. For abdominal in vivo investigations, curved transducers are typically used due to its large aperture. Sectorial phased array transducers provide the following advantages:

- Provides an effective way of accessing cardiac imaging through a narrow gap between 5th rib known as cardiac window or simply 5^{th} intercostal space
- Yields a divergent beam in the far field region (critical advantage for partial view) and narrow beam in near field.
- Allows a compromise view of resolution over depth. A disadvantage of poor resolution in the divergence far field view than the near field view.

Figure 3.6: Typical Transducer Probes. Different vendors bear their own trademarks. Example General Electric (GE) probes are common in many clinical practises.

The underpinning concept of echocardiographic imaging is based on the echo-pulse principle of the selected transducer probes. The process starts with the generation of acoustic radiated time-harmonic pressure field which is usually represented by the Rayleigh–Sommerfeld diffraction integral (Zeng and McGough, 2008) described in equation (3.7), followed by the spatial impulse response of the transducer probe (pressure field) which account for the intensity of propagated pressure beam, can be described as emitted pressure field in a homogenous

medium in equation (3.8), Acoustic pressure is then propagated, transmitted and received through the transducer equation (3.12). The resulting time delay between transmitted pressure field, the impulse response field and reflected signal can thus be summed up in equation (3.13) as the close loop detection to determine the transducer distance from the object which is required for image reconstruction purposes. Hence, the received scatter field pressure in the transducer, provides the information about the depth and angle of received signal are converted from polar to cartesian coordinates which is necessary for display. The acoustic spatial impulse response pressure field for wave equation is represented by the Rayleigh-Sommerfeld diffraction integral equation (3.7) as:

$$p(x,y,z;t) = j\rho cke^{jwt}\int_{S'} v(r')\,\frac{e^{-jk|\vec{r}|}}{2\pi|\vec{r}|}dS' \qquad (3.7)$$

The equation factors in the material or medium density ρ, speed of sound c, acoustic wavenumber k, driving velocity w, distribution of normal velocity v, on the rigid baffled radiator surface area S' and the distance between the source and observation coordinate (*x, y, z; t*). While $\vec{r}$ represents the position of the field point in space and j remains the imaginary electronic unit. The emitted pressure field, in a homogenous medium is then calculated using the surface velocity v_n of the transducer in equation (3.8) thus:

$$p(\vec{r},t) = \rho\,\frac{\partial v_n(t)}{\partial t} * p(x,y,z;t) \qquad (3.8)$$

The spatial impulse response of the whole transducer for a linear array can thus be expressed in equation (3.9) as:

$$P_r(\vec{r},t) = V_{pe}(t)t * \ f_m(\vec{r})\,r * \ T_{pe}(\vec{r},t) \qquad (3.9)$$

Where V_{pe} is the impulse-echo impulse, excitation impulse response during transmission and reception of echo; f_m is account for the tissue inhomogeneities to density propagation and T_{pe} as the pulse-echo spatial impulse response for transducer geometry and scattered field. This can be rewritten explicitly in equations (3.10 and 3.11) as:

$$V_{pe}(t) = \frac{\rho}{2c^2} E_m(t) * \frac{\partial^3(t)}{\partial t^3} \quad (3.10)$$

$$f_m(\vec{r}) = \frac{\delta\rho(\vec{r})}{\rho} - \frac{2\delta c(\vec{r})}{c} \quad (3.11)$$

$$T_{pe} = p_t(\vec{r}, t) * p_r(\vec{r}, t) \quad (3.12)$$

$$h(x, y, z; t) = \sum_{i=0}^{N-1} a_i h_i(x, y, z; t) \quad (3.13)$$

Therefore, the received response for a single RF line can thus be calculated by summing the response from a collection of scatterers (equation 3.13) which aggregates the impulse response from transmitting and receiving transducer and then convolving with the impulse response of the transducer (Jensen and Munk, 1997). Therefore, the spatial impulse response of the whole transducer surface area; *i = 0,1,2 …, N-1* is the index of the element of the transducer array; a_i represents the weighting level at element *i*: $h_i(x, y, z; t)$ is the spatial impulse response between the element *i* and the field point *(x, y, z)*. Generally, the scattering strength is determined by the density and speed of sound perturbations in the tissue and the surrounding (Cheng *et al.*, 2011).

3.3 Echocardiogram Image Resolutions

The ability of ultrasound probes and processing units to distinguish pulse-echo in spatial and temporal space is defined by its resolution component and plays a significant role in the choice of probes, and the quality of the generated echocardiogram.

But to acquire high quality images and accurate measurements require operator's skilled competence (Aschkenasy *et al.*, 2006); (Sprawls, 2014) and many years of training experience without which exacerbates the reliability issues and increase possibility for misdiagnosis (Yoon, Kim and Chang, 2021). Ultrasound image resolution is defined by its spatial and temporal resolutions (illustrated in Figure 3.8) which further define the critical ability of echocardiogram to accurately show the underlying anatomical changes over time.

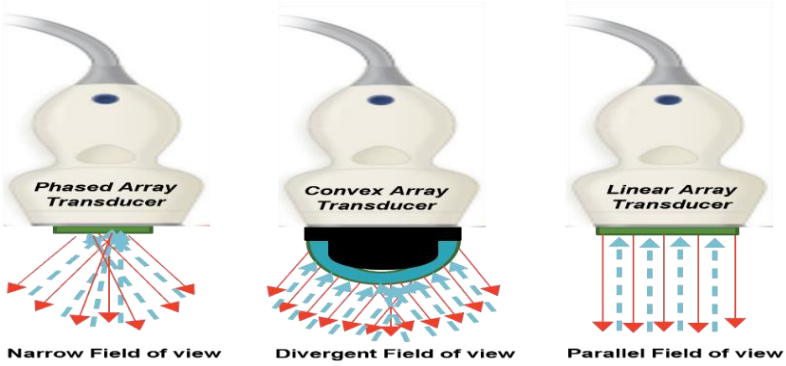


Figure 3.7: Probes showing typical field of View. (Left): Phased-Array Probe with small aperture to yield a divergent far field view, useful application in cardiology, (Middle): Convex Array Probe (curvilinear) with divergent beam, useful in abdominal exams (right): Linear Array Probe with wide field of view is suitable for vascular and tissue other than heart. (Merritt, no date)

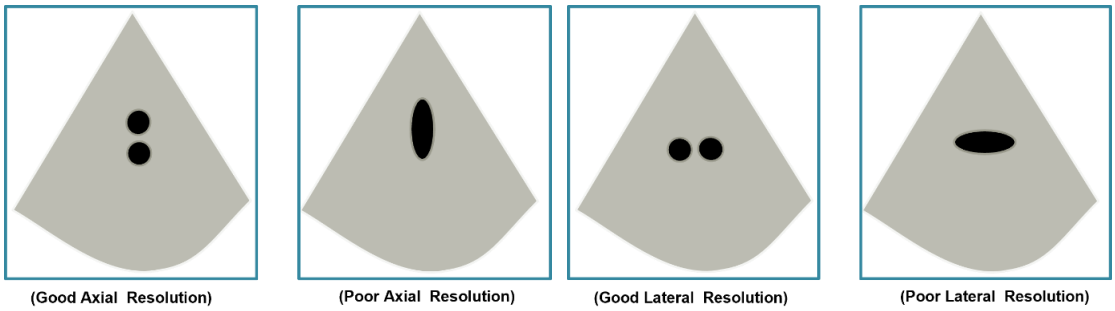


Figure 3.8: Illustration of spatial resolution: good and poor axial resolution describes the ability to display small targets as separate when two targets are on the path of ultrasound beam. Axial resolution is defined as a half of spatial pulse length (SPL) – the shorter the beam the better the axial resolution and lateral resolution, showing good and poor resolution respectively. Ability of echocardiograms to present two separate targets perpendicular to the beam can be classified as poor, average, or good. The wider the beam, the poorer the lateral resolution.

According to the Consortium for the Accreditation of Sonographic Education (CASE), a recognised body that accredits ultrasound courses in UK Universities, it takes a minimum of three (3) years to train an operator on freehand probe manoeuvring for high quality image acquisition (*BMUS*, 2021); (CASE, 2019) much more challenging when operators must deal with multiple patients in clinical settings (SCoR & BMUS Guidelines, 2020). Evidently, there is a direct relationship between the quality of acquired echo images and the operators' experience, objectivity, and consistency in probe manoeuvring, unfortunately, this is a persistence occurrence that imposes limitation on optimum image acquisition hence, the development of artificial intelligence-based tool for cardiac image optimization discussed in this thesis would be of a great importance to cardiologists' clinical practice.

3.4 Deep Convolutional Neural Networks

The concept of deep convolutional neural network (DCNN) enables the possibility of identifying discrete or continuous pattern which is associated with our natural and abstract worlds. Technically, DCNN consists of several blocks of hierarchical functional nodes (layers) which provide logical transformation and each layer's input going to its adjacent layers. The simplest architecture block arrangement usually starts with input and convolution (ConvNet) layer and a combination of one or many of the following data transforming layers: Activation Layer, Pooling Layer, Batch Normalisation (BN) Layer, Flattening, and Fully Connected (FC) Layer (Howard *et al.*, 2017). The art of combining these layers (model architecture) and the choice of spatial parameters defined for each layer constitutes a design task usually meant for experienced researchers and represents a distinguishing factor between several other model architectures. The building block layers can be summarised as follows:

- **Input Layer:** The input layer holds the raw pixel values of the input data that in the case of the echocardiographic data; the width and height of the input layer are the spatial dimensions of a single frame which will be exposed to the network.

- **Convolution Layer:** Convolution layer consists of 2 major components; Kernel (2D Filter), stride and can accept input data of 1D, 2D, 3D. While 1D is suitable for single dimensional input data like voice, text, the latter are suitable for images and videos. This layer holds the raw pixel values of the images known as spatial input and consists of a kernel or filter of fixed size parameters which slides in a window fashion to perform convolution operation on the windowed image thereby extracting spatial features. Several nodes could be stacked together depending on the application and perceived complexity in the image, for example, medical images do not present clear edges as compared to camera images. Also, padding could be applied to the size of the input image to overcome uneven mapping with filter size. Consequently, the output of convolution layer thus yields a feature map $a^{l}_{i,\,jk}$ as described by the equation 3.14 where $a^{l}_{i,\,jk}$ is the output feature map of the i^{th} kernels, of convnet l layer when the weight matrix $w^{l}_{i,mn}$ convolved with the input feature map $a^{l-1}_{(j+m)}$ of the current convolutional layer.

$$a^{l}_{i,\,jk} = \sum_{m=1}^{M} \sum_{n=1}^{N} w^{l}_{i,mn}\, a^{l-1}_{(j+m)(k+n)} \qquad (3.14)$$

- **Activation Layer**: The activation layer is a nonlinear function to convert the output of the convolutional layer to an output that can be used in the next adjacent ConvNet layer. This is a part of regularisation technique used in forward propagation training phase to make a nonlinear transformation which allows the estimate of complex functions by mapping the estimated values between 0 to 1. Activation acts on sum of weighted input in the form of cumulative weight and bias ($w1 * 1x + w2 * x2 + w3 * x3 + 1b$) and return the output neuron with range constraints i.e., 0 to 1, -1 to 0 as illustrated in the equation 3.15.

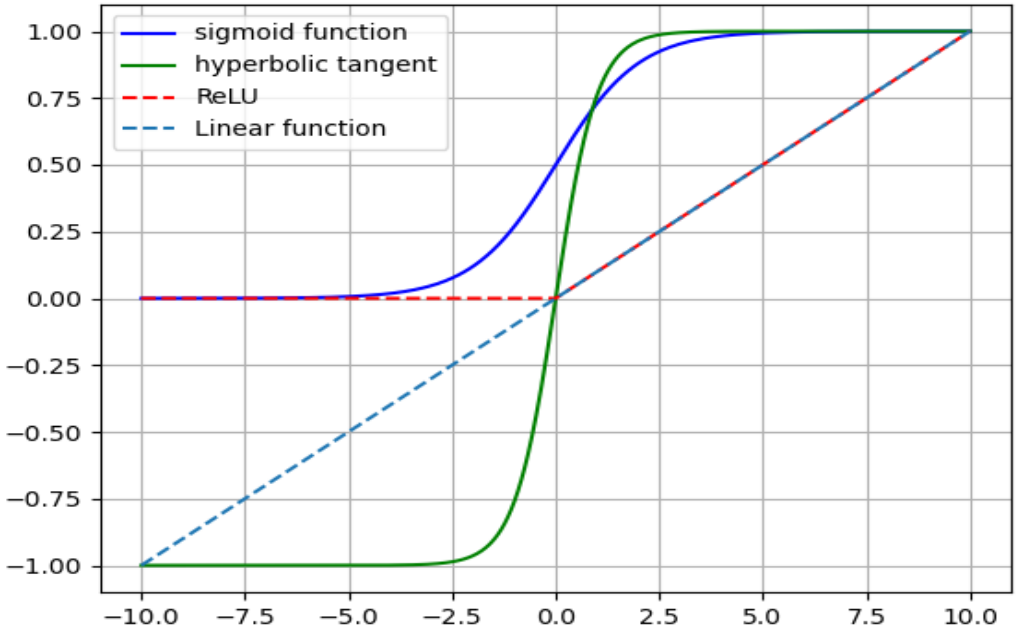


Figure 3.9: Illustrating popular activation functions used in this research CNN models Function flatten rather quickly, mapping the estimated values between zero and one (0, 1).

This research work has experimented with different types of activation functions (illustrated in Figure 3.9) with ReLU as a best choice yielding best convergence and efficiency for the algorithm and pipeline use case. Consensus on ReLU [0 to inf), or its variant LeakyReLU (-inf to inf) proved these as monotonic functions. Sigmoid (0 to 1) with smooth gradient but steeply characteristic along x-axis or SoftMax (0 to 1), best in handling multiple classes and predicting output multi-class probability. Also, TanH (-1 to 1) which functions like sigmoid but zero cantered, good for modelling inputs with strong negative, neutral, and strong positives values.

$$activation = f(\sum_{i=0}^{N} w_i\, x_i) \qquad (3.15)$$

Nevertheless, ReLU activation function appears to be the most employed activations in deep learning models as it can learn fast in the large neural networks (Nair and Hinton, 2010).

- **Pooling Layer:** Pooling layer object, usually is applied before or after a nonlinearity function to convolutional feature map. The choice of pooling filter size is usually smaller than the size of the input feature map. Most research experiments go by filter size 2×2 pixels with a stride of 2 pixels in both x, y direction. Pooling layers use either Average/Maximum/Global kernel function which slides like a window across input data left to right, top to button fashion to extract dominant features that are rotational and positional invariant in the convolved data. The main function is to reduce the spatial variance of the convolutional features, speeds up convergence and in turn reduces the computational power required to process data. Average pooling calculates the average for each patch of the feature map. This means that each 2×2 square of the feature map is down sampled to the average value in the square while max pooling is a pooling operation that calculates the maximum value in each patch of each feature map. MaxPool has been found in experiment to work better than average pooling especially for cardiac image classification and regression problems. Max and Average pooling expressed respectively as:

$$z_f = \max\{\mathbf{s}\} = \max\{s_1, s_2, \ldots, s_n\} \qquad (3.16)$$

$$z_f = mean\{\mathbf{s}\} = mean\{s_1, s_2, \ldots, s_n\} \qquad (3.17)$$

- **Batch Normalisation:** This layer is also known as Batch Norm (BN) is a layer with learnable parameters defined with BatchNorm2d and only active when (batch_norm = true). Its primary function is to accelerate convergence by reducing internal covariate shift inside each batch. Batch Normalization performs well in reducing the chances for overfitting during training phase and comes handy in preventing error propagation into adjacent neural network layer. During the model learning phase when the parameters begin to drift, the mean and std output by each layer will also change. A change in mean and std of the output of one layer will cause a change in mean and std for all following layers. Unfortunately, the model has no way of correcting these changes in means and std once training has begun. Consequently, the individual observation in the batch becomes significantly different, the

gradient updates could appear clumsy and take longer to converge. The batch norm layer normalises the incoming activations, scale, and shift across the channel dimension of the input then outputs a new batch where the mean equals 0 and standard deviation equals 1 (Ioffe and Szegedy, 2015). It computes the mean B_μ and variance B_σ , $then$ subtracts the mean B_μ and divides by the standard deviation of the batch equation (3.20). BN algorithm process is computed as follows:

$$B_\mu = \frac{1}{N}(\sum_{i=1}^{N} x_i) \qquad (3.18)$$

$$B_\sigma = \frac{1}{N}\sum_{i=1}^{N}(x_i - B_\mu)^2 \qquad (3.19)$$

$$\vec{x}_i = \frac{x_i - B_\mu}{\sqrt{B_\sigma + \epsilon}} \qquad (3.20)$$

Finally, BN then scales and shift each channel of the normalised batch of input $\vec{x}_i$ using the learnable parameter filter, denoted as $\alpha, and\ \beta$ to yield normalised output in equation (3.21).

$$y_i = BN_{\alpha,\ \beta}(x_i) \qquad (3.21)$$

- **Dropout:** A dropout layer is useful during the training phase to prevent overfitting in neural networks. Dropout accepts layer's activation input and randomly sets a certain fraction of inputs referred to as (dropout rate) depicted in Figure (3.10b) are rationalised by setting their output activation values to 0 at each training update. The values of inputs that are retained are scaled up, so that their sum is unchanged during forward propagation phase. The dropout rate is the tune-able hyper parameter that is adjusted to measure performance with different values. It is one of the regularisation techniques like cross validation, batch size, data augmentation or early stopping. Typical dropout values are set between 0.2 and 0.5 percent.

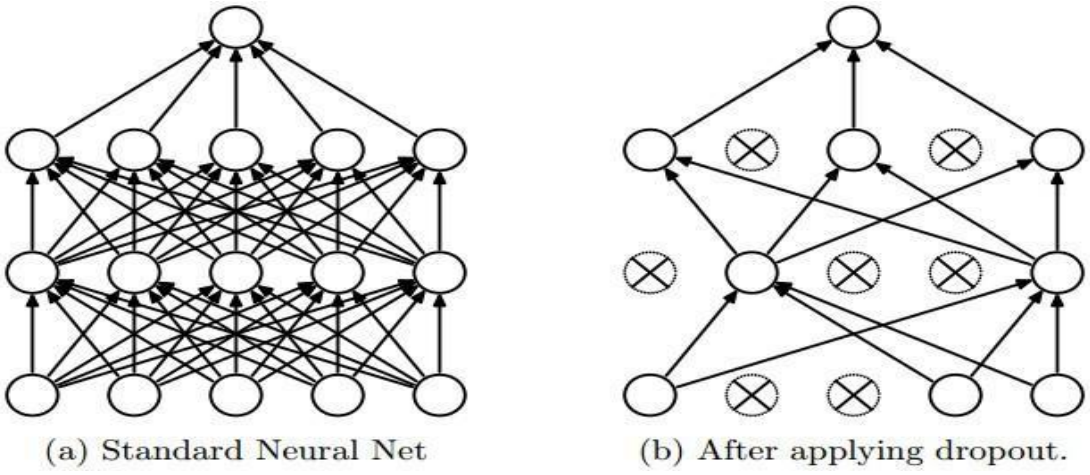


Figure 3.10: Illustrating a standard neural net with 2 hidden layers before and after randomly applying dropout during training phase, reproduced from (Srivastava et al., 2015)

- **Fully Connected Layer**: Fully connected layer (FC), also known as DENSE layer, where each neuron in the input is connected to each neuron in the output layer. Illustrated in Figure (3.10a), the FC layer is responsible for computing the predicted layer l, score by calculating the inner products of their input a_j^{l-1} with their associated weight parameters $w_{i,j}^l$ and biases (equation 3.22). FC layer is preceded by FLATTEN operation, an object component that converts the feature map of the final pool layer from two-dimensional features into one-dimension and is fed into a fully connected layer.

$$f_{fci}^l(a^{l-1}) = \sum_{j=1}^{n} w_{i,j}^l \, . a_j^{l-1} + b_i^l \qquad (3.22)$$

a) **Loss Functions**: One of the ways to evaluate the performance of objective function, deep learning algorithm either on discrete or continuous values is using loss functions. It provides visibility on algorithm current output and the expected output values. Also, can be described as feedback signal to guide algorithm performance and for back propagation and optimizations. This research used cross entropy loss for classification task, MAE or MSE for regression tasks or a combination of both. Cross entropy J loss is a measure of the difference of the randomness between two random variables.

$$L = -\sum_{i=1}^{N} y_i \log\big(h_\theta(x_i)\big) + (1 - y_i) \log\big(1 - h_\theta(x_i)\big) \quad (3.23)$$

3.5 Deep Learning Model Architectures

Deep learning models consist of few or several layers of convolution neural networks which are made up of hierarchical nodes (layers) of non-linear processing networks for supervised learning, automatic feature extractions, transformations, and classification. The combination of these network layers forms the basis for different model architecture in the public domain. They are capable of being trained on large sets of labelled data (supervised training) and learn complex features directly from data (self-supervised), without the need for manual feature extractions. In research literature, neural architectures are created to provide informative features for classification or regression tasks. Historically, including LeNet (LeCun, Bottou et al., 1998), AlexNet (Krizhevsky, Sutskever and G. E. Hinton, 2014), VGG'16 (Simonyan and Zisserman, 2014), GoogleNet (Szegedy, Wei Liu et al., 2015), ResNets (He et al., 2016), DenseNet (G. Huang et al., 2017), etc. The summary of the evolutionary trends in common DCNN architectures that have been widely used:

- The **LeNet** architecture proposed by (LeCun, Bottou et al., 1998) for handwritten and machine-printed character recognition in the 1990s. The LeNet neural network contains two sets of convolution, activation, and pooling layers, followed by a Fully Connected (FC) layer, activation, and another FC layer and finally a SoftMax layer

- The **AlexNet** architecture includes eight layers such as five convolutional layers and three FC layers. After each convolution and FC layer, ReLU is applied (Krizhevsky, Sutskever and G. E. Hinton, 2012); (Srivastava *et al.*, 2014).

- The **VGGNet** is a well-known convolutional neural network proposed by (Simonyan and Zisserman, 2015) and was used to win ILSVR (Large Scale Visual Recognition Challenge, 2014) competition. This model makes an improvement over AlexNet architecture (Krizhevsky, Sutskever and G. E. Hinton, 2012) by increasing the depth of the network. The architecture consists of building blocks of two convolutional layers followed by a pooling layer. This block is repeated multiple times, whilst all the convolution kernels are of size 3×3. Finally, a stack of convolutional layers is followed by three FC layers. Also, by introducing the number of layers (i.e., 11, 16 and 19) different architectures were proposed and the VGG '16 with a total of 16 layers as shown in Figure 3.11 recommended to have the best performance.

- The **GoogLeNet** neural network was introduced by (Szegedy, Wei Liu et al., 2015) who proposed an Inception Module that reduces the number of parameters in the architecture. The Inception Module applies multiple convolutional filters for the same input and concatenates the result. The network consists of 22 layers.

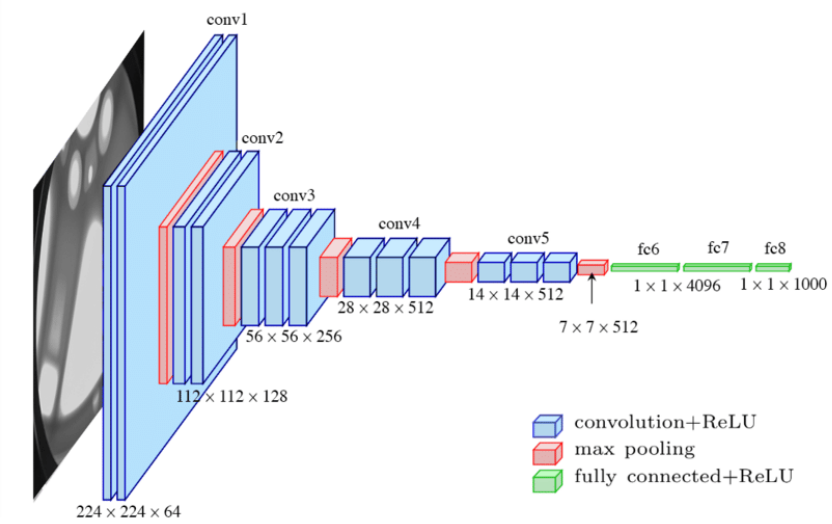


Figure 3.11: VGG16 Model Architecture. Reproduced from (Simonyan & Zisserman, 2015).

- **ResNet** Residual Network (ResNet) was inspired by VGGNet and won the ImageNet contest in 2015, a year after VGGNet opened the floodgates of deep convolutional neural networks. Unlike VGGNet with 16 to 19 layers, ResNet allows extremely deep neural networks of 150 layers and is considered a fundamental breakthrough but at computational cost of memory. The least version of ResNet in ResNet18 has a total trainable parameter of 11.174 million and is not regarded as a lightweight embedded architecture. Deep learning experiments have revealed that stacking many network layers without changing the network structure would diminish the model performance because gradients of network parameters will vanish as ConvNet's depth is increasing. ResNet, which suggested a residual learning framework through adding identity-mapping shortcuts (Xie *et al.*, 2017) was a better response to such experimental challenges. The ResNet model utilised four modules comprising residual blocks, each of which uses several referred blocks with the same number of output channels. The number of channels in the first module is the same as the number of input channels. Each residual block has two 3 × 3 convolutional layers with the same number of output channels. Each convolutional layer is followed by a batch

normalisation layer and a ReLU activation function, except the last operation of a block that does not have the ReLU. There are 18 layers in total. Therefore, this model is commonly known as ResNet-18. By configuring different numbers of channels and residual numbers of channels and residual blocks in the module, different ResNet models have been created such as ResNet-152 (He *et al.*, 2015). Figure 3.12 illustrates the architecture of ResNet-18.

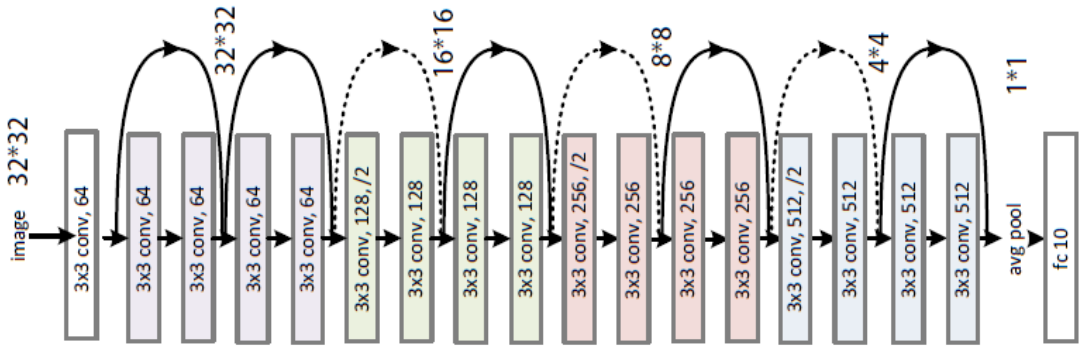


Figure 3.12: Illustration of the ResNet-18 architecture adapted from (He et al., 2016).

- The **DenseNet** presented by (Huang *et al.*, 2017) that in a feed-forward fashion, connects each layer to every other layer. It includes a convolution operation or pooling layers, batch normalisation, and an activation function. DenseNet concatenates the output feature maps of the layer with the incoming feature maps as illustrated in Figure 3.13.

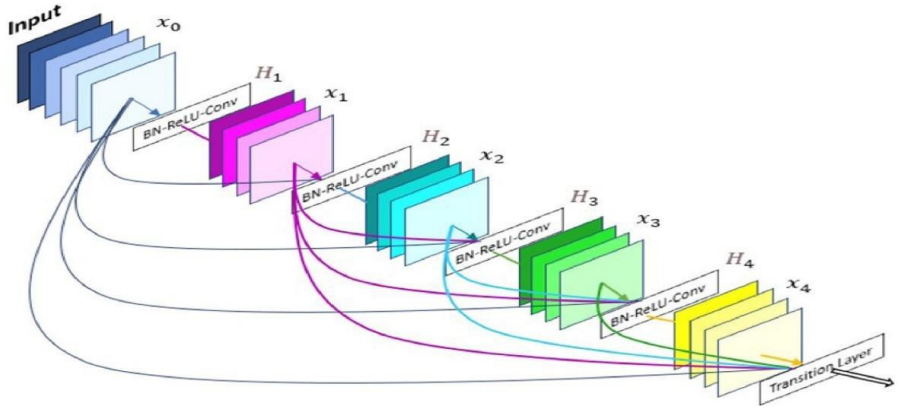


Figure 3.12: A 5-layer dense block showing layers' input connection [Con = L(L+1)/2]. Each layer accepts preceding future-maps as input while its own feature maps become input to subsequent layers. Image adapted from (Huang, G., et al., 2017).

The significant connection between these architectures is their application in image classification, identification, and regression tasks. Although, performance of these architectures varies significantly depending on the dataset applied, but they can reduce development time by avoiding training the model from scratch. This saves valuable times in modelling and can yield impressive accuracy in transfer learning modes.

Furthermore, the depth of the DCNN network have shown significant impact on the performance of the model. Therefore, getting deeper without applying changes in the structure can potentially retard model's performance, leading to loss of information and vanishing-gradient problem (Liu *et al.*, 2018).

For classification and regression tasks, this study investigated the performance of most of the state-of-the-art models but based on performance and inference times, settled on VGG16, ResNet50, and MobileNetV2 suitable for obtained dataset on PACS, PACS-1 and CAMUS public dataset.

3.6 Time Series Regression Model

Regression techniques are widely employed to solve tasks where the goal is to predict a continuous value or several continuous values simultaneously. Regression models assume some sort of linear or polynomial relationship between dependent and independent variables. Although, some tasks, like multivariate regression implemented in this research, do require prediction more than a single numeric value, consequently require a complex build of parallel stream of subnets known as multi-output regression model. In computer vision, regression techniques span a large ensemble of application scenarios such as: image classification (Nafchi and Cheriet, 2018), human pose estimation, age estimation, object detection, quality estimation (Liao *et al.*, 2019) or analysis of image virality. Besides the normal classification architectures, regression could also be formulated as a classification problem (Lathuiliere *et al.*, 2020). In that case, the output space is generally discretized to obtain class labels, and a multi-class loss is minimised. Conversely, a ConvNet with a fully connected layer replaced with linear or sigmoid activation as its final activation layer will function as regression architecture (Lathuiliere *et al.*, 2020). Several configurations are possible which can extend ConvNets' input function into spatial, and temporal (time) space dimension as utilised in experiment in Chapter 5, 6, 7 and are summarised as follows:

- **Recurrent Neural Network** (RNN), introduced by (Elman, J., 1990) is a standard neural network that has been extended across time-space by having edges which feed into the next time step instead of into the next layer in the same time step (Figure 3.14). RNN algorithm is best at recognising sequence data for example, a video frames sequence, speech signal, or text (Donahue *et al.*, 2016). It has become the most popular method of performing classification or regression analysis on sequences data. Nevertheless, a RNN, expressed as $h_t = \sigma(Wx_t + Bh_{t-1})$, where W and B are the weight matrices connecting the inputs and the recurrent outputs Y, shows a major drawback during backpropagation through time. The gradient accumulates and explodes or vanishes down to nothing. Consequently, in practical applications, a subset of RNN like long-short term memory networks (LSTM networks) is highly recommended.

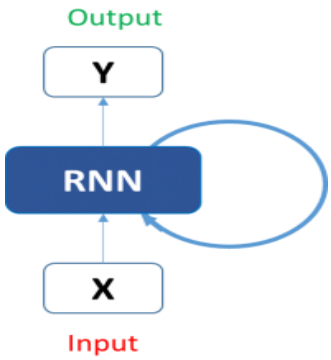


Figure 3.13: A Cell structure of RNN, showing how weight matrices connecting the input and the recurrent output.

- **Long Short-Term Memory** (LSTM), unlike RNN with inherent gradient problems, LSTM consists of a unit of logical cell specifically designed to reduce the vanishing gradient problem sufficiently to make recurrent neural networks more useful for long-term memory tasks like sequence predictions. Thus, allowing deeper networks and recurrent neural networks to perform well in practical application. There is a need to reduce the multiplication of gradients which are less than zero. LSTM reduces the multiplication effect of small gradients by adding processed input to its internal memory state. Therefore, the time dependence and effects of previous inputs are again controlled by the forget gate, which determines which states are remembered or forgotten (Ullah *et al.*, 2018). Figure 3.15 illustrate an LSTM cell with two other memory state or gates, i.e., the input gate and output gate.

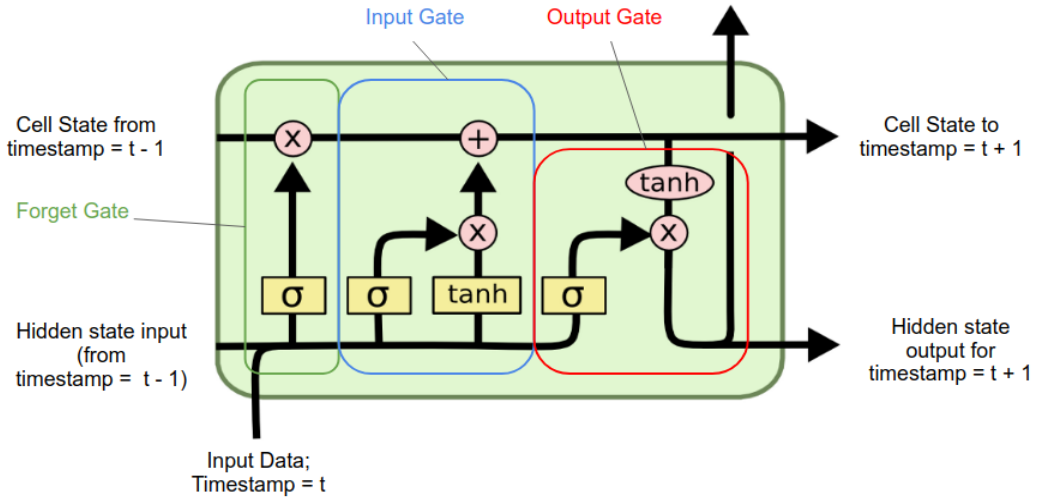


Figure 3.14: Structure of LSTM cell diagram showing three inputs: Previous hidden state (Ht-1), Previous Cell state (Ct-1) and current cell input (X_t).

- **Hybrid Network**: A combination of several layers of ConvNet's with LSTM or C-LSTM is often considered especially when the generalisation ability of the default classifier is not very satisfying on image sequence data. Consequently, combining the spatial feature-learning ability of DCNN model architecture with LSTM architecture's high-level temporal dependent features provides best performance for predicting video sequences (Luong *et al.*, 2021). For 2D echocardiography, deep convolutional neural networks could be used as a base model to learn low-level spatial features of 2D cardiac specimen image quality while feeding the resultant vector into an LSTM or any of its variants' model architecture, to learn high-level temporal features. This combination provides a combined spatio-temporal value estimation or prediction values for sequence of echocardiogram frames. The implementation of a hybrid spatio-temporal model architecture is illustrated in Figure 3.16. Echocardiograms are sequence frame data stream usually in varying time stamp present a spatio-temporal features that is best suited for time series classification. This advantage can be harnessed since echocardiogram presents fast moving frames of the myocardium. Consequently, justifying the adopting of hybrid network on quality assessment task.

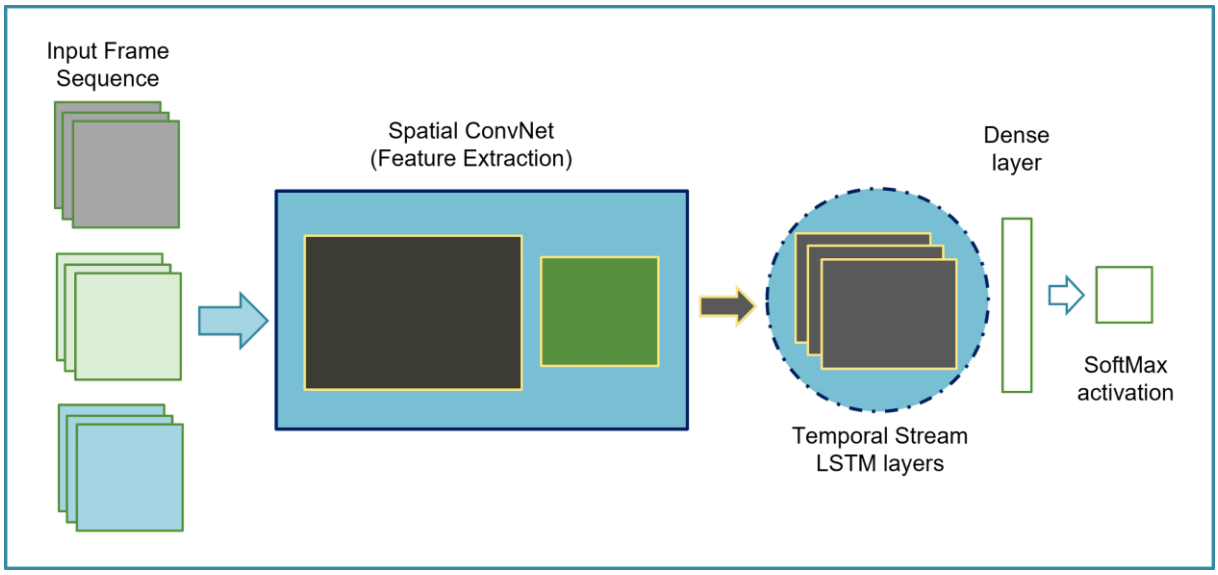


Figure 3.15: Illustrating a Hybrid Model Architecture consisting of DCNN and LSTM models to achieve superior generalisation, performance, and computational efficiency on sequence of images or video frames.

3.7 Semi-Supervised Ensemble Models

Semi-supervised machine learning is a combination of supervised and unsupervised learning methods. It uses a small amount of labelled data (Ouali *et al.*, 2020) and a large amount of unlabelled data, which provides the benefits of both unsupervised and supervised learning while avoiding the challenges of finding a large amount of labelled data. Generally, the motivation behind semi-supervised classification is to employ many unlabelled data to help build a better classifier from the labelled data (Shiming Xiang, Feiping Nie, and Changshui Zhang, 2010) but this is not always the case. Semi-supervised is applicable to classification problem where the objective is to pseudo-label unclassified data samples with little available labelled data due to challenges in cost, time-tied limitation, data protection, or turn-around overhead. Semi-supervised learning (SSL) presents a layer of complexity and challenge to 2D echocardiograms, neither supervised nor unsupervised learning algorithms can make effective use of the mixtures of labelled and unlabelled data. Hence, specialised semis-supervised learning algorithms are required. This approach to machine learning is a combination of supervised machine learning, which uses labelled training data, and unsupervised learning (Figure 3.17), which uses unlabelled training data. This work presents the implementation of a semi-supervised classification algorithm using stackable model in ensemble learning framework to implement clustering and label predictions.

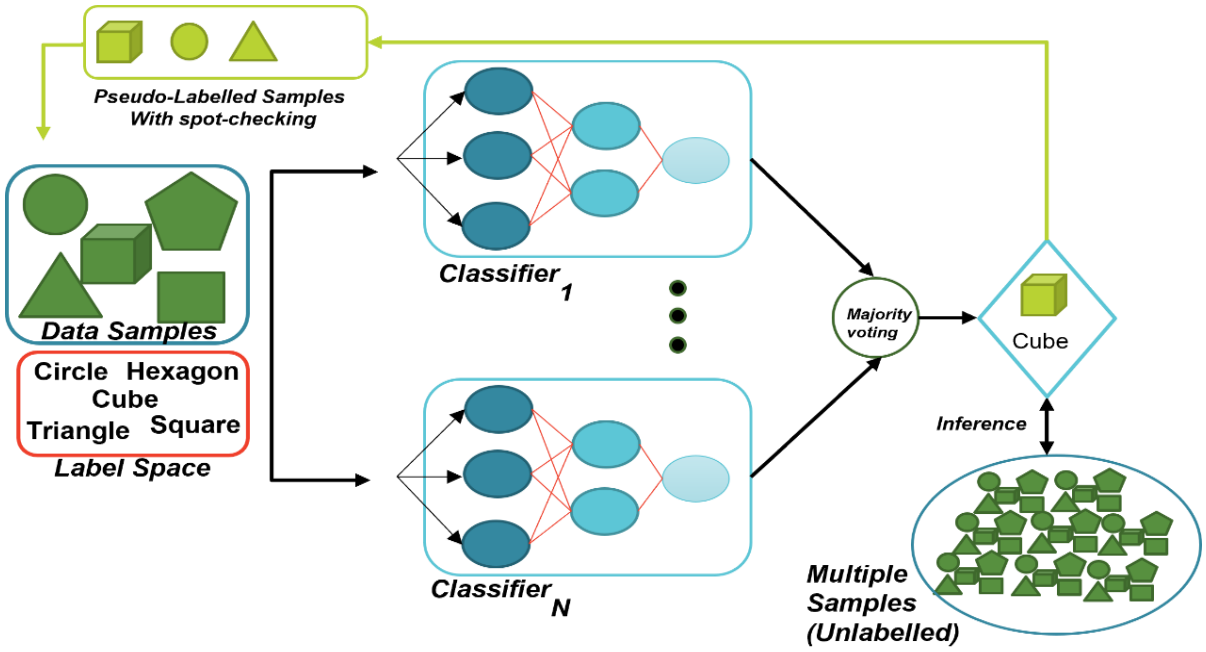


Figure 3.16: Illustration of Semi-supervised with ensemble learning Architecture adapted for the characterization of 2D echocardiograms and pseudo-labelling process.

The ensemble method in semi-supervised provides some sort of confidence and robustness in the pipeline. Its general architecture is illustrated in (Figure 3.17) and iteration processes as follows:

(a) The ensemble models are trained on small number of labelled samples that represents a minute sample across the domain of label space

(b) Model enters second iteration where it predicts on unlabelled samples to yield pseudo-labelled samples since they may not be accurate

(c) Predicted labels are linked to the label space and data inputs are linked to the unlabelled data samples and are added to the separate buffer to allow spot checking prior to addition into labelled samples

(d) The iteration then pauses prompting action for spot checking on predicted label samples before they are finally added into the initial pool of labelled samples. The successful samples are then deducted from the pool of unlabelled samples.

(e) Finally, the model retrain with additional new label samples and this process repeats until no unlabelled samples remain.

3.8 Overview of DCNN Design Methods

Designing deep convolutional models has been described as a non-trivial task however, requires a skilled expertise and experience to design and train. For example, an important consideration for adopting any type of model architecture on ultrasound protocol would depend on technical details that are specific to general integration and hardware requirements. Design methods are essential considerations and components for system integration done through the system engineering process. The aim is to unify product's components and process components into a whole unit with the hope of achieving overall system purpose, satisfy use case, and functional efficiency.

A couple of design approaches have been considered, this could be a traditional method of calculating parameters for each DCCN component/layers while automating the optimization of layers' learning rate (Semi-Automatic) or using any auto search methods like AutoML or neural architectures search method (NAS). Both methods are implemented with choice for adaptation based on the evaluation outcomes.

3.9 Semi-Automatic Approach in NAS

The publicly available state-of-the-art models provides a baseline performance for any application using DCNN. For light-weight applications (Vaseli et al., 2019), the search strategy adopted was to imitate a physical life scenario where deployment of a specific solution is urgently required and where few hours of delay can spell clinical disaster. Such a scenario was seen to have played out in the 2019, the beginning of Covid-19 Pandemic. Since automatic neural search could take days and weeks to wrap up with suitable architecture models, a semi-manual approach could become relevant and save the day. The advantage of translating into quick deployment and comparative performance accuracy can make a huge difference in life scenarios.

Technical justification for a lightweight model is based on the set of requirements envisaged for the deployment of quality assessment algorithm on either mobile platform, for example, point of care ultrasound POCUS or on system-on-chip SoC platforms where memory requirement is limited to few kilobytes. For the application of clinical echocardiography, the justification for a lightweight model can be summed up as follows:

- Requirement for low memory application, e.g., POCUS, (<100MB)
- Low storage pipeline requirements, (i.e., # of parameters < 20 million)
- High performance data access specification (> 16GB),
- CPU inference speed (frame per second) (<25ms)
- Dataset requirements & scalability
- Low latency remote connection

3.9.1 Lightweight Model Architecture (CardioQNet)

Manual architecture design which could involve minor to major structural changes of any existing state-of-the-art models for adaptation or creating entirely new architecture to solve specific problems (Abdi, Luong, Tsang, Allan, *et al.*, 2017). Although, the major disadvantages with manually designed models is that they are incapable of continuous improvement in terms of data space, after the tuning hyper-parameters converge. However, the advantage could yield a lightweight model where numbers of redundant parameters are stripped of to achieve specific memory requirement and inference speed. like Figure 3.10, that works on low hardware memory specification, improved performance and saves implementation overhead because of specific adaptation to a given hardware and problems. Manual search methods implemented include a few layered components which can generalize on cardiac dataset (PACS) comparable to ResNet18 performance.

This study explored the manual design of a lightweight model architecture development using semi-automatic approach. The design of cell parameters was first implemented in excel worksheet to gain insight into the spatial parameters of the input images and memory requirements on a given batch size. The architecture search space consisted of a normal cell and a reduction normal cell were manually chosen based on persistent evaluation of set of filters sizes that defined the convolutional layers. The outcomes of manually chosen parameters for each cell was then evaluated and either rejected or chosen based on the performance on the specific dataset (PACS or its variants). This process is repeated with hyperparameter optimization until no appreciable increase in performance is recorded for the given architecture.

Although, manual design tool as explained is incapable of predicting the model's cell performance unless it is executed, the numbers of training parameters and required memory for

input image can be evaluated without running the model which is a design advantage compared to AutoML methods. This process yielded a derived 3 layered architecture we referred to as 'CardioQNet' model, depicted in Figure 3.18.

To evaluate the derived architecture requires running and fine-tuning the model's hyper parameters until the performance ceiling is reached. Memory requirements is based on the spatial size of the input image to each convolution layer and calculated using:

$$spatialWidth = \frac{(W - F + 2 * P)}{S} + 1 \qquad (3.24)$$

Where W is the spatial width of the image to be convolved with kernel size F and padding P values divided by the stride S values. This is valid for layers of convolution and max pool layers which are used to down sample the spatial dimensions for faster convergence.

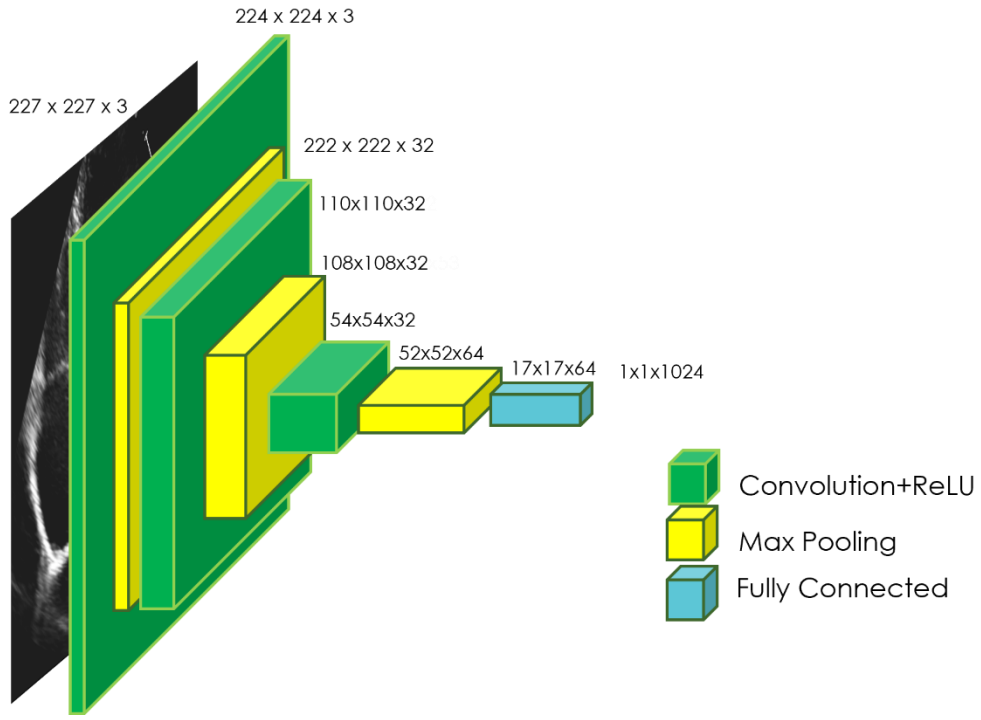
Hyper parameter optimization: Automated tuning and optimization of CardioQNet hyper parameters depends on selecting appropriate hyper-parameters values. However, it is not an easy task because it requires time and expertise to tune the hyper-parameters to fit the machine learning model (Domhan, Springenberg and Hutter, 2015). This study has considered the grid search method in the optimization of Hyper-Parameter (HP) for CardioQNet, the proposed lightweight model.

The grid search method is considered as the brute-force way of searching HPs, with defined lower and higher bound along with specific steps. Grid search works based on the cartesian product of the different set of values, evaluates every configuration and returns the combination with the best performance (Zahedi *et al.*, 2021). Its implementation is simple but inefficient for large search spaces due to its adjacent boundary values. The problem of inefficiency on a very large dataset can grow worse with increased data dimensionality or when label space increases.

Random search is another common optimization standard method of searching HPs. In the random search method, instead of evaluating every configuration, optimization values are chosen randomly and repeat this process until the defined resources are over. Random search method is by experiment, much faster than grid search but presents no clear path on how optimal parameters are obtained (Greff *et al.*, 2017).

Another practical search method known as manual search is the most basic HP tuning method and a typical approach among researchers and students. Different starting values are repeated

in the search experiment on an ad-hoc automated basis. An ad-hoc experimentation values could benefit from certain domain knowledge or experience, nevertheless, the training process is automated and repeated at each given value until a satisfactory or improved result is obtained (Elsken, Hutter and Metzen, 2018). Cardiac dataset however, presents anatomical complexities with large search spaces making basic manual search an impractical approach. The effect of manual parameter tuning on performance was earlier explored by Mohammadi et al (2021) where different embedded parameters were combined which improved the accuracy of semantic auto-encoder in classification problems (Zahedi, Mohammadi and Amini, 2021).



ConvNet Layer1	ConvNet Layer2	ConvNet Layer3
32: 11x11	32:7x7	64:32x32
BN	BN	BN
32:4x4	32:4x4	64:4x4

Figure 3.17: CardioQNet - The ultralight-weight architecture, derived using semi-automatic neural architecture search method. This is optimised for 2-dimensional echocardiographic classification and regression tasks.

With automated hyper parameters tuning procedures, the derived model was then evaluated against model performance and accuracy with the publicly available DCNN architectures deemed suitable for medical imaging classification tasks. VGG16, ResNet50 and MobileNetV2 were among those evaluated for baseline performance, and assessment of data, and memory

requirements. Memory requirements for ResNet50 were found to be the highest, while the least memory requirement was for CardioQNet and are summarised in Table 1.1. Also, the tool allows evaluation of trainable parameters to display memory requirements for a given input image sizes and epoch's batch sizes. This was done for several conceptual architectures. Hence, the 3-layered convolutional architectures named 'CardioQNet' became the derived architectures after intensive optimization on the given dataset. Nevertheless, the semiNAS derived model, CardioQNet, adequately compared in performance with the selected state-of-the-art (ResNet50, MobileNetV2 and Vgg16) Architectures over the given dataset. While 2Cell-DARTS architectures claimed to show significant decrease in user memory requirements, CardioQNet illustrated in Figure 3.18, shows 93.73% decrease in memory requirement with faster inference time and a significant improvement in overall model performance relative to ResNet50 architecture as shown in chapter 5 (Table 4.1).

In terms of data space scaling, semi-automatic architecture search is less preferred to fully automatic neural search method however, it has been proven that significant advantages in semiNAS derived models include immediate conceptualisation and feasibility for agile deployment in an inevitable case like pandemic and ad-hoc development. Furthermore, the semiNAS method saves design time, computational overhead (expensive to own) and provides the ability to check the memory requirement (on a given data samples) for training and back propagation which is useful for targeted hardware.

3.10 Automatic Neural Architecture Search

The automatic search method otherwise known as neural architecture search (NAS) is regarded as a subfield of auto machine learning (AutoML), which application relates significantly to hyperparameter optimisation (Feurer, H. et al., 2019) and meta-learning (Vanschoren, J., 2019). As expected, ultrasound hardware has stringent memory and storage limitations hence, the realisation of an efficient network architecture (NAS), with low integrate-able memory requirements, lightweight architecture, and best performance with less human intervention comparable to state-of-the-art models is a technological milestone. In experimental outcomes, evidence that NAS techniques have outperformed manually designed architectures on some tasks such as image classification (Zoph, B., et al., 2016;) or segmentation (Chen, L.C. et al., 2018) has been established. This precipitates the idea that automated network architecture

design is feasible for regression problems, especially for 2D echocardiographic image quality estimations. NAS method involves three sequential procedures namely: search space, search strategy, and performance estimation strategy (Elsken, J.H., et al., 2018) and enumerated as follows:

(i) **Search Space:** One of the useable methods in neural architecture search space includes all sets of convolution layer configurations stacked on each other in a chain-structure which include a sequence of n layers, where the nth layer *ln* receives its input from layer n – 1 and its output will be as the input for layer n + 1. Then, the search space is parameterized by the number of layers, (n), type of operations such as dilation (Yu and Koltun, 2015), element of deep-wise (Chollet, 2017) convolutions, and hyper-parameters related operations (Baker et al., 2016).

Possible implementation of skip connections in NAS has yielded more complex architecture with multi-level NAS network solutions (Zoph, Vasudevan et al., 2018; Real et al., 2019a) which compares to ResNet, and DenseNet, where previous layers outputs are summed (Huang et al., 2017) for improved performance.

Furthermore, the cell-based method uses two cells as a basic dimensional array; one keeps the original input's spatial dimension and the other keeps the input's dimension that is decreased at every iteration. This method yields significant advantages of epoch-times, performance, and scalability in comparison with chain-structured methods.

Another method is based on hierarchical search space (Liu, T, et al., 2018) which includes three levels of operations: set of fundamental operations, the second level connects the fundamental operations through a directed acyclic graph, and the third level encodes how to connect the second levels and so on. This provides an advantage over cell-based methods. This thesis adopted the cell-based search space method for view classification model and hierarchical search space approach for regression model.

(ii) **Search Strategy:** Neural architecture space can be investigated using random, Bayesian optimization, Reinforcement learning (RL) (Zoph and Le, 2016; Zoph, Vasudevan et al., 2018), gradient-based (Williams, 1992), and evolutionary (Miller, Todd and Hegde, 1989) search strategies.

Experimental records show the competitive performance of Bayesian optimization and RL search strategies on CIFAR-10 dataset, even though they are significantly high in computational overhead. Research efforts to reduce this overhead have been in the public domain since 2013 (Bergstra, Y., and Cox, D., 2013; Zoph, B., and Le, Q.V., 2016).

In evolutionary algorithms, a population of the possibly trained network is developed and, in every step, at least one model will be sampled from the population and serves as baseline where offspring would be generated by adding or removing some layers and possibly altering the layer's hyperparameters, adding skip connections, a process known as local operations. Finally, the performance of all the offspring is evaluated and are added to the final model structure. (Elsken, J. H. Metzen and Hutter, 2018b).

It has been affirmed that RL and evolution strategies performed significantly well in finding smaller models and in terms of performance accuracy preferred to random search strategy (Real et al., 2019a). Therefore, this thesis implemented the evolution search strategy for the view classification model.

(iii) **Performance Estimation Strategy:** One of the performance estimate strategies is based on the comparison of the model's predictions (validation data) with the known values of the dependent variable in a dataset (training data). Although, its highly improbable that predictions and dependent-variable values are often equal in most cases, hence the need to estimate the disagreement in terms of how well the model performed on a given dataset. This research work investigated four (4) strategies: lower fidelities, learning curve extrapolation, network morphisms or weight inheritance, and one-shot weight sharing strategies.

a. ***Low fidelities:*** Low fidelities strategy refers to performance measuring strategy where computational overhead is downscale to improve turn around. This strategy involves: (i) reduced training (epoch) times, (ii) training with fewer cells and kernels per layer, (iii) training on lower-resolution images or (iv) training on a sample of the dataset (Zoph, V., et al., 2018); (Klein, et al., 2017); (Chrabaszcz, L., and Hutter, F., 2017). Consequently, a risk that computed performance will underestimate the quality of architecture predictive performance becomes imminent.

b. ***Learning curve extrapolation:*** Learning curve extrapolation strategy (Domhan, T., et al., 2015) is based on the implementation of probabilistic model that is capable of extrapolating model's performance as a function of its number of iterations or training time with a view of minimises early termination consequences that could occur during model's training phase. Another variant of this strategy is the performance as a function of available training dataset. Learning curve extrapolation has been demonstrated to speed up hyper-parameter optimization.

c. ***Network morphisms:*** Involves the adaptation and inheritance of a novel model's weights without altering its network functions (Wei, T., et al., 2016). This allows substantial increase in network capacity and maintaining high performance without needing training from scratch. Although strict network morphisms may precipitate complex architectures by making architectures larger that it needs to be, this can be attenuated by using approximate shrinking network morphisms (Elsken, T., et al., 2018). Overall, it provides the benefit of allowing search spaces without an inherent upper bound on the architecture's size.

3.11 NAS for Classification Model

The evolution of Network Architecture Search (NAS) methods (Pham, H., et al., 2018; Xie, S., et al., 2018) have accomplished highly competitive performance in discrete label classification tasks (Zoph, B., and Le, QV., 2016). One of the earliest methods applied by Zoph (2016) was RNN, as the controller to compose neural network architecture successfully searched variable-length architecture space. Although RNN can be trained with a policy of gradient method to maximise model's accuracy on architecture space, it does require around 800 GPUs for 600 hours making this impracticable for real-life scenarios while other alternatives are feasible.

Another NAS implementation alternative includes learning the structure of CNN using the RL and evolutionary algorithms (Liu, C., et al. 2018). This approach uses a sequential model-based optimisation strategy, where structures are searched in order of increasing complexity, and learning a surrogate model simultaneously, to guide the search through structure space. This method proved 8x faster in comparison to Zolph's RNN method (Zoph, B., et al. 2018).

Another competitive model using an evolutionary algorithm known as *AmoebaNet-A* (Real, E., et al. 2019) presented the first controlled comparison of RL algorithms for image classifier architecture search, among other methods. Implementing NAS approaches, such as Efficient Neural Architecture Search (ENAS) (Pham, E., et al., 2019) or the new Differential ARchiTecture Search (DARTS) (Liu, H., et al., 2018), have shown an advantage in reduction of cost of search requiring fewer GPU hours compared to ENAS and RL.

Since CNN architectures mostly contain a repeat of stackable blocks, finding a small optimal computational cell with considerably fewer layers, which makes up the building block of the final architecture, rather than searching for a complete network constitutes a specific advantage in DARTS implementation. Consequently, the search space size is therefore minimised, making DARTS a preferred choice for solving real-life problems. Although the success of NAS implementation was majorly reported on ImageNet and CIFAR, objective performance of DARTS on medical images and other scientific datasets have been significant. The DARTS method has been demonstrated to outperform ENAS, since it requires low GPU hours in its search process (Liu, H., et al., 2018).

The application of DARTS method for designing customised architectures to classify echo view images was investigated and published in a joint publication 'Neural architecture search of echocardiography view classifiers.' Authors contributions were duly acknowledged.

3.12 NAS for Regression Model

It is becoming increasingly important to design complex neural network architectures which can enable effective training by stochastic gradient descent to achieve competitive performance capable of comparing or exceeding state-of-the-art models. However, developing such an architecture model constitutes significant challenges especially, in medical imaging with Spatial and temporal features. Regression models are widely employed to tackle tasks where the goal is to predict continuous values (Lathuiliere *et al.*, 2020). Just like classification models, a regression algorithm consists of several convolutional layers, followed by a few fully connected layers, and a linear layer as its final activation function. For regression tasks, the loss function employed would either be the mean absolute error (MAE) or mean squared error (MSE) used in contrast to the cross-entropy loss function usually associated with classification

models. The overall architecture is referred to as a convolutional neural network (ConvNet) capable of extracting spatial features, which can be quantified for image's quality scores.

It is important to note that echocardiograms include both spatial and time dependent features therefore, general models for image classification are les suitable for extracting time dependent temporary features that are synonymous with echocardiographic frames. Due to this inherent level of complexity posed by echocardiogram frames, neural network algorithm that can explore the advantages of both ConvNet and LSTM models for spatio-temporal features is an ideal model. Therefore, ConvNet is capable of generalising feature selection for regression (Hu *et al.*, 2019) on echocardiograms with limited capability for computational complexity involving temporal features unlike the LSTM model or at best, the combination of both derived ConvNet and LSTM. This is regarded as a hybrid model of echocardiogram image quality assessment.

In the search for a competitive model for regression tasks, this research is aware that most NAS-derived networks for classification or for the prediction of continuous quality value are designed in DCNN and LSTM hybrid models with a view of exploring the repeatable cells to construct the model's backbone. However, the special case of image quality where several values are expected for a frame over time would require a multi-output regression model. In multi-output regression, two or more outputs are required for each input sample, and the outputs are required simultaneously.

Since DARTS-based method was the choice for view classification model in section 3.4, consequently, it could be rewarding and necessary to use the performance of specific downstream tasks to evaluate and search for good neural architectures for regression tasks. Therefore, this research has been adopted to encode neural networks with a differential architecture search (DARTS) and present an algorithm for hybrid architecture in the regression task of echocardiogram image quality. Furthermore, DARTS provided the most competitive architecture search methods, it is considered a best compromise since the LSTM model presents no elaborate or complicated layers than itself as a single convolutional layer. This analytical compromise can also generate satisfactory results in classification or regression tasks. The evolved network architectures require less space for network parameters and given the same amount of time, yielded a significantly lower error on average.

3.13 Conclusion

This chapter has provided some technical background on the principle of ultrasound echocardiography and the fundamental theory that underpins the formation of B-Mode image. Also, highlighted the advantage of the only in vivo imaging modality compared to MRI, X-ray or CT scan imaging. The complexities and challenges of poor ultrasound image resolution and the contrasting advantages of 2D echocardiograms in analytical modelling and myocardial assessment were discussed.

Deep convolutional neural network methodology along with the components of model architectures was discussed. Also, the state-of-the-art model in public domain were highlighted in comparison to a lightweight model architecture where inference speed can support real-time assessment was highlighted. Technical justification and the need for designing of new model architectures fitting the spatio-temporal requirement of echocardiograms was enumerated.

Details of relevant algorithm for classification, and regression tasks were discussed along with the advantages of supervised and semi-supervised model in medical ultrasound imaging. Moreover, an overview of the NAS including search space, search strategy, and performance estimation strategy has been discussed.

Chapter 4 applied semi-automatic search techniques to design an ultra-lightweight, efficient neural network (CardioQNet) deployable for real-time emergency scenarios in echocardiogram view classification. Then compares the model's classification accuracy with ResNet50, VggNet16, and MobileNetV2.

Chapter 5 presents the novel formulation of echocardiogram objective quality attributes, state-of-the-art methods of image quality assessment and proposes a robust architecture to quantify the quality score of echocardiograms (B-Mode) frame using multi-stream, multi-output regression model.

Chapter 6 focused on implementing a novel characterization of echocardiogram specimens (in explicit clinical terms) for anatomical global feature classification and automatic pseudo-labelling. The pipeline was implemented in the PyTorch backend to feature ensemble architecture for robust and high-performance classification.

Chapter 4

Echocardiography Frame Classification

4.1 Introduction

In transthoracic examinations, typical clinical protocol demands that the acquisition of several echocardiograms obtained at diverse probe positions and orientations is paramount to building a complete picture of patient's pathology and providing several views of the heart' anatomy. During the clinical acquisition workflow, specific probe positions are required to interrogate heart tissues in either apical, subcostal, suprasternal or parasternal acoustic windows (Ebadollahi, Shih-Fu Chang and Wu, 2004). Consequently, several different orientations of echocardiograms are produced, which can be classified into views such as apical four-chamber (A4C), apical four-chamber (A2C), parasternal long axis (PLAX) and other relevant views which are recommended for linear quantification or clinical measurements (Lang *et al.*, 2015).

A concerted effort to advance accurate analysis and diagnosis of cardiac functions has therefore elicited the inherent gaps in the image acquisition standards, incoherent interpretations, closely related features in cardiac specimen and ultimately suboptimal image quality. Hence, the analysis of echocardiogram begins with apical view identification or classification. This is done in a manual and time-consuming process that requires specialised training and is prone to observer variability, inaccurate measurement, and misdiagnosis.

A. Significance

Ultrasound echocardiography fall within two major classes, a two-dimensional (2DE) or three-dimensional echocardiography. Although, recent development has resulted in smarter transducers and fast image acquisition, 3DE is far less in use in clinical practice compared to

2DE. Except in the assessment of left ventricle (LV) where 3DE provided additional information where optimum quality images were assured (Ruddox V, et al. 2013), a 2DE provides better acquisition response for general practice and during clinical emergencies. Today, larger percentage of clinical workflow still depends on 2DE. Hence, 2DE is the preferred choice of assessment tool for most myocardial diagnosis. In 2DE, echo images in the same apical views compare similarly to each other unless to a well-trained eye therefore, distinguishing the elements that constitutes difference can be particularly challenging for an operator (Khamis, H., et al., 2017). Therefore, accurate automatic classification of echocardiogram views is paramount to several clinical assessments in echocardiography.

With the application of machine learning and computer vision has improved the accuracy and time-efficiency of automated image analysis, particularly automated interpretation of medical images (Smistad et al., 2020); (Zhou, S.K., et al., 2006); (Sassaroli et al., 2019). However, traditional machine learning methods are implemented using complex processes and tend to have restricted scope and effectiveness (Stoitsis et al., 2006) in contrast to deep learning models.

B. Justification

Experimental investigations have shown that a normal image classification model applied on 2DE/3DE, echocardiogram specimens could yield a phenomenon performance but significantly misleading in terms of classification results. While 3DE attract computational cost in contrast to 2DE, several operational and performance factors along with wider acceptance of 2DE have obtained research attention and the development of smart solution for 2DE. Therefore, this study focuses on 2DE. It is worth to note that echocardiograms consist of background noise which are not easily discernible by humans, computational cost in solving such problem is lesser and faster on 2DE.

In development algorithm, the use of successive echocardiogram images feeding a classification model in quicker successive frames (known as cine-loop), whose length could vary between 1.5 to 2.5 seconds have yielded significant results. Since hundreds of such cine-loops need to be classified, the convention where some extracted frames are classified could be technically insufficient unless the frames are chosen to represent all the inert actions present in the cine-loop length. Therefore, a normal image classification task would be unsuitable to capture the transient or temporal features in the cine-loop length which is required for adequate classification (Ren *et al.*, 2021); (Yang *et al.*, 2018).

To prevent miss classification and increase model confidence with stable prediction, time series classification algorithm (hybrid model) is thus apparent and required for echocardiography frame classification (Barros *et al.*, 2021).

Current advances in the automated design and application of deep neural networks have resulted in increased possibilities when automating medical image-based diagnosis (Coates et al., 2013; Bai et al., 2017). In most current clinical practice, images from different modalities are by regulation, managed and stored in picture archiving and communication systems (PACS). Moreover, add-on echo software packages, such as EchoPAC (GE Healthcare) and QLAB (Philips), are available to clinicians at extra costs. These applications claim to allow the automation of clinical analysis and diagnosis processes. These add-on packages do this in retrospect and may be unsuitable for real-time acquisition and analysis since they necessitate human involvement in detecting relevant views. Furthermore, the opportunity to improve workflow, guiding inexperienced users, reducing inter-user discrepancy, and improving accuracy for high throughput of echo data and subsequent diagnosis are highly advanced functions that add-ons package rarely contains in a single version. Therefore, automatic apical view classification could be clinically beneficial for pre-labelling large datasets of unclassified cardiac specimens (Zhang *et al.*, 2018).

This chapter focuses on the design method and dataset used to train an ultra-lightweight model on classification task using spatio-temporal model derived from semi-automatic NAS procedure. Firstly, it presents an overview of time spaced PACS dataset, detailed architecture, and model training, then highlighted the analysis of classification baseline studies using state-of-the-art models and performance of semiNAS lightweight model. Finally, the chapter compares model performances of state-of-the-art spatio-temporal models with semiNAS derived models. This chapter also highlighted main contributions achieved on apical view classification using 2-Cell DARTS and a detailed description of the proposed semiNAS classification models. Finally, the experimental setup, results, and discussion are presented including the summary.

4.2 Related Work

Majority of earlier research on apical view classification of echocardiographic images were based on traditional machine learning approach; this is synonymous to manual feature engineering process where hand-crafted image features are defined using traditional machine learning (ML) techniques. Although, traditional appearance-based have achieved varying degrees of classification success but with limited number of cardiac specimens among various anatomical views. This is because patients' heart structures vary based on anatomical profile and prognosis (Ebadollahi, Shih-Fu Chang and Wu, 2004).

Currently, recent success of deep convolutional neural networks in computer vision, and particularly for image classification tasks has proved beyond doubt that complex cardiac images are capable of objective identification and analytical adequacy. (G.N. Balaji, Subashini and Chidambaram, 2015). In the application of deep learning, which is found in newer works, deep networks architecture attempts to learn object's features majorly using discriminative learning approaches.

Some recent work includes Gao et al., who proposed a fused DCNN architecture by integrating a deep learning network along the spatial direction, and a hand-engineered feature network along the temporal dimension (Gao *et al.*, 2017). The final classification result for the two strand-network was obtained through a linear combination of the classification scores obtained from each network. Their dataset was obtained from 93 patients and consists of 432 image sequences. For each strand of the DCNN network, it took 48 hours to process the derived images with an average accuracy rate of 92.1% when classifying 8 different echocardiographic views.

In a similar study, the view identification pipeline was designed to aid automatic interpretation of echocardiograms. The state-of-the-art architecture (VGGNet) was implemented as its backbone (Deo et al., 2017). Although six different echocardiographic views were included in the study, the class label for each video was assigned by taking the majority decision of predicted view labels on the 10 frames extracted from the video. Using the confusion matrix, their model accuracy was 97.7%, and no results for single image classification were reported. However, in another follow up study, (Zhang *et al.*, 2018), they included 23 views (9 of which were 3 apical planes, each one divided into 'no occlusions', 'occluded left atrium (LA)', and left ventricle (LV)' categories) from 277 echocardiograms. Unfortunately, the reported overall accuracy of the model dropped by 14.5% at an individual image level.

Another research work proposed a DCNN model to classify 12 standard B-mode echocardiographic views (15 views, including Doppler modalities) using a dataset of 267 transthoracic studies (90% used for training-validation, and 10% for testing) achieved overall accuracy of 91.7% for classifying single frames (Madani, A., et al., 2018), but later, reported an improved classification accuracy of 93.64% (Ronneberger, O., et al., 2015).

Also in a recent study, where a light-weight model was employed (Vaseli et al., 2019), three state-of-the-art networks (VGG16, ResNet and MobileNet) were ensembled for classification of 12 echocardiographic views. Effective accuracy recorded was 88.1% with a minimum inference time of 52ms for echocardiographic images of 80×80 pixels in spatial size. Although, the reported accuracies were computed from the average predictions for all constituent frames in each cine loop. There was no specific indication of the number of frames present in the cine loop. Therefore, further critical evaluation of their models may be necessary and as apparent from the provided confusion matrices, a great majority of the reported misclassifications, seen as a failure of the models, occurred for parasternal short-axis views.

In a more recent study by (Azarmehr *et al.*, 2021), NAS-based method (DARTS) was implemented on 14 apical standard views. The method uses normal and reduction cells and evaluating several spatial input sizes with best image size of 128 x 128 x 3 pixels took 96hours of GPU search. The best architecture of 2-cell-DARTS with 0.5million trainable parameters became the derived model that yielded prediction accuracy of 96% while the evaluation of model with more than 2-cell produced an increased number of trainable parameters that consequently account for redundancy and increased inference time. Nevertheless, the implemented DARTs method did not account for classification that combined temporal space features (Mesmakhosroshahi and Kim, 2012) which is significant and synonymous with 2D echocardiograms. Processing cardiac specimens in video rather than single image in label space, consist of essential temporal and spatial features which, unless a model accounts for, could inhibit model's final performance (Ren *et al.*, 2021).

4.3 Main Contributions

The application of the NAS-based method for Spatio-temporal feature in cardiac specimen classification is beginning to attract substantial attention. Although this is common in complex action recognition in videos (Ullah *et al.,* 2018), two-dimensional echocardiograms present far more complexity in computation overhead and recognition of essential anatomic and physiological features. Since transthoracic examination workflow majorly operates b-mode cine loop also known as collection of sequential frames in time, they present significant spatial and temporal interest of which any inaccurate representation on echocardiogram specimen would spell undesirable clinical scenario.

However, given the primary objectives of minimising the neural network size and maximising prediction accuracy of a space-time convolution architecture search, this study aims at implementing hybrid-architectures method which assess the performance of the new semiNAS-derived architecture and the classical DCNN with LSTM model for spatio-temporal feature extraction and predictions of fourteen (14) different apical view echocardiography standards. Finally makes comparison with 2Cell-DART architecture (Azarmehr, N., et al., 2020), discussion on translatory advantages and system on chip (SoC) implementations.

Although some NAS methods have reduced the number of GPU search days required to single digits in the image field, directly using 3D convolution to extend NAS to the video field is still likely to produce a surge in computing volume (Ren, P., et al., 2021). Therefore, the choice of spatio-temporal DARTs method would implement 2D convolution instead of 3D which is popular with high-fidelity video action recognition.

This study used a private dataset (PACS) to design customised network architectures for the task of cardiac view classification with inclusion of Spatio-temporal elements. Unlike most previous studies, where a specific selection of an image resolution sample was not explained, the impact of different input image resolutions on the performance of the model has been investigated. Hence, the accuracy of deep learning classifiers is collectively dependent on the size of high-quality training datasets and Spatio-temporal features (in the case of echocardiogram's cine loop specimen). Furthermore, this research presents the findings on the influence that the size of training data wields on model's performance for each of the networks examined.

No matter how ingenious the deep learning model, image quality places a ceiling on the reliability of any automated image analysis. Echocardiograms inherently suffer from relatively poor image quality, which in effect, impact classification performance. To the best of author's knowledge, no other study has extended the DARTs or semiNAS method to spatio-temporal features of two-dimensional echocardiographic frame classification. Considering the above points, the main contributions of this chapter are summarised as follows:

- Inclusion of 14 different anatomical echocardiographic views (outlined in section 4.4 Figure 4.1); larger than any previous study. Additional cases when only 7 or 5 different views were included to investigate the impact of the number of views on the detection accuracy were also investigated.
- Analysis of three well-known network topologies and of a proposed neural network, obtained from applying manually derived techniques to design network topologies with far fewer trainable parameters and better accuracy for Spatio-temporal view classification.
- Analysis of computational and accuracy performance of the developed models using our large-scale test dataset.
- Analysis of the impact of the input image resolution; 4 different image sizes were investigated.
- Analysis of the influence of the size of training data on the model's performance for all investigated networks.
- Analysis of the correlation between the image quality and accuracy of the model for view detection.
- Analysis of ultra-lightweight model architecture in Spatio-temporal classification and the derived NAS-based classification model.

4.4 Methodology

In this section, the introduction of private dataset PACS used for apical frame classification are presented, followed by the description of the expert annotation process. This research investigates the performance of an optimised lightweight architecture (CardioQNet and 2Cell-DARTS) described in chapter three section 3.9.1 and three state-of-the-art models (VggNet16, ResNet50, MobileNet v2), for echocardiography frame classification. Details of the neural network model, training parameters, and evaluation metrics are then provided.

4.4.1 Private Dataset Source (PACS)

This section introduces PACS private dataset used for the 2D echocardiographic view classification in this thesis. A random sample of 374 echocardiographic examinations of different patients and performed between 2010 and 2020 was extracted from Imperial College Healthcare NHS Trust's echocardiogram database. The acquisition of the images was performed by experienced echocardiographers and according to standard protocols, using ultrasound equipment from GE and Philips manufacturers. Ethical approval was obtained from the Health Regulatory Agency (Integrated Research Application System identifier 243023). Only studies with full patient demographic data and without intravenous contrast administration were included. Automated anonymisation was performed to remove all patient-identifiable information. The videos were annotated manually by an expert cardiologist (JPH), categorising each video into one of 14 classes which are outlined in Figure 4.1.

Videos thought to show no identifiable echocardiographic features, or which depicted more than one view, were excluded. Altogether, this resulted in 9,098 echocardiographic videos. Of these, 8,732 (96.0%) videos could be classified as one of the 14 views by the human expert. The remaining 366 videos were not classifiable as a single view, either because the view changed during the video loop, or because the images were completely unrecognisable. The cardiologist's annotations of the videos were used as the ground truth for all constituent frames of that video. DICOM-formatted videos were then split into constituent frames. From each video, three frames were randomly selected to account for end-systole and end-diastole frames, representing arbitrary stages of the heart cycle, resulting in 41,321 images. The study had earlier experimented with 15, 20, 25 and found that 3 frames did not alter the performance of the model in the overall assessment. This could be due fair representation of sample space in the dataset.

The dataset was then randomly split into training (24791 images), validation (8265 images), and testing (8265 images) sub-datasets in a 60:20:20 ratio. Each sub-datasets contained frames from separate echo studies to maintain sample independence. The relative distribution of echo view classes labelled by the expert cardiologist is displayed in Figure 4.2 and indicates an imbalanced dataset, with a ratio of 3% (Subcostal-IVC view as the least represented class) to 13% (PSAX-AV view as the dominant one).

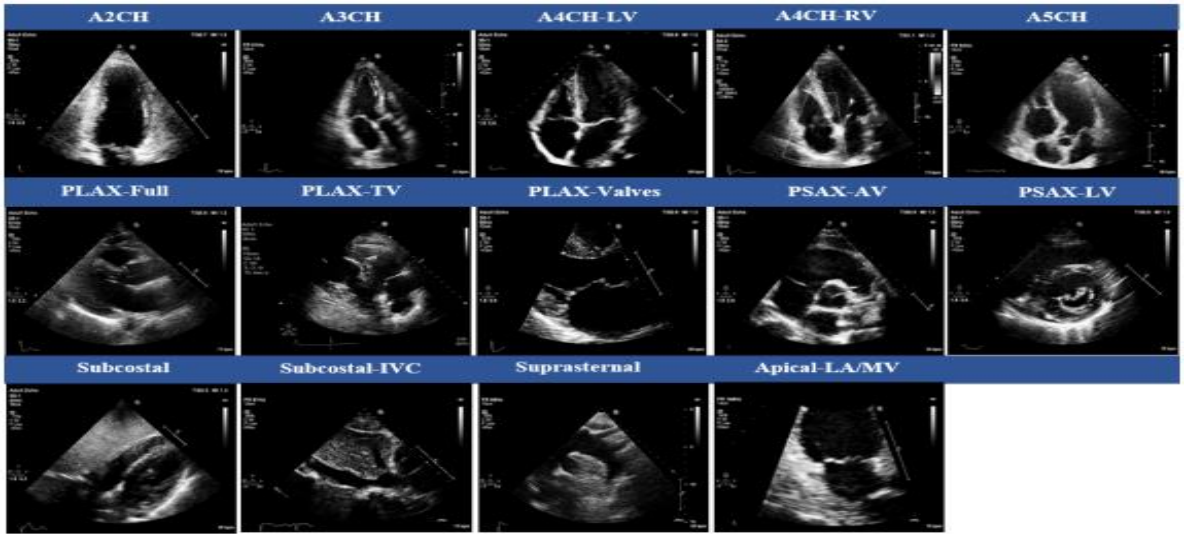


Figure 4.1: Samples of 14 cardiac views used in this study: apical two chamber (A2C), apical three-chamber (A3CH), apical four-chamber left ventricle focused (A4C-LV), apical four-chamber right ventricle focused (A4C-RV), apical five-chamber (A5C), parasternal long-axis (PLAX-Full), parasternal long-axis tricuspid valve focused (PLAX-TV), parasternal long-axis valves focused (PLAX-Valves), parasternal short-axis aortic valve focused (PSAX-AV), parasternal short-axis left ventricle focused (PSAX-LV), subcostal (Subcostal), subcostal view of the inferior vena cava (Subcostal-IVC), suprasternal (Suprasternal), and apical left atrium mitral valve focused (LA-MV).

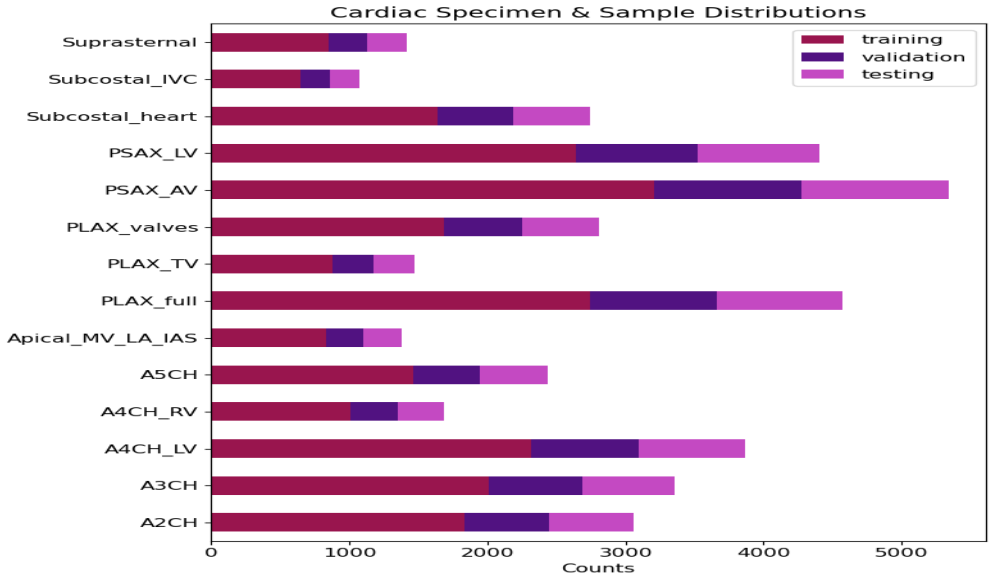


Figure 4.2: Distribution of data used in apical view classification showing class imbalanced on training, validation, and test dataset; values shown represents the number of frames per given class

4.4.2 Ultra-lightweight Architecture Model

Echocardiogram's data comes in a sequence of images known as frames arranged sequentially in timesteps, of 1 to 2 seconds frame length. Although image classification models have shown great potential in medical imaging (Luong *et al.*, 2021), however, the application of normal image classification in echocardiography is limited to 2D echocardiography. This is because classical image classification could yield a misleading prediction especially on cardiac frames due to the inherent temporal features usually associated with 2D echocardiograms. Therefore, the classification task of predicting echocardiographic views is insufficient for clinical advantage in the real world. To take advantage of souring performance of deep convolutional neural models, the appropriate data for view classification should include images taken in timesteps rather than random snapshot of cardiac views.

Time series classification model hence becomes the essential model used for accurate Spatio-temporal classification and can increase model confidence on prediction, provide prediction stability of a fast-moving echocardiogram frame (Ullah *et al.*, 2018).

The challenge to post implementation integration engineers however, remains mostly in hardware configuration, scalability, storage, speed of operations and memory requirement. Majority of the state-of-the-art models each have its memory requirement to function on mobile or desktop applications. Since automatic neural architecture search is about finding the best possible model architecture to efficiently handle classification tasks, these models are highly personalised and may perform poorly on a different dataset compared to the dataset originally trained or designed for. Besides, each model with different characteristics in terms of trainable parameters, storage requirements, memory requirement and inference speed all should be carefully considered for medical imaging application.

There are several AutoML methods in the automatic search for neural architecture and use cases are best to justify the choice of any NAS (Siems *et al.*, 2020). As enumerated in section 3.3.1, the justification for the semi-NAS method for lightweight architecture is to replicate a physical life-event scenario where deployment of a specific solution is urgently required and where a few hours of delay can spell clinical disaster. Such a scenario was seen to have played out in the recent Covid-19 Pandemic. Since automatic neural search could take days and weeks of GPU's time to conclude on the search for suitable architecture models, a semi-automatic approach, where a model is designed manually, or follow a structural but improved pattern of an existing model could become relevant and save the day. The advantage of quick

development, meeting stringent hardware integration requirements, and memory limitations could become significant and huge differences in clinical practice and real-life situations.

Partly inspired by the work of (Abdi, Luong, Tsang, Jue, *et al.*, 2017a) light-weight model, a further reduction in model learnable parameters and memory requirement was investigated and implemented using semi-NAS method described in chapter 3. The new ultra-lightweight architecture was developed using a combination of manual methods in excel worksheet, and automated optimization of learning parameters. The derived model is herein referred to as 'CardioQNet', which is a semiNAS-based derived regression hybrid model. This is divided into two logical blocks, deep convolutional neural network DCNN block for spatial feature extraction and LSTM classifier block to elicit temporal features in the cardiac frame sequences. The feature extraction process allows weight sharing across convolutional layers while extracting the hierarchical feature in the image sequence. Instead of feeding the model with a random, single frame per time which is synonymous with classical CNN classification, a batch sequence of frames determined by sequence length is fed into semiNAS derived, 3-layer CNN architecture which accepts 2D cardiac frames of spatial size 224 x 224 x 3 pixels. The frame sequence convolved with the feature extractor and a resultant feature map are fed with timestamp into the recurrent layer using Rectifier Linear Units (ReLUs) activation function. The recurrent layer consists of a Long Short-Term Memory (LSTM) classification model. The purpose of this last layer is to extract temporal features for the sequence of cardiac images as it slides through LSTM windows for Spatio-temporal predictions of average weighted probabilities. The model was trained on PACS pre annotated dataset using cross entropy cost function and adaptive moment estimation (ADAM) as weight's optimizer. Since LSTM uses sigmoid for its internal gating, a fully connected layer is added to act as a classifier layer with SoftMax activation function. The architecture is depicted in Figure 4.3.

.

4.4.3 Model Training

The architecture listed in Table 4.1, include top-3 award-winning classification models namely: VGG16Net, ResNet, and MobileNet. Each of the model was configured in a CNN-LSTM hybrid fashion to extract low and high temporal features from echocardiographic frames. The output feature map of the ConvNet feature extraction is flattened and fed into LSTM layer input to extract the time-dependent temporal features with the final probabilities vector. Each of the models was then trained on PACS dataset including the newly semiNAS derived hybrid model,

'CardioQNet' illustrated in Figure 4.3. The consideration for selection of the state-of-the-art model was based on the number of trainable parameters, least memory requirement for forward and back propagations, and performance history in medical imaging classification.

Image Resolution: The model's input has a spatial size of [224 x 224 x 3] pixels and is trained using a 5-fold cross validation technique to ensure adequate learning on the dataset and performance was recorded for each model. Since PACS dataset is slightly imbalanced, an online over sampling strategy was implemented to prevent model bias toward the minority class while early stopping was implemented to reduce potential overfitting.

Hyper-parameter optimization: The hyper parameters learning rate was set at 0.002 with high momentum 0.95 and decay rate of 0.1every 25 steps and were reproducibly initialised to minimise possible deviation in score performance. Optimization of network parameters was done via stochastic gradient descent (SGD) with cross entropy loss function at each run. Training was initialised and completed as learning curves converged around 50 epochs.

Batch Selection: To ensure that each batch has a decent chance of containing each class samples, a batch size of 128 was chosen but at the expense of computational overhead cost which ran high and limiting operational memory. Consequently, a batch selection of 64 and 32 were experimented, memory utilisation becomes significantly apparent at batch selection of 64 at a fixed length sequence of 3 than running a batch size of 32 at the same fixed length sequence. Hardware performance difference of 0.18% in terms of computational speed was a negligible trade-off that did not affect model's ability to properly generalise new test samples.

Data Augmentation: Augmentation was applied to allow optimum learning sequences for the models; a maximum translation of [-0.05, +0.05] pixels and maximum rotation of 10 degrees were applied randomly for horizontal, vertical, and rotational angles respectively. To prevent overfitting in the training phase, we applied batch normalisation at each convolution layer, early stopping and dropout (rate 0.30) for the training samples. Batch normalisation also helps stabilise and speeds up convergence during the training phase.

Model training was implemented using PyTorch. The experiment was carried out on GPU GeForce GTX 970 chipset's Maxwell architecture, featuring 4GB RAM coupled to 1,664 CUDA cores. Figure 4.1 shows the performance evaluation by confusion matrix.

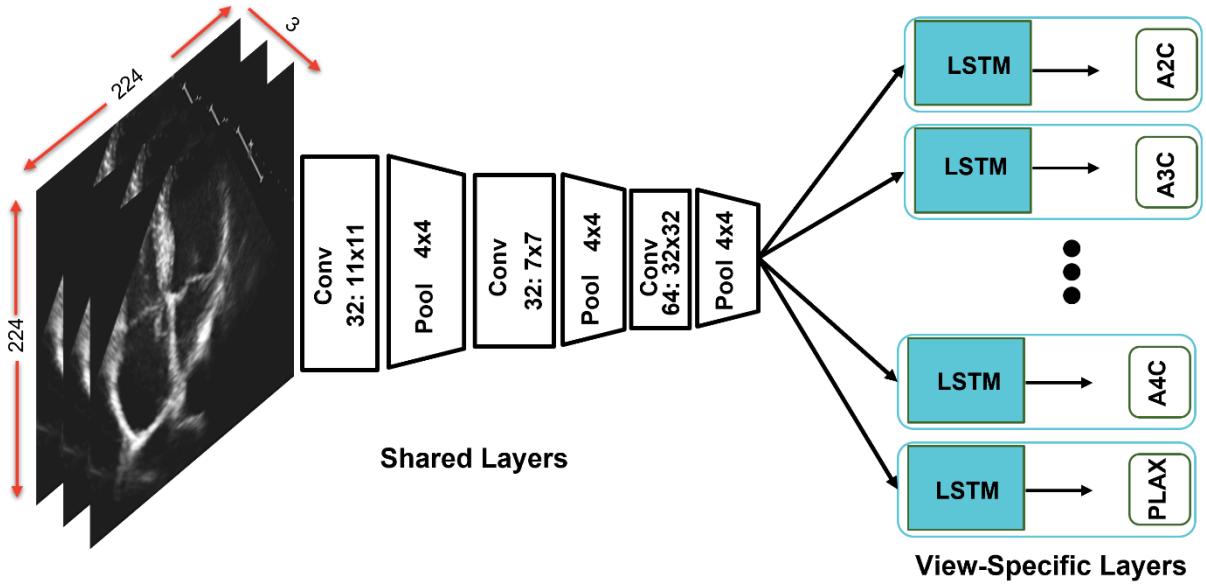


Figure 4.3: The lightweight hybrid architecture in a multi-stream configuration for echocardiography frame classification. CardioQNet is used for weight sharing feeding 14 view-specific layers in the LSTM. Numbers of kernels in each layer and their corresponding sizes are shown accordingly.

4.5 Evaluation Metrics

The model performance was evaluated using most popular classification metrics as enumerated as follows:

- **Model Accuracy:** The model performance was evaluated as the number of correctly classified samples over the total number of all labelled samples, computed as

$$Acc_{\mu} = \frac{\sum_{i=1}^{k} \frac{tp_i + tn_i}{tp_i + tn_i + fp_i + fn_i}}{k} \qquad (4.1)$$

- **Recall Macro Average (RecM):** Recall is the fraction of instances of correctly predicted class *(tp)*, and in the binary classification problem recall can be computed by the equation 4.2 where *tp* represents the model correctly predicts the positive class, and *fn* (false negative) is where the model incorrectly predicts the negative class. But for multi-class cases, the macro-averaged recall will be computed by an equation 4.3 which calculates metrics for each class *(k)* and finds the mean.

$$Recall = \frac{\sum_{i=1}^{k} tp_i}{\sum_{i=1}^{k} tp_i + fn_i} \quad (4.2)$$

$$Recall_{\mu} = \frac{Recall}{k} \quad (4.3)$$

- **Precision Macro Average:** This metric is defined as the fraction of correct predictions for a certain class *(tp),* and in the binary classification case precision can be calculated by the equation 4.4 where *tp* (true positive) represents the model's correctly predicted test samples, and *fp* (false positive) is the total number of prediction errors. Precision can be extended to capture class imbalance. However, for multi-class cases, macro-averaged precision is computed in equation 4.5 which gives equal weight to each class and averages the overall performance.

$$Precision = \frac{\sum_{i=1}^{k} tp_i}{\sum_{i=1}^{k} tp_i + fp_i} \quad (4.4)$$

$$Precision_{\mu} = \frac{Precision}{k} \quad (4.5)$$

- **Confusion Matrix:** Confusion matrix is a two-dimensional graphical matrix that presents a performance overview of the classifier on a test dataset. This provides an insight into the error being made by a classifier and more importantly the types and location of errors that were made in the label space. In its first dimension, list the true classes of the test dataset, and in the other dimension, the prediction results assigned by the classifier. Effectively, the confusion matrix displays the number of correct and incorrect predictions broken down by each class in label space (Grandini, Bagli and Visani, 2020).

- **Number of trainable parameters:** There are two ways in which model reduction is beneficial in deep learning. In the first case, after the model weights are obtained via training, model reduction can be used to reduce the resources needed for inference, (Wu, 2019). There exists larger redundancy with increased number of trainable parameters in the model and so does the inference time for view classification. Therefore, the models with larger trainable parameters do require larger storage and memory space hence the

reduced number of trainable parameters afford integration benefits of faster convergence, shorter training times and inference time.

- **Inference latency:** Latency is a measure of successful round-trip time for some data across the processing medium, in this case of inference, it's a measure of time a model yields a prediction of view classes per image. This is measured on a test dataset, and it is dependent on the combination of the central processing unit (CPU), graphic processing unit (GPU) (Khan *et al.*, 2019) and number of O(n) complexity in the algorithm. Inference time is therefore computed on the number of image processing in a loop.

4.6 Results and Discussion

The ultra-lightweight semiNAS derived model achieved an average overall test accuracy of **91.25%** on sequence frame classification, without overfitting and test accuracy of **87%** on image classification with minimum convergence time. Figure 4.4, showing the confusion matrices showing actual view labels on *y*-axis, and neural network-predicted view labels on the *x*-axis by view category for (a) frame classification (b) image classification on semiNAS derived model. There is a clearly positive correlation (**4.89%)** between model performance accuracy and frame classification, establishing the hypothesis for video or frame classification task preferred to image classification task in 2D echocardiography. Although, the ultra-lightweight model classification performance lags behind ResNet50, the best performing state-of-the-art model by **0.02**%, and **5.4**% in comparison to 2Cell-DARTS models, nevertheless, it achieved an interesting performance and computational potentials for on-the-go true lightweight model in classification tasks. The CardioQNet model clearly outperformed 2Cell-DARTS model in terms of inference time, even though 2Cell-DARTS' memory requirements were not assessed.

Moreover, Table 4.1 detailed the comparison of accuracy by view category for model's frame classification performance, total number of trainable parameters, memory storage requirements, and varying training time exhibited by the chosen model for echocardiographic specimen. Memory wise, CardioQNet achieved a reduction of **93.72%** in comparison to the highest

memory demand of Resnet50 which required 7.01 gigabyte memory for 23million trainable params, making CardioQNet a possible candidate for mobile integration and point of care echocardiography. On the other hand, convergence and training time was high for VGGNet with 14.06 minutes per epoch time, while CardioQNet achieved a reduction by **54%** in epoch time, hence in convergence and inference time.

On a single frame drawn from all 14 views, the model achieved an average overall accuracy of **91.25%** (Table 4.1), compared to an average of **87.0%** on classical image classification. Accuracy was decently consistent across the views except with apical four-chamber with apical three chamber A3C, PSAX_LV, PSAX_AV, PLAX_TV as a view that is less visually distinct from A2CH in Figure 4.6 while Figure 4.5 provides the visualisation of example filters learned on PACS Dataset.

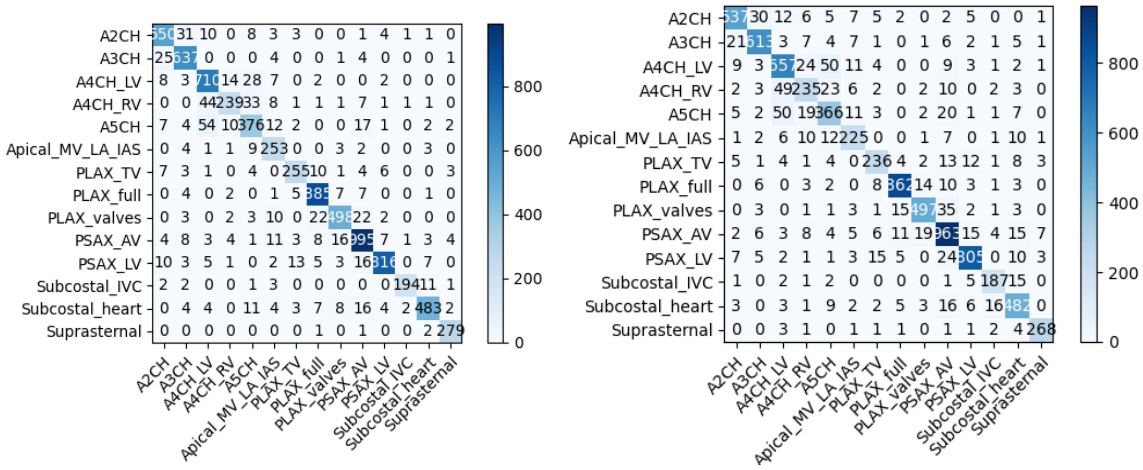


Figure 4.4: Confusion matrices for semiNAS derived model with input resolution of 224 x 224 pixels. The figure showing actual view labels on y-axis and predicted view labels on the x-axis by view category for (Left) frame classification (Right) image classification on CardioQNet.

Table 4.1: Comparison of Model's design properties and performance. The values in bold indicate the best performance for each model. CPU Inference carried out on Intel Xenon® CPU X5650 @ 2.67GHz.*

Evaluated Models @ batch size 24	# of trainable parameters (million)	Memory requirements (GB)	Training Time(s) per epoch	CPU* Inference Time (ms)	GPU Inference Time (ms)	Model Accuracy (%)
VggNet 16	17.93	5.35	844	394.19	10.60	89.25
ResNet 50	23.80	7.01	795	194.70	8.71	92.62
MobileNet V2	12.71	4.58	443	81.00	6.69	92.52
2Cell-DARTS	0.55	-	380	-	11.8	96.00
CardioQNet	**2.69**	**0.44**	**392**	**83.79**	**6.44**	**91.25**

On inference time, 2Cell-DARTS model was evaluated on GPU inference which shows little competitive advantage compared to the rest of the models. CardioQNet proved to be effective both in GPU and CPU inference time.

This is a significant advantage in evaluating and contrasting model performances. Even the least thoughtful model can exceed performance expectation as shown in this case of CardioNet.

Table 4.2: Details of class performance accuracy for 14 classes of selected apical standard classification.

Class Performance	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Spatial Input Size	**[224 x 224]**			
A2CH	90.3	90.0	90.0	90.0
A3CH	90.4	90.0	95.0	92.0
A4CH_LV	90.3	85.0	92.0	88.0
A4CH_RV	89.4	88.0	71.0	78.0
A5CH	90.4	79.0	77.0	78.0
Apical_MV_LA_IAS	90.3	80.0	92.0	85.0
PLAX_TV	90.4	89.0	87.0	88.0
PLAX_full	90.4	94.0	97.0	96.0
PLAX_valves	90.3	93.0	89.0	91.0
PSAX_AV	90.3	91.0	93.0	92.0
PSAX_LV	90.4	97.0	93.0	95.0
Subcostal_IVC	90.3	97.0	91.0	94.0
Subcostal_heart	90.3	94.0	88.0	91.0
Suprasternal	90.4	96.0	99.0	97.0

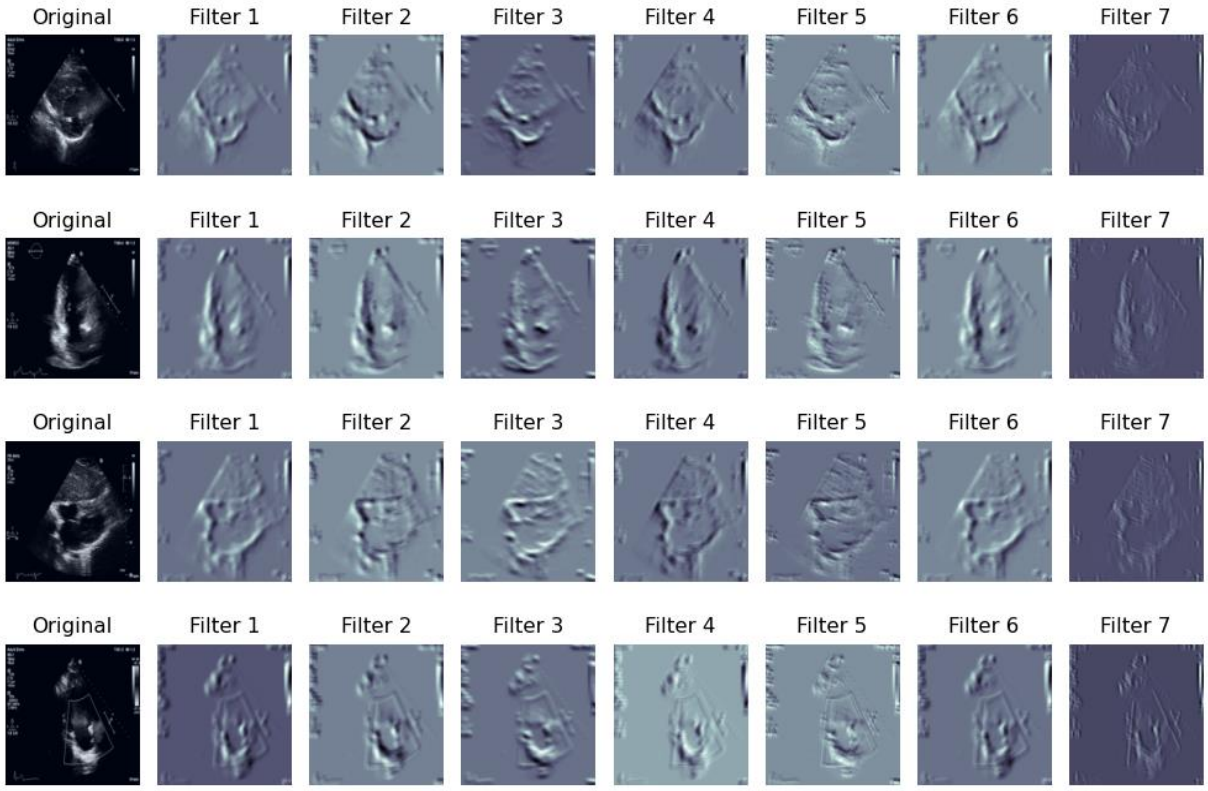


Figure 4.5: Visualisation of some learning filters for 2D echocardiographic example on PACS Dataset. Learning filters provide insight to what features are detectable in the given specimens.

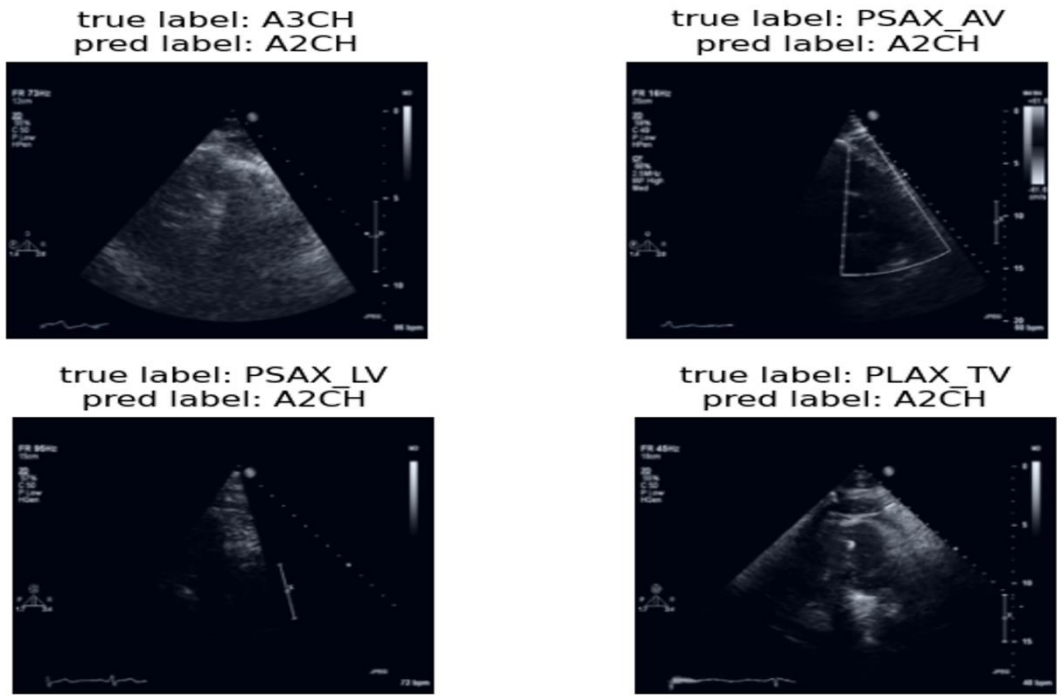


Figure 4.6: Showing four misclassified samples prediction by semiNAS derived model (CardioQNet) for spatial input resolution of 224 x 224 pixels disagree with expert annotation.

4.7 Conclusion

In this chapter, efficient CNN architectures and ultra-lightweight (2 Cell-DARTS, CardioQNet) performance were proposed for the automated identification of the 2D echocardiographic frame. Both DARTS and semi-automatic NAS methods were considered in designing optimised architectures for rapid inference while chosen state-of-the-art models were assessed for memory suitability and performance accuracy towards integration of echocardiography pipeline.

Compared with the standard classification DCNN architectures, the proposed models are faster in terms of inference, with 93.00% reduction in memory usage but achieved 5.4% less comparable in classification performance to DARTS and 2.5% comparable in classification performance to ResNet50, MobileNetV2 and VggNet models.

Such ultra-lightweight models can thus be used for real-time detection of the standard echo views. Similar contributions on the 2Cell-DARTS model provide detailed impact of image quality and input image spatial size with a conclusion for larger input size resolution yielding better accuracy performance. Neural network models with many redundant trainable

parameters, require more training data and extended training time to achieve similar performances. On the 2Cell-DARTS model, a direct correlation between the image quality and classification accuracy was observed while frame classification yielded a significant performance accuracy and best inference time in the CardioQNet model.

Furthermore, it appears that model performance decreases when the class space is scaled up and increases with fewer class space. More class category presents additional complexity and computational overhead in distinguishing cardiac views. Therefore, it can be concluded that the extent of class space in frame and image classification tasks have a direct impact on the performance of the deep learning models. This is true for all deep learning models and a point of realistic consideration for a fair comparison of classification models.

For 2D echocardiographic images as input to a model require a minimum of 128 x 128 spatial size as established in DARTS model, therefore, lower spatial input size could result in inherent loss of relevant anatomical and pathological features, a severe consequence in clinical workflow and model accuracy. On the other hand, larger images are thus required for spatio-temporal frame classification and other fine-grained applications such feature segmentation, or cardiac global features. In this case, 2 Cell-DARTS will fail to compare favourably in terms of inference speed and memory requirement and perhaps model accuracy.

In echocardiograms assessment, inference time is considered a critical factor. Firstly, to provide real-time score visualization and updates as frames advanced in cardiac cycles, prediction algorithm must be able to keep up with standard rate of frame acceleration otherwise known as frames per seconds (FPS). Secondly, to prevent latency or drag on predictions performance when deployed in remote or during clinical emergencies. The best candidate on inference time is CardioQNet model with GPU-6.44ms, and CPU-83.79ms. This raises the question of technical morality associated with technological trends, obsolete process, and exhaustive view of automated network architecture search (NAS) and traditional manual approach of DCNN design methods. It's clear that model accuracy of CardioQNet ranked 4th out of five models evaluated, it outperformed VggNet16 model. This impressive and objective record of a semi-automated derived model indicates the possibility for achieving a state-of-the-art performance on echocardiogram frame classification when AutoML NAS based model is not part of the options.

Chapter 5

Image Quality & Assessment Methods

5.1 Introduction

A two-dimension (2D) echocardiogram has become the de-facto standard of assessing cardiac functions because it presents rich anatomical details of the myocardium. Also, for its non-ionizing in-vivo advantages have found its significance in antenatal, obstetric, and general diagnosis of myocardial infarction. Nevertheless, images produced though scattering centres do not come with crisp edges unlike natural images. This inherent limitation in echo image resolutions poses a challenge to clinical measurement and interpretation of image features, the reason it is solely considered suitable for experts. Therefore, a quality image acquisition is a priced clinical enterprise and mostly required for expert's assessment, quantification of cardiac functions and diagnosis of myocardial infarction.

Acquisition Problem: Acquisition of echocardiographic images requires significant experts' skill which vary with experience and patients' pathological profiles. In clinical transthoracic examination workflow, strong indication for quantification of systolic function in apical four (A4C) and parasternal long axis (PLAX) view is a recommended standard practice (Thomas *et al.*, 2005). According to Lang et al (2015), these are recommended quantification standards in clinical practice because their spatial orientations are congruent in nature, both, thus offering complementary advantages on heart functional measurement (Lang *et al.*, 2015). Nonetheless, the clinical process of image acquisition comes with inherent challenges of operator skill/experience which cascade many successive analytical problems including diagnostic reliability output. These drawbacks remain significant, consequently inhibiting the adoption of echocardiograms as a reliable imaging modality for cardiac diagnosis despites its many advantages.

***Definition of Image Quality*:** Specifying echocardiography image quality and methods of its assessment is not a trivial issue because image quality is critical to clinical measurement and accurate diagnosis (Liao *et al.*, 2019). As the in vivo examination of heart structures gained prominence in cardiac diagnosis, it has been affirmed that accurate diagnosis of the left ventricle functions is hugely dependent on the quality of echo images. Up till now, visual assessment of echo images is highly subjective and what constitutes the element of two-dimensional image quality is largely unknown (Sassaroli *et al.*, 2019) hence, the requirement for a domain-specific definition under clinical pathologies. While poor-quality images impair quantifications and diagnosis, the inherent variations in echocardiographic image quality standards indicates the complexity faced among different observers and provides apparent evidence for incoherent assessment under clinical trials, especially with less experienced cardiologists.

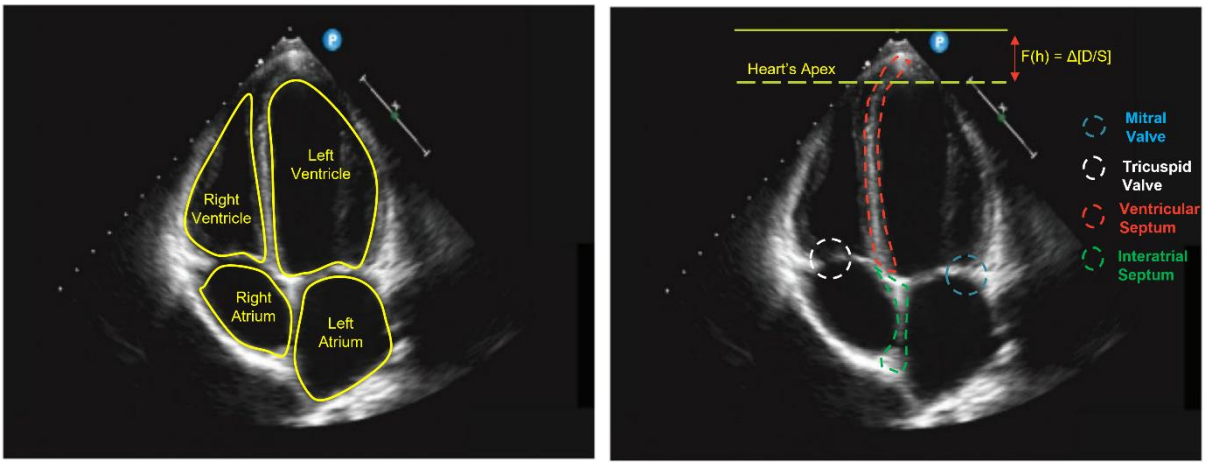


Figure 5.1: Illustrating tissues identification of Apical-four quality images. The ideal perspective of A4C and the position of chamber cavities, valves, and interatrial septum in perspective of clinician. Images show clear delineation of cavities for linear measurement and quantifications.

Figure 5.1 are two separate echocardiogram showing apical-four chamber and parasternal long axis views. While the relevant anatomical cavities and tissues have been correctly delineated in yellow, it not always the case for other images of the same views. This is the case of clarity, visibility, and accessibility. The significance of this relates to clinical measurement and quantification of cardiac functions. For example, to measure A4C's left ventricle (LV) cavity, say in the diagnosis of cardiac myopathy where cardiologist needs to evaluate the abnormal changes of wall or cavities, a linear measurement of LV is one of the required clinical methods. but when the cavity edges become blurry in a poor-quality image as illustrated in Figure 5.2, measurement then becomes a guess work, and this results in misdiagnosis.

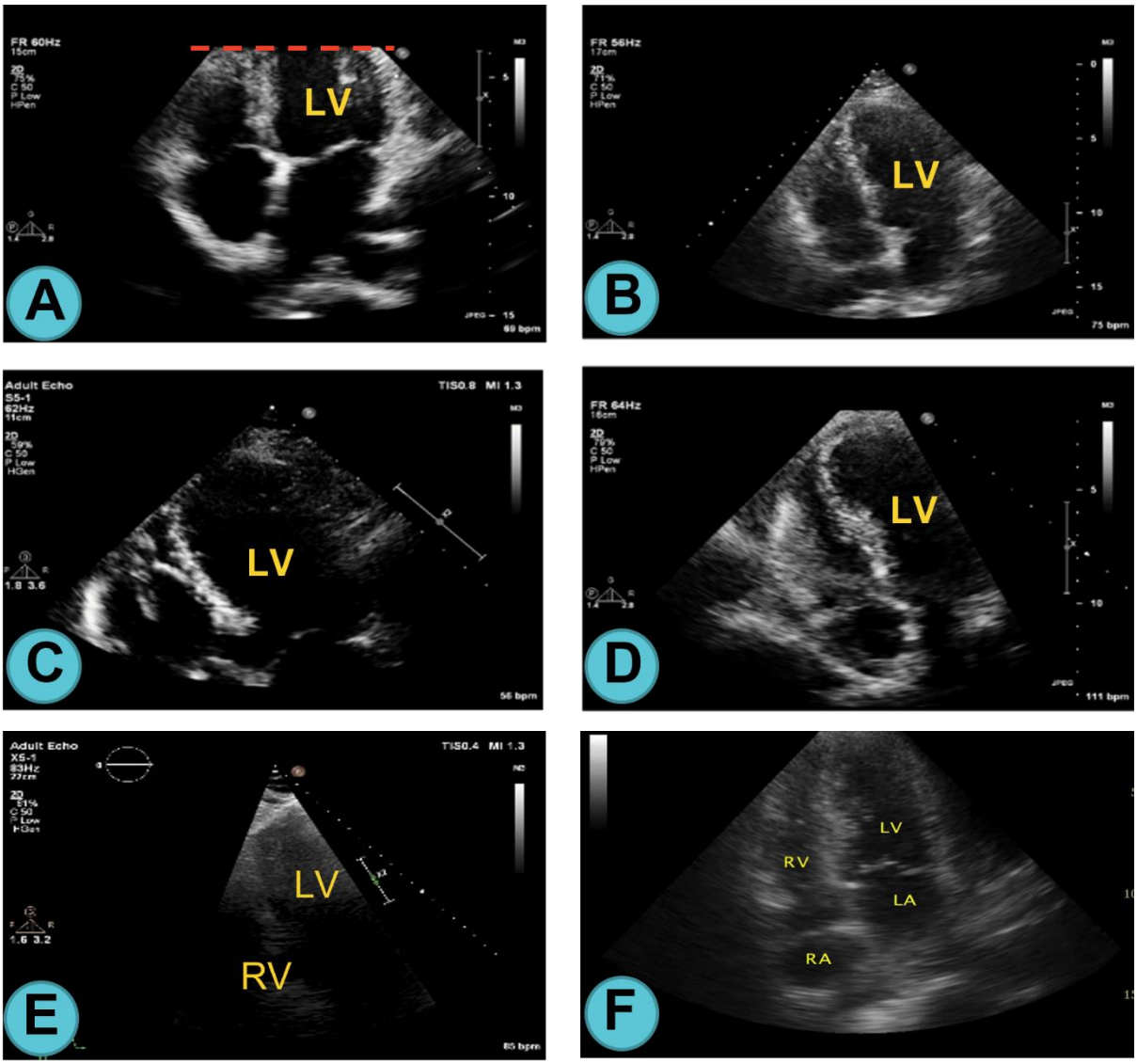


Figure 5.2: Typical levels of quality samples in A4C images. (A) Poor quality image due to no visibility on image's apex, may raise pathological concern. (B) Poor quality image due to low chamber clarity, linear measurement is estimated can cause misdiagnosis. (C) Poor quality due to projection and missing interatrial septum leading to incorrect LV volume depiction. (D) Poor quality due to LV significant off-axis and foreshortedness, would yield wrong volume and wall measurement. (E) Poor quality due to blurry cavities, wall, or poor probe selection. (F) Poor quality due to indistinguishable cavities, this quality possesses a challenge to linear measurement and quantification of myocardial function.

Real-Time Assessment Methods: The existing method of image quality assessment is a subjective manual process, where an echocardiography specialist visually inspects the images and decides on what anatomical features present in the image to be pathologically and clinically relevant. This process is laced with a spread spectrum of opinion and decision variability during assessment of multiple specimens (Nagata *et al.*, 2018) and when an image is retrospectively reassessed by the same operator. These variabilities and uncertainties are found to impair

quantification accuracy of cardiac functions, diagnosis, and the overall quality of patient care. Moreover, subjective processes are synonymous with management flaws and cannot be implemented in a real-time automated system. The combined effects of these flaws constitute a greater risk to life saving intentions of the only non-ionized imaging system. Many research efforts to move assessment method into automatic method have not yielded any translatory advantages because, the assessment proposed in many literatures are based on weighted average assessment which majorly applied in retrospect and would not allow image's significant optimization during acquisition.

5.2 Related Work

Prior to the use of deep convolutional models, several researchers proposed a series of methods for quantifying echocardiographic image quality and assessment methods which were enhanced by the advent of deep convolutional neural networks. Image quality assessment is generally approached by defining a reference image and calculating the deviation of any given image to this reference (Wang *et al.*, 2004). However, in echocardiography, this method is not practical, since images vary significantly from patient to patient, and it is difficult to define an image with perfect quality. Therefore, it is necessary to develop a blind image quality assessment algorithm, which does not depend on a reference image. Studies have been carried out on blind image quality assessment (Zhang *et al.*, 2016); (Wang and Bovik, 2002) largely focusing on the distortion of images due to compression, with some implementing machine learning algorithms using edge sharpness and random/structural noise level to evaluate image quality (Nafchi and Cheriet, 2018). This approach is difficult to apply to echocardiography because cardiac ultrasound does not present well defined edges due to two facts: 2D cardiac images are formed by interference pattern of scattering centres presenting an inherent poor resolution; and anatomical features do not present crisp edges because the endocardium is trabeculated there are papillary muscles, the external purkinje network. Also, epicardium does not present as a crisp boundary, either because it is joined to the myocardium on one side, and to the pericardium in several layers, including pericardial fluid, on the other. So there exist relatively subtle acoustic impedance transitions next to larger ones (Labs *et al.*, 2020). Hence, new measures of image quality need to be developed and tested based on the global properties of the Echocardiographic images. Meanwhile, the recent success of deep Convolutional Neural

Networks (DCNN) in computer vision tasks (Ungru *et al.*, 2014), there have been few reports on the application of deep learning for echocardiographic image quality assessment.

One of the earliest works on objective assessment of cardiac image quality is Abdi et al (2017). He demonstrated the feasibility of quality assessment using convolutional neural network model in five apical views using six (6) quality criteria scoring methods (Abdi, Luong, Tsang, Jue, *et al.*, 2017a). Since there was no publicly available cardiac dataset to model, the author relied on expert's knowledge for its feature engineering, a high resource intensive process. Abdi et al research yielded 85% model accuracy, plausible outcome it was but quality criteria are limited and clinically insufficient for transthoracic standard examination practice. The second reason being that the defined quality features are assessed in retrospect and do not represent experts' global characteristics for cardia specimens in 2D echocardiographic real-time frames. The authors admitted that algorithm based on cardiac view detection of cardiac chambers in apical A4C echocardiography did not guarantee good performance in clinical standard of healthcare because medical images in 2D consist significant noise and are plagued with features of low amplitude signals (contrast) that yield significant challenges in interpreting clinical pathologies and quantifications. The approach yields a contradicting detection result on images with low contrast-gain and high contrast-gain which reinforce the conclusion on a model that combines spatial and temporal extraction to guarantee better classification accuracy. Among the most recognised work on automated quality assessment of 2D echocardiograms, are (Abdi, Luong, Tsang, Jue, *et al.*, 2017a), (Luong *et al.*, 2021), (Dong *et al.*, 2020), and (Labs, Zolgharni and Loo, 2021). Initially, Abdi presented a deep convolution algorithm for detection of echocardiogram cardiac view classification in apical four chambers (A4C). The approach successfully modelled and generalises cardiac specimens with four apical chambers (A4C) demonstrating the feasibility of neural network on 2D echocardiographic images. Although, in his follow up publication, which was based on a much better method of continuous values regression model, five apical standard views were used, quality scores were estimated based on the sequence of echo cine loops which include the end systole, end diastole to produce a single quality score per frame per view. The results were impressive with a prediction accuracy of 86% and a computation time of <1000 sec on a desktop computer. Unfortunately, Abdi's work used a weighted average of quality measures hence, the scores do not provide precise guidance to the aspect of image quality that needs to be optimised. On the other hand, Luong's et al (2021) which investigated the mechanically ventilated TTE on hospitalised patients, yielded certain improvement in performance but with a much larger dataset to represent wider

population distribution. Luong et al., (2021), defined twelve (12) criteria to grade each of the nine apical standard view, while computing a continuous single variable score to represent objective quality scores for respective apical views. Luong's regression model achieved overall accuracy of 87% with regards to four expert ground truths and sufficiently demonstrated the impact of image quality on diagnostic utility. However, Luong's method of quality assessment was similar to Abdi's in the sense that the unified model produces a single score per image view across the nine apical standards considered. Luong admitted there exist no reference quality standard for the evaluation of echocardiographic image quality (Luong *et al.*, 2021) but a scale of criteria used in many publications did not represent experts' visual assessment and consensus on 2D echocardiographic image quality. Similarly, Dong et al (2020) proposed a generic quality control framework on A4C. It considered application of image quality to fetal ultrasound to alleviate the challenges in antenatal investigation. The proposed method detailed the assessment of image quality using two features namely Gain and Zoom. It was considered as the first comprehensive quality control system but significantly lacks adequacy for generalisation of quality attributes required for wider use case. Hence, suitability for quality assessment is inherently impaired.

To the best of our knowledge, Dong et al, (Dong *et al.*, 2020) represents the most recent study on objective quality assessment, however the study was limited to apical four-chamber plane (A4C) and did not include PLAX view nor similar score criteria independently assessable in clinical practice. For this reason, assessment could be suitable for quantifying image quality for fetal ultrasounds rather than for adult patients. Dong's argument for focus/zoom attributes emanated from fetal cardiology where specific tissue became the focus of an investigation. However, these attributes, though important, should be described as elements of clarity. Therefore, a zoomed section of myocardium should exhibit the attributes of clarity, instead of being considered as an independent factor

This thesis details the novel definition of clinical standard of domain-specific quality attributes for 2D echocardiogram specimen and proposed deep convolutional neural network architecture to model four quality attributes as this separately provides the most relevant quality assessment information for operators' feedback during the image acquisition phase. This work is based on a multi-stream, multi-output regression model with selective qualitative attributes which are progressively distinguished from earlier related work (Labs, Zolgharni and Loo, 2021). The advantage of this novel method of quality assessment is that the specific component of quality can be optimised, thus potentially guaranteeing clinical real-time feedback and optimization for

obtaining optimum image quality in clinical practice. This means that during the acquisition phase, the operators can assess specific quality elements independently, as would be indicated on each four attributes rather than obtaining a weighted average of quality components which is the existing and current assessment method obtainable in the most recent research papers. To the author's knowledge, there are no published methods on attributes of quality and its assessment method in echocardiography modelling. This novel approach and quality formulations can be used to assess, optimise, and quantify echo images surgically, in real-time.

5.3 Main Contributions

Interpreting the results of the proposed architectures in the literature is not straightforward. This is because a direct comparison of the models' performance would require access to the same patient dataset. At present, no echocardiography dataset and the corresponding annotations for the image quality assessment is publicly available. We, therefore, aimed at evaluating the performance of deep learning models for the automated image quality assessment using an independent echocardiography dataset (PACS) available on request. Although the inference time reported in the previous studies reviewed was short enough the make it feasible for real-time applications, the utility of such systems in the clinical practice would be limited.

This is because only an overall predicted image quality score is provided by the models. Therefore, if employed as part of an operator guidance system, the operator is provided with no clues as to why the image is being tagged as low quality, and how to improve it to obtain optimal images. A practical quality control report should contain such information.

In the light of the above related works, we conclude that, although they represent plausible contributions, all the criteria used to define assessment are limited in scope and insufficient for translatory clinical relevance. However, this research examined all existing quality criteria and additional criteria that can easily translate to experts' subjective assessment. Finally, we defined, for the first time, a novel, most comprehensive criteria, and objective attributes by which cardiac images can be optimally assessed. The main contribution of this chapter can be thus summarised as follows:

- Demonstration of the feasibility for a novel, coherent and clinically relevant objective standards for the assessment of 2D echocardiographic image which account for relevant anatomical profiles, linear and volumetric quantifications of myocardial functions
- Fresh insight to real-time assessment method that provides access to specific element of cardiac image quality for the purpose of image optimization, accurate quantification, and diagnosis
- Annotation of an independent echocardiography patient dataset showing four attributes of image quality namely: anatomical visibility, chamber clarity, depth-gain, and foreshortening attributes for A4C, PLAX apical standard views
- Public release of experts annotated patient dataset to allow future studies and external validation of the new approach or methods
- Proposed detailed implementation of multi-stream deep learning architecture pipeline to process and allow access to specific image attributes in A4C and PLAX view of echo cine loop.

5.4 Methodology I

In this section, a detailed account of two sources of dataset namely CAMUS and PACS-1 were explored separately under methodology I and II. The research critically investigated the possibilities of and performance of two objective functions separately, in the assessment of echocardiographic image quality. It is hoped that the results in each approach can lead to an important discovery and establish technical justification on the best and final approach.

The first part of scientific methodology is based on CAMUS public dataset and quality assessment by classification. CAMUS consist of 2D echocardiographic specimens with end systolic (ES) and end diastolic (ED) images and ships with experts' single score annotation. This section details the modelling process of the quality by classification method using the public dataset and the use case scenario for clinical advantages. Several state-of-the-art models were also investigated along with the ultra-lightweight CardioQNet model.

Finally, the section compares the result outcomes from state-of-the- art models and the ultra-lightweight CardioQNet model which was created for this purpose.

5.4.1 Quality Assessment by Classification

This chapter investigates the use of CAMUS image quality assessment using classification approach. Cardiac Acquisition for Multi-structure Ultrasound Segmentation (CAMUS) is a fully annotated dataset claimed to be the largest publicly available (Leclerc *et al.*, 2019) two-dimensional (2D) echocardiographic dataset for clinical quality and segmentation assessment. The full dataset was acquired from GE Vivid E95 ultrasound scanners (GE Vingmed Ultrasound, Horten Norway), with a GE M5S probe (GE Healthcare, US). No additional protocol than the one used in clinical routine was put in place. For each patient, 2D apical four-chamber and two chamber view sequences were exported from EchoPAC analysis software (GE Vingmed Ultrasound, Horten, Norway).

CAMUS Public Dataset Description

CAMUS public dataset is intended for the evaluation of cardiac segmentation, consists of clinical exams from 450 patients, clinically acquired at the University Hospital of St Etienne (France). The dataset comprises a wide variation of acquisition settings. For instance, for some patients, parts of the wall were not visible in the images; this produced a highly heterogeneous dataset, in terms of image quality and pathological cases, which is typical of daily clinical practice data. The full dataset was acquired using GE Vivid E95 ultrasound scanners (GE Vingmed Ultrasound, Horten Norway), with a GE M5S probe (GE Healthcare, US). For each patient, A2C and A4C view sequences were exported from EchoPAC analysis software (GE Vingmed Ultrasound, Horten, Norway). At least one full cardiac cycle was acquired for each patient in each view, allowing manual annotation of cardiac structures at ED and ES.

In total, the dataset includes 1800 2D ultrasound sequences (2 chamber and 4 chamber views of 450 patients) along with the provided multi-structure annotation (i.e., endocardium (LV-Endo), the myocardium (epicardium contour more specifically, named LV-Epi), and the left atrium (LA)) by one expert at the ED and ES instants (Leclerc, Smistad, Pedrosa et al., 2019). An example of images from a public CAMUS dataset with a different range of quality (i.e., good, medium, and poor) are illustrated in Figure 5.3.

Even though the intended use of CAMUS is segmentation, it has valuable contribution to quality assessment hence, the adoption of CAMUS to evaluate quality assessment classification.

5.4.2 Weighted Average Scores Annotations

Normal classification problem presents the requirement to predict a discrete label sample within the label space, this idea can be explored for discretized integers value if thresholds are set for numerical boundaries. CAMUS dataset annotation ships with quality score values represented by integer values from 0 to 9 for each image specimen. Although, classifying discretized samples is not alien to deep learning methods but 2D cardiac samples rated in integer values pose significant challenges to both annotator and machine learning model because echocardiogram specimens can appear to be similar despite having subtle difference that may be considered significant to anatomical and diagnostic considerations. To set the threshold values for discretization, cardiologists were consulted, who having examined the specimen specified three emerging categories as defined below:

- Poor Quality with score ranges from 0.0 – 4.5,
- Average Quality with score ranges between 4.6 – 7.8
- Good Quality with score ranges between 7.9 – 9

Figure 5.3 illustrates the emerging discretized class samples for A4C and A2C. Consequently, the emerged distribution set apportioned for training and development subsets are illustrated in histogram distribution shown in Figure 5.4.

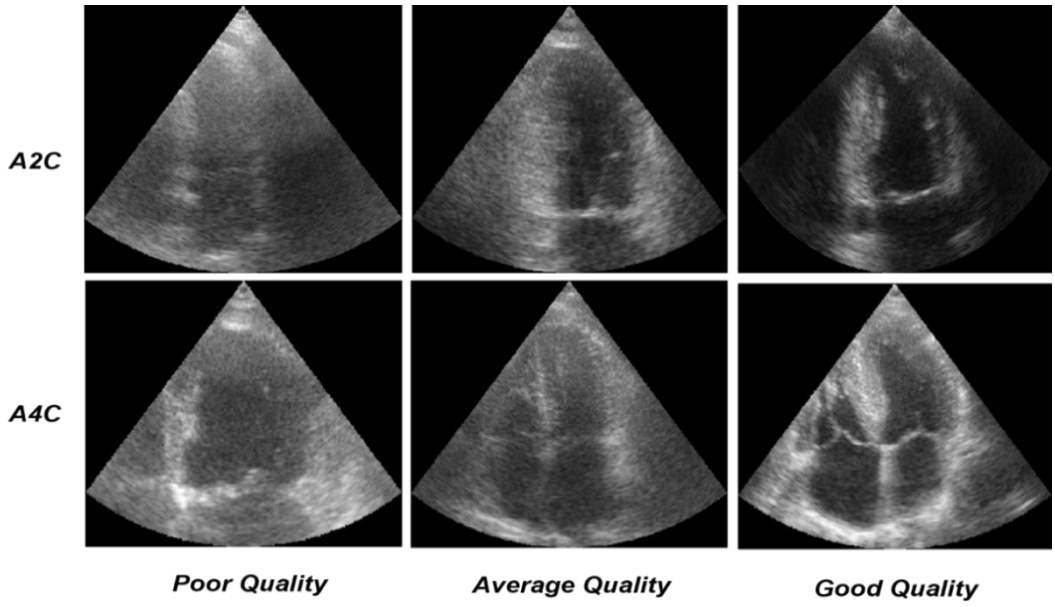


Figure 5.3: CAMUS A4C samples of three (3) classes of quality detected on end-systole (ES) frame, 3 classes for end-diastole (ED), making a total of 12 classes. Each class ranges from poor quality to high quality with image size 256 x 256 x 1.

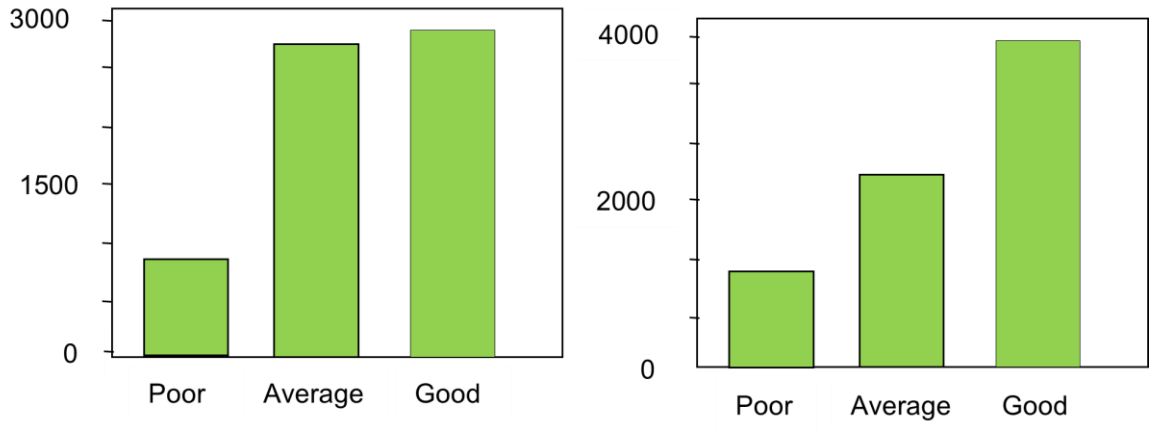


Figure 5.4: Histogram Distribution of CAMUS (20% Test Set), for combined ES/ED A2C and A4C respectively. A highly imbalance distribution exists for both apical two chamber (A2C) left and apical four chamber (A4C) right.

5.4.3 Classification Model Training

In this section, a description of a proposed nested model that has been achieved by semi-manual optimization technique was partly inspired by the work of Abdi, et al. (2018) is provided.

The architecture used is an ultra-lightweight deep convolutional model derived from semiNAS. This is a 3-layer DCNN architecture which accepts fixed length images of spatial size 224 x 224. Each of the input images convolved with the three-layered architecture with its respective Max Pooling layers ReLU activation to form a 2D feature map. The model operates in the temporal domain where image temporal features are extracted for classification of image quality. While cross entropy was chosen for loss function equation (5.0), stochastic gradient descent function was adequate for weight optimization. The architecture is depicted under technical background in chapter 3.

Batch selection: Computational cost during training phase was not significant with batch selection of 14 but shows significant reduction in accuracy and overfitting on higher batch rates. The training phase on CAMUS batch of 14 of 6 classes is considered as optimum since this produced the best model accuracy than selecting higher batch rates. Also, memory utilisation becomes significantly apparent at higher batch selection of 64 than at optimum batch rate. Hardware performance in terms of computational speed showed a negligible increase in training times and did not affect model's ability to properly generalise new test samples.

$$CE_{LOSS} = -\sum_{i=1}^{N} y_i \log\big(h_\theta(x_i)\big) + (1 - y_i) \log\big(1 - h_\theta(x_i)\big) \quad (5.0)$$

Data augmentation: Data augmentation was applied to allow optimum learning sequences for the models; a maximum translation of [-0.05, +0.05] pixels and maximum rotation of 10 degrees were applied randomly for horizontal, vertical, and rotational angles respectively. To prevent overfitting in the training phase, batch normalisation was applied at each convolution layer with early stopping and dropout (rate 0.30) for the training samples. Batch normalisation also helps stabilises and speeds up convergence during the training phase.

Over sampling: As clearly illustrated in Figure 5.4, CAMUS dataset was highly unbalanced, and would impair model's training result if appropriate generalisation methods were not applied during the training process. The solution was to implement a generalisation algorithm known as (over-sampling); a technique commonly used to allow minority class to get as much sampling as the majority class. This means that all minority classes would be oversampled during the training phase, while the majority class would gain as much as sampling rate as the minority class under sampled in achieving a balanced dataset. Over-sampling was implemented in PyTorch and applied online to all minority classes with good success.

Hardware resources: The model was implemented on the PyTorch backend. The experiment was carried out on GPU GeForce GTX 970 chipset's Maxwell architecture, featuring 4GB RAM coupled to 1,664 CUDA cores.

5.4.4 Model Evaluation Metrics

Similarly, to (Ornstein and Adam, 2021) and (Leclerc *et al.*, 2019), the quality of the data fit via model accuracy per fold was assessed using the following metrics.

Performance Accuracy: Overall accuracy was calculated as the number of correctly classified images as a fraction of the total number of images which can be computed as follows:

$$Acc_{\mu} = \frac{\sum_{i=1}^{k} \frac{tp_i + tn_i}{tp_i + tn_i + fp_i + fn_i}}{k} \tag{5.1}$$

$$Recall_{\mu} = \frac{\sum_{i=1}^{k} tp_i}{\sum_{i=1}^{k} tp_i + fn_i} \tag{5.2}$$

Confusion matrix plot: Confusion matrix is a two-dimensional matrix that presents a brief overview of the classification performance of a classifier on a test dataset which gives us insight not only into the error being made by a classifier but more importantly the types of errors that were made. In one dimension the true classes of the test dataset, and in the other dimension, the prediction results by the classifier will be assigned. In other words, the confusion matrix displays the number of correct and incorrect predictions broken down by each class (Grandini, Bagli and Visani, 2020).

$$Precision_{\mu} = \frac{\sum_{i=1}^{k} tp_i}{\sum_{i=1}^{k} tp_i + fp_i} \tag{5.3}$$

$$Error\ Rate = \frac{\sum_{i=1}^{k} \frac{fp_i + fn_i}{tp_i + tn_i + fp_i + fn_i}}{k} \tag{5.4}$$

Quality classification methods using CAMUS dataset suffer many drawbacks in combined classes for ED and ES images. Model achieved **69.23%** (Figure 5.6) but **91.23%** for apical four (A4C) and apical two (A2C) end systolic images combined (Figure 5.5).

5.4.5 CAMUS – Results and Analysis

Some metrics employed in the evaluation of classification performance of the examined and proposed classification models in this study includes classification accuracy, confusion matrix. Accuracy expressed in terms of number of correct predictions over total number of samples in label space while confusion matrix presents graphical overview of classification performance on the given dataset. Confusion matrix (Figures 5.5, 5.6) provides insight into what errors were made and where in the label space the error exists (Grandini, Bagli and Visani, 2020). Furthermore, the model's performance was compared to VGG16, ResNet50, and MobileNetV2 in Table 5.1.

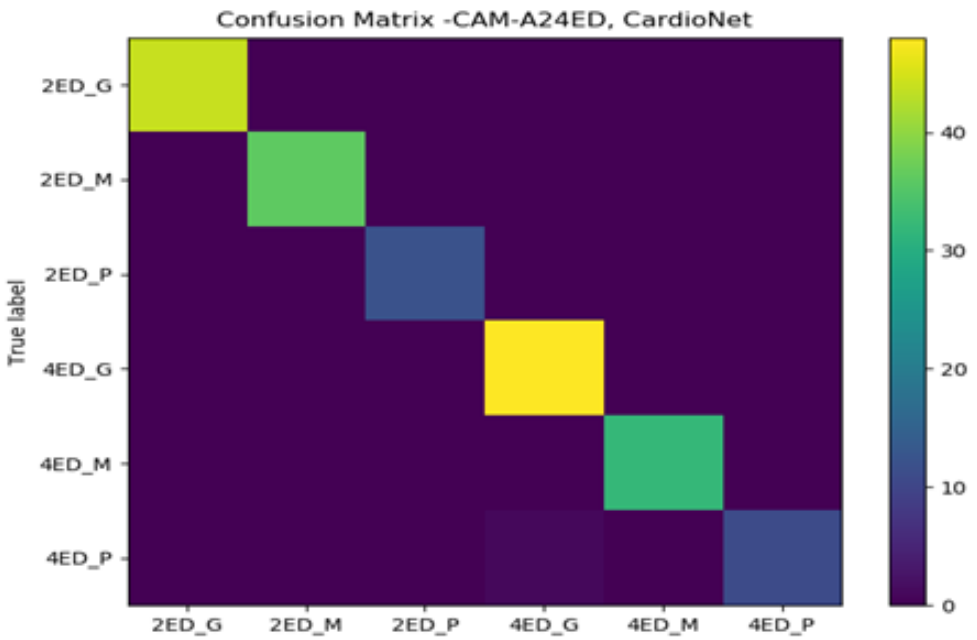


Figure 5.5: Performance Evaluation: Confusion Matrix for only (ED) cardiac specimen for A2C+A4C on CAMUS dataset. The proposed architecture achieved 91.23 percent accuracy on end diastole specimens only.

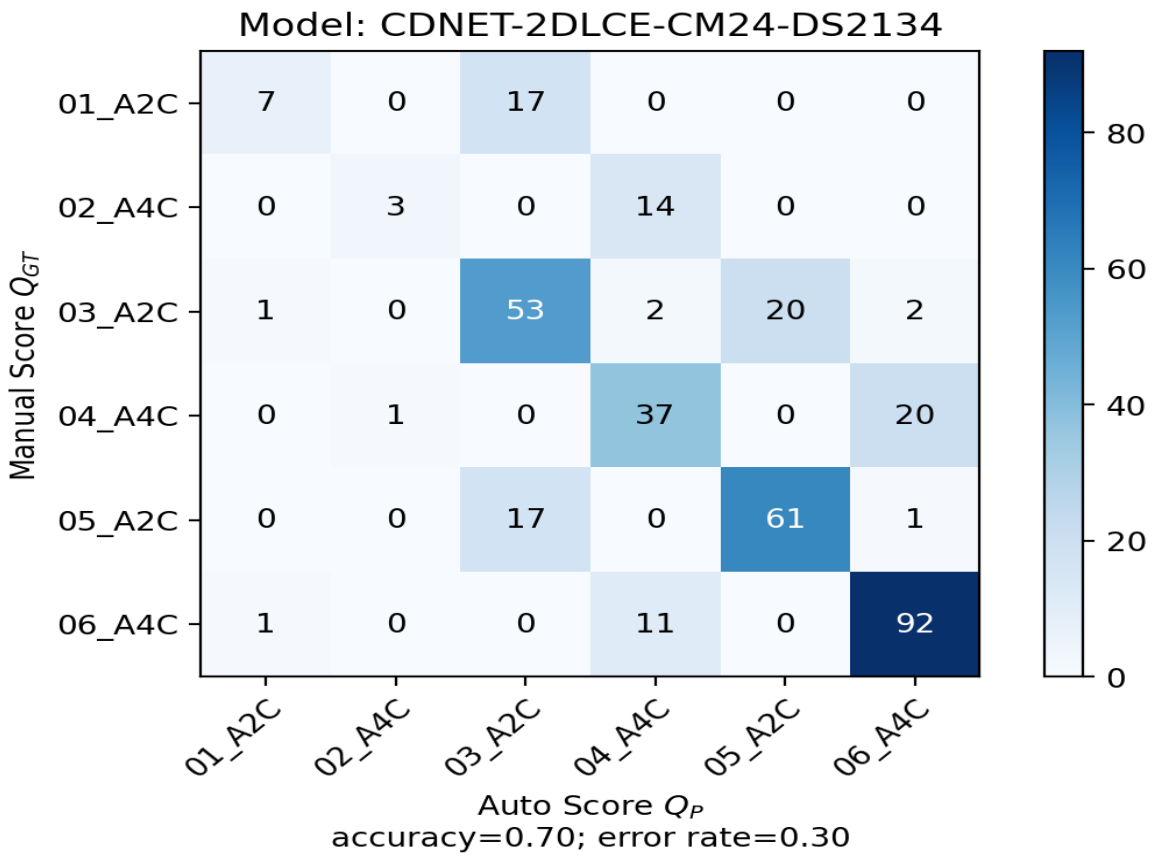


Figure 5.6: Performance Evaluation: Confusion Matrix for the Combined End-Systole and End-Diastole (ED+ES) frames on CAMUS (A2C, A4C) dataset. The proposed quality classification model achieved 70.00% accuracy on end systole and end diastole cardiac specimens.

Table 5.1: The comparison of model performance on combined A2C and A4C of CAMUS public dataset. Echo images consist of end-systole and end diastole frames (ES+ED). The ultralightweight (CardioQNet) model shows significant appreciation performance over VggNet and better inference speed over all the state-of-the-art models evaluated. Performance comparison based on mini-batch size of 24, and 50 epochs per model.

	VGGNet16		ResNet50		MobileNetV2		CardioQNet	
Optimizer	ADAM	**SGD**	**ADAM**	SGD	ADAM	**SGD**	**ADAM**	SGD
Test Accuracy	**63.24**	**70.00**	**70.00**	**68.07**	**69.15**	**72.18**	**70.00**	**69.44**
Test Loss	**0.0325**	0.0363	0.0297	0.0301	0.0293	0.0300	0.0306	0.0309
Network Depth	16		50		53		6	
Total Trainable Parameters	17.93 million		23.77 million		12.70 million		2.7 million	
Run-Time / epoch (sec)	**50.38**		**38.40**		**27.26**		**22.90**	
Inference Time/Frame (ms)	**11.852**		**9.307**		**7.261**		**6.925**	

Model Performance: The state-of-the-art models (ResNet50, VGG16 and MobileNetV2) evaluated on CAMUS quality classification yielded comparative results on Adaptive Moment Estimation (ADAM) optimizer for ResNet50 to CardioQNet model. Except for MobileNetV2 and VGGNet which thrived on SGD optimizer for the same observations. Obvious results from Table 5.1, the comparison of model's accuracy indicated a significantly low performance on end-systole (ES/ED) frames. This was not the case when considering the end diastolic frame (ED) separately from the rests. The end-systolic frames usually indicate the compression of the myocardium, this action last over a couple of minutes. Therefore, a varying or dissimilar pixels formation can only be represented in a discrete sample within the activity frame length, this is known as echo frame, and it contains varying spatial noise. Models' performance accuracy therefore, reflects on the complexities of varying spatial noises, limited volume of data space, class imbalanced and model generalizability.

Test loss values are indicative parameters that training did converge, although each model shows a level of overfitting which was controlled by early stopping. This explains why stochastic gradient descent (SGD) optimizer slightly appreciated on VGGNet16 and MobileNetV2 as opposed to ADAM on ResNet50 and CardioQNet models.

While MobileNet and CardioQNet showed fast convergence on the dataset, MobileNet's performance accuracy was highest on CAMUS, exceeded its counterparts by 2.8%. Although some aspects of echocardiographic specimens considered in this chapter still precipitate stringent questions, especially in terms of quality estimation and weighted average score predictions, the approach to quality classification would fit a regression task that can provide adequate generalization with multiple score values per frame.

Also, classical model classification without the advantage of a spatio-temporal feature extraction would significantly limit a model's ability to generalise cardiac specimens with artefacts and label noise which is synonymous with cardiac samples.

Apart from highly imbalanced dataset which could potentially impair model's accuracy, complex data structures in the end-systolic (ES) frames, and very closely related features in cardiac frames could further lower models' performance as the case was for all models investigated as results indicated in Table 5.1.

Inference time: Although, the inference times achieved by each model is significantly low to allow more than 60FPS processing, however, in practice, this may not accurately be so, especially when used for remote assessment. Therefore, the lower the inference time quoted the better. CardioQNet is in a stronger position on this with capacity for 144 frames per seconds. Furthermore, CardioQNet leads in epochs training time as a direct result of having the lowest trainable parameters, which effectively avoid redundancy in the trainable parameters in the architecture and network depths.

Finally, the use case scenario for an automatic image quality assessment algorithm in clinical setting would consequently require more than a single quality score (known as weighted average score) value on cardiac frame and solving the quality classification by regression method would be suitable for unified workflow or translatory advantages in the healthcare industry. Going forward, it's obvious that a new cardiac dataset consisting of multivariate annotations per echo cine loop is highly desirable for the implementation of automatic quality assessment in clinical practice.

In this scenario, each specimen would be classified based on the experts' opinion score for aspect of quality of an image. This novel idea was conceived, discussed further, and implemented in method II.

5.5 Methodology II

In this section, a detailed account of a variant of picture archiving and communication system dataset known as PACS-1 dataset is given. PACS-1 dataset contains echocardiographic samples of A4C, and PLAX specimens randomly drawn to constitutes essential dataset for research purpose. This second part methodology provides analytical justification for the assessment of quality elements by regression and modelling of the anatomical characteristics for objective quality standard framework.

It is evident by experimental results that prediction of quality scores using classification model does not enjoy high level accuracy in terms of model performance and inference, hence, regression models that predict continuous values with the specific breakdown of quality elements would require a new set of criteria that is measured by multiple numeric scores. These are significantly relevant for echocardiographic image quality assessment. Therefore, PACS-1 dataset with multiple score annotations, ultimately became relevant for image quality assessment.

5.5.1 Quality Assessment by Regression

In this section, a detailed account of a variant of picture archiving and communication system dataset known as PACS-1 dataset is given. The original PACS dataset contain 14 apical views out of which A4C, and PLAX specimens were randomly drawn to constitutes PACS-1 dataset.

At present, no echocardiogram dataset with the corresponding multivariate scoring method for A4C/PLAX image quality assessment is publicly available. This novel scoring approach provides specifically new insight to the aspect of echocardiographic image element lacking optimum quality characteristic instead of a single score annotation. Therefore, this research work aimed at preparing our own dataset (echocardiograms and corresponding ground-truth) for model developments and training. Importantly, are the essential and novel definition of quality attributes that are necessary for objective assessment of 2D image quality using PACS-1 private dataset and the clinical justification for the collective elements required for objective standard assessment of image quality in A4C and PLAX views.

PACS-1 Private Dataset Description

The study population consisted of a large random sample of echocardiographic studies from different patients (age ranges from 17 and 85 years), who were recruited and had undergone echocardiography between 2010 and 2020. A total of 11,262 DICOM formatted videos was extracted from Imperial College Healthcare NHS Trust's echocardiogram database. Ethical approval was obtained from the Health Regulatory Agency for the anonymized export of large quantities of imaging data. It was not necessary to approach patients individually for consent of data originally acquired for clinical purposes. The images were acquired during examinations performed by experienced echocardiographers, according to the standard protocols for using ultrasound equipment from GE Healthcare and Philips Healthcare manufacturers. Automated anonymization was performed to remove the patient-identifiable information from echocardiographic videos. A neural network model, previously developed and detailed in chapter 4 was then used in identifying and separating different echocardiographic view to obtain A4C and PLAX required for quality assessment modelling. This resulted in a total of 33,784 frames from different patients: 15,476 and 18,308 frames for A4C and PLAX, respectively.

5.5.2 Multivariate Quality Attributes

The inherent limitations imposed by classification model on 2D image quality and results of the evaluation on CAMUS' quality assessment method has precipitated the discovery of multivariate quality assessment method as a proposed and alternative solution in image quality assessment to the weighted average quality assessment (WAQ) method, which imposed limitation on objective models. The usage of a weighted average score system which evaluate the overall quality prediction of an image is terribly insufficient for real-time objective grading, optimization or as feedback for operators' guidance system. In weighted average scoring system investigated, objective model can only predict a single score value to depicts overall quality score for an echo image. This is grossly lacking in specific quality element of the image under assessment, hence its insufficient in real life practices. In other words, there are no associated clinical advantage in such approach.

Therefore, for the first time in the history of echocardiogram objective quality assessment, images with view-specific scoring criteria were proposed and defined to encapsulate the myocardial anatomical and pathological profile. These are known as domain-specific criteria of two-dimensional (2DE) echocardiogram image quality. This proposal was put forward to the

consulting clinical experts and considered a plausible solution for at National Heart and Lung Institute, London.

Unlike a single weighted average method which, after modelling, could not provide the indication for specific quality and reason why cardiac specimen is tagged as poor quality or provide critical path to optimization other than relying on operator subjective skill for manual optimization.

The novel modality provides new insight into how 2D echocardiograms should be quantified or graded. Sine our focus was on A4C and PLAX with reason given above, we defined 23 domain-specific criteria (Table 5.2) which was grouped under four (4) pathological and anatomical descriptions namely: Anatomical Visibility, Anatomical Clarity, Signal Depth-Gain and Foreshortedness. These are enumerated below as follows:

(a) Anatomical Visibility:

Unlike photographic images, ultrasound images are formed by interference patterns of scattering centres that do not present clear edges, but inherently poor lateral and axial resolutions (Labs *et al.*, 2020). Hence, the magnitude of visibility on chamber cavities for both A4C and PLAX frames can be thought of in terms of correct slicing of heart's apex, within the acceptable clinical projection of anatomical structures either with sharp or blurred edges of amplitude structures. Equations (5.5), (5.6) describe the rotation of a frame vector in two-dimensional spatial distribution where $x1y1$ represent on-axis projection, taking arbitrary centre xc, yc, off-axis xp, yp can thus be mitigated from β known angle to improve anatomical visibility. In A4C, emphasis is placed on apical orientation, echogenicity of the left ventricle chamber, mitral and atrium valves (Mitchell *et al.*, 2019). Although the LV apex is not visualised in PLAX, emphasis is placed on the anatomical echogenicity and clinical orientation of the right ventricle, left ventricle, the pericardium positions, and the aortic valves. These are clinically relevant features experts rely on for quantification, clinical assessment, and diagnosis.

$$x_1 = (x_p - x_c)\, cos\beta - (y_p - y_c)\, sin\beta + x_c \qquad (5.5)$$

$$y_1 = (x_p - x_c)\, sin\beta - (y_p - y_c)\, cos\beta + y_c \qquad (5.6)$$

(b) Anatomical Clarity:

Left ventricle clarity is a define the contrast between the tissue and image background noise. Echocardiogram's apical chambers of any zoomed region can only present rough boundaries and contractive edges. Kurt *et al.*, (2009), have demonstrated the impact of contrast echocardiography, however, with respect to quantification, cavity clarity is visualised by several distinguishable fast-moving pixel's formations during cardiac cycles. This attribute, therefore, addresses the degree of distinguishable pixel element representing the endocardial border cavities or clear distinction between the intraventricular septum, valves, any trabeculated pericardial fluids and endocardial walls. Cardiac frames with very high contrast or very low contrast represent the extreme end of the spectrum and pose significant challenges (Nagata *et al.*, 2018), (Kurt *et al.*, 2009) with newly qualified clinicians. Equation (5.7) describes the root mean squared (RMS) contrast, $C_{i,j}$ which does not depend on angular frequency content or spatial distribution as suited for 2D cardiac frames and given by the difference between the standard deviation of normalised pixel intensity $I_{i,j,}$ and mean normalised intensity $\widehat{I,}$ of a given anatomical pathology; where (i, j) represents the $i^{\text{-th}}$ and $j^{\text{-th}}$ element of 2D image size M, N; A very high contrast could generate artefacts and potentially obscured essential anatomical details. Unfortunately, some images with low contrast do have significant anatomical details required for clinical measurement hence the need to assess each image on the merits of clarity.

$$C_{i,j} = [\frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(I_{i,j} - \hat{I})^2]^{\frac{1}{2}} \qquad (5.7)$$

(c) Signal Depth-Gain:

Depth-gain is peculiar to 2D echocardiography, and it represents a measure of intensity of discrete signal samples of a specific region of interest. The intensity of the image signals becomes susceptible to depth changes, sector width and patient anatomical differences. Although the use of high frequency probes can yield better resolution intensity at shallow tissue depth penetration (Sassaroli *et al.*, 2019), low frequency probes give the opposite effect. Consequently, signal gain at the image apex (near field) usually possesses strong intensity of high amplitude and could become excessively low at the far field region of the cardiac frame. In the same way, excessive gain can present as pulmonary fluid in some cases (Dong *et al.*, 2020) and images with very low gain attributes but bear significant anatomical details or

noticeable artefact are not ignored in clinical practice. Equation (5.8) describes the intensity of reflected beam, which is associated with depth gain; where $d^2\phi$ represent the luminous flux of the infinitesimal area of source $d\Sigma$, dividing by the product of $d\Sigma$, infinitesimal solid angle $d\Omega_\xi$ and θ_ξ angle between the normal Ω_ξ to the source $d\Sigma$. While luminance is the photometric measure of the pixel luminous intensity per unit area of light at a given area of interest, brightness therefore is the subjective impression of the object of luminance $I_{i,j}$ and is measured in candela per square meters cd/m2. Clinical significance therefore assesses and score's potential introduction of artefact from excessive gain, incorrect depiction of true anatomical tissues or obscurity of relevant anatomical details which is relevant for measurements.

$$I_{i,j} = \frac{d^2\phi}{d\Sigma \,.\, d\Omega_\xi \, cos\theta_\xi} \qquad (5.8)$$

(d) LV Foreshortening:

Apical foreshortening is a distortion of heart's apical visibility that occurs during image acquisition. Basic definition of apical foreshortening specified as central misalignment of image plane with LV apex, However, this seems too simplistic when considering the possibility of skewed projection of image plane and LV's apex with respect to perspective transformation. This potentially alters LV volume and shapes become geometrically incongruent (Ünlü *et al.*, 2019). To add to the complexity, foreshortening could occur in either of the cycles hence, both diastolic and systolic cycles are considered during frame real-time assessment. Smistad et al., (Smistad *et al.*, 2020), have described the importance of real-time detection of apical foreshortening. For instance, foreshortening can result in inaccurate quantification of ejection fraction (EF) (Labs *et al.*, 2020) or prevent the detection of crucial pathology especially in the apical region. We refer to this undesirable transformation in terms of the product of homogenous properties given in equation (5.9). In the PLAX view, however, where LV apex visibility is not required, visible apex of the LV could be taken as 'false-apex' (Mitchell *et al.*, 2019), therefore counts as LV foreshortening. From a clinical standpoint, eliminating foreshortedness is paramount for anatomical assessment and diagnosis of many ailments including cardiomyopathy.

$$I_{x,y,z} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{d} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ -\frac{1}{d} \end{bmatrix} => \left(-d\,\frac{x}{z}, \qquad d\frac{y}{z}\right) \quad (5.9)$$

5.5.3 Multi-Layer Annotation Process

To establish the ground-truth scoring for neural network developments and testing, A manual score criteria was established based on the specific pathological and anatomical features, these were done under the supervisory team consisting of cardiologist, sonographers, and an atrioventricular (AV) specialist in what we describe as the consortium.

The entire A4C and PLAX echo cine loops were independently studied and manually annotated by two professionals: an experienced cardiology expert and an accredited professional. Using a custom-made program (Figure 5.7), purposely developed in MATLAB which closely replicated the interface of echo hardware. The program allows each echo cine loop to playback the length of frames allowing visual inspection of the cine loops and functions to control the animation of the loops. Each cine loop is visually evaluated against the criteria in Table 5.2 while using score slider to assign visual score on each attribute. Thus, quality score is assigned and recorded. This is repeated for each defined attributes in Table 5.2 yielding four different scores per frame. Consequently, annotation records are created automatically in MATLAB file (.mat) format after each cine loop. Each echo cine has a corresponding .MAT file that is stored in a protected folder or drive.

Score attributes range from 0 to 9 to allow specificity and fair assessment of A4C/PLAX apical standard. Consequently, three frames were randomly drawn from each video and split into training (27,028 frames), and testing (6,756 frames) sub-datasets in 80:20 ratios. Figure 5.9 summarises the sample distributions for A4C and PLAX with categorical characteristic using experts' score range values of 0 - 4.5, 4.6 - 6.5, 6.7 - 9.9 classified as poor, average, and good quality respectively. This level of quality classes is illustrated in Figure 5.8.

Figure 5.7: Visual-aid utility program – Implemented in MATLAB for subjective assessment of ground truth annotations, labelling and scoring: 1,600 echo cines were independently studied and scored prior to extraction. Cine with scores less than 0, or negative score were discarded. The average scores for the entire cine loop were automatically recorded. This process repeats for each of the four quality attributes defined.

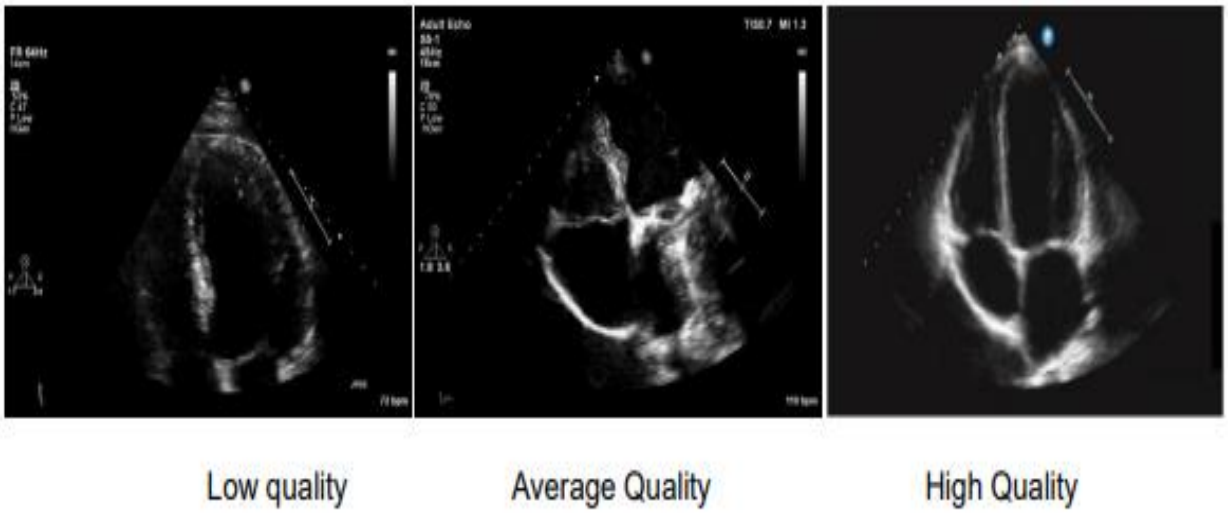


Figure 5.8: PACS Dataset: The three emerging quality levels used. Poor Quality, Average Quality and Good Quality represented in the legend for both A4C and PLAX standard view specimens.

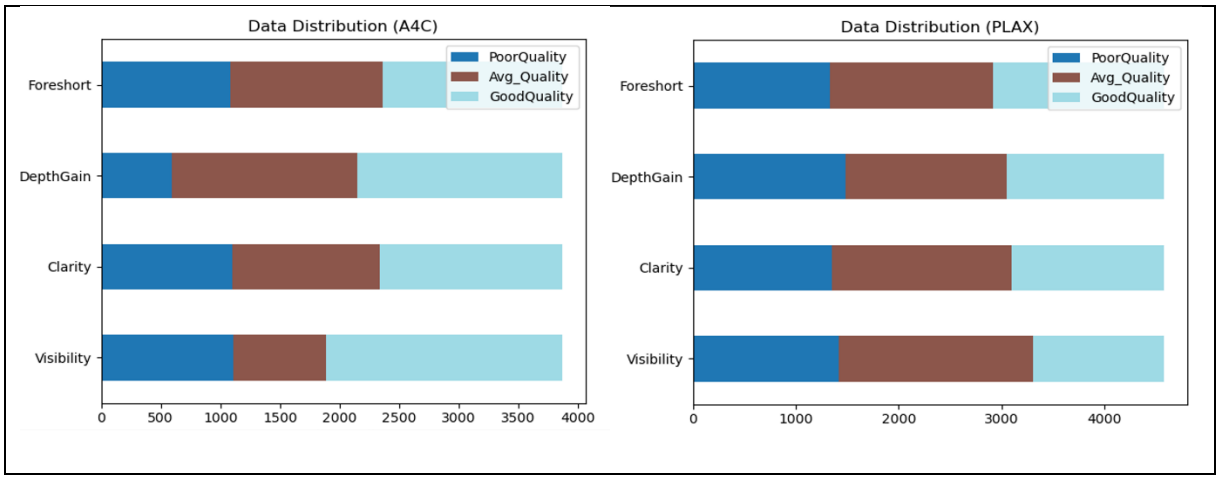


Figure 5.9: Summary of data distribution for A4C and PLAX cardiac frames, indicating three categories of quality levels based on experts scores values: Frames with max scores of 4.5, 6.5 and 9.9, are classified as Poor Quality, Avg. Quality, and Good Quality, respectively.

Table 5.2: View-specific scoring definition. The quality of each view was evaluated according to several criteria; each criterion consisted of several attributes with independent scores but yielding a maximum score of 10 points for each criterion.

A4C			**PLAX**	
Assessed Element per Attributes	Manual Score awarded		Assessed Element per Attributes	Manual Score awarded
ANATOMICAL VISIBILITY:			**ANATOMICAL VISIBILITY:**	
Correct Axis, Apical Segment	6		Left Ventricle (LV) Visible	5
Interventricular Septum Visible	2		Right Ventricle (RV) Visible	3
Interatrial Septum Visible	2		Full Segment Pericardium Visible	2
ANATOMICAL CLARITY:			**ANATOMICAL CLARITY:**	
LV Cavity clarity, clear edges	4		LV Cavity Clarity (distinguishable border)	4
Distinguishable Valves	3		LV Anteroseptal Wall Clarity	3
Distinguishable Septum Wall	3		LV Inferolateral Wall Clarity	3
SIGNAL DEPTH-GAIN:			**SIGNAL DEPTH-GAIN:**	
Image Sectorial Gain	4		Sectorial Gain	4
No Excess Gain	3		No Excess Gain	3
Minimum Artefacts	3		Minimum Artefacts	3
LV FORESHORTEN:			**CAVITY FORESHORT:**	
LV Apical Segment present	4		No-Apex Diastole	5
Normal-Shaped Diastole	3		No-Apex Systole	5
Normal-Shaped Systole	3			

5.5.4 Regression Model Training

Details of the well-known and state-of-the-art network architectures investigated in this study can be found in relevant resources: MobileNet (Howard *et al.*, 2020), ResNet (He *et al.*, 2015) and VggNet (Simonyan and Zisserman, 2015) along with its hybrid versions.

The architecture used in this study, is based on a multi-stream, multi-label regression model, featuring CardioQNet as weight weight-sharing module and four (4) sub-node model architectures fused together in TensorFlow API to simultaneously train and make predictions in a multi-label, multi-class fashion. The model accepts variable/fixed length sequence frames of spatial size 224 × 224 x 3 sequence convolved with each parallel convolutional layer of each sub-model to compute spatial feature map, F given by equation (5.10), which is fed into a time sequence module (LSTM) for temporal feature extraction. Each convolution layer features an activation function of type - Rectifier Linear Units (ReLU) (Nair and Hinton, 2021), equation (5.11). Each node was dedicated to extract specific anatomical features relating to criteria defined for A4C and PLAX standard views in Table 5.2. The components of the sub-node architecture are logically optimised for specific quality attribute and adapted based on best performing architecture against each quality attribute. The model is illustrated in Figure 5.10, and detailed as follows:

(a) The spatial module consists of four parallel convolutional layers, dedicated to extraction of hierarchical features on each of the defined quality attributes. These are termed as quality-specific layers, receiving their inputs from a weight sharing convolution architecture CardioQNet. Each quality-specific layer features batch normalisation (Ioffe and Szegedy, 2017), between each ConvNet layer except the third layer which missed out on max pooling (Wu and Gu, 2015), and dropout of 0.5 (Srivastava *et al.*, 2014). The output is flattened, and the sequence is fed into the temporal module.

(b) Temporal module consists of an LSTM layer, used to extract temporal features. It accepts vector data from each adjacent module to compute mean score on frames' sequence data. Features are based on fast changing pixel intensity between consecutive frames, which could result in spatial noise where sudden increase in varnishing gradients can be apparent, especially during training phase. Therefore, output layers were configured differently to include dense layer, batch normalisation and dropouts of 0.5. This was noted to offer resilience against noisy labels and reduce variance in the image/frame data.

The choice of architecture construct Table 5.3 was based on the performance data, memory requirement, and inference speed which are significant for real-time quality assessment and future operators feedback implementation. The model was trained with a k-fold cross validation technique to ensure adequate learning on the dataset and performance was recorded for each model. The hyper parameters learning rate was set at 0.002 with high momentum 0.95 and decay rate of 0.1 every 24 steps and were reproducibly initialised to minimise possible deviation in score performance. Training was initialised with 32 batch size and completed as learning curves converged around 40 epochs.

Data augmentation was applied to allow optimum learning sequences for the models; a maximum translation of [-0.05, +0.05] pixels and maximum rotation of 5 degrees were applied randomly for horizontal, vertical, and rotational angles, respectively. To prevent overfitting in the training phase, we applied batch normalisation and dropout. A multi-label optimization approach was adopted (Díez *et al.*, 2015), and the model was trained simultaneously using four quality attributes with mean absolute error as the cost function.

$$F^{l}_{(i,j)k} = \sum_{i=0}^{n}\sum_{j=0}^{m} w^{l}_{i,mn}\, F^{l-1}_{(j+m)(k+n)} \qquad (5.10)$$

$$f(x)_{relu} = max(0, x) \qquad (5.11)$$

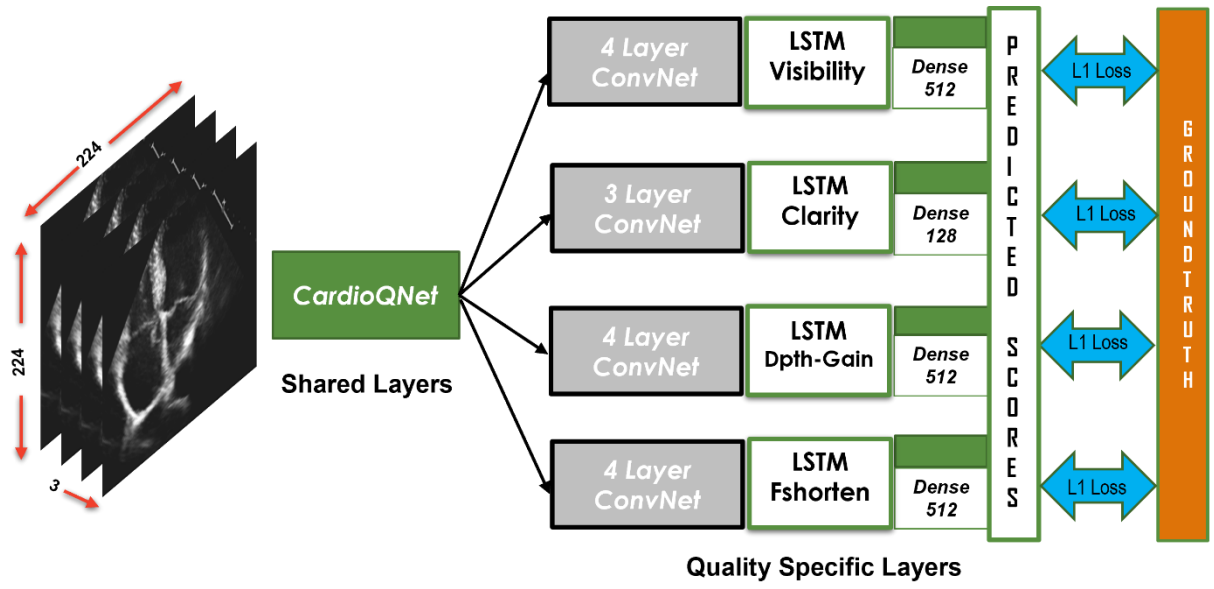


Figure 5.10: Multi-Stream Neural Network Architecture used in this study, returning a prediction for each of the four quality attributes. Each ConvNet is identified by its objective function: visibility, clarity, depth-gain and foreshortening attributes respectively.

Table 5.3: Network architecture summary. Numbers in the _K: w x h represent the number of kernels per layer, and kernel size in w x h respectively.

CardioQNet	Quality-Specific Layers			
Shared Layers	**Visibility**	**Clarity**	**Depth-Gain**	**Foreshorten**
32: 11x11	32: 3x3	16: 3x3	32: 3x3	16: 3x3
MaxPool	32: 3x3	Batch Norm	32: 3x3	16: 3x3
32: 7x7	MaxPool	MaxPool	Batch Norm	Batch Norm
MaxPool	64: 3x3	32: 3x3	MaxPool	MaxPool
64: 32x32	64: 3x3	Batch Norm	64: 3x3	32: 3x3
MaxPool	Batch Norm	MaxPool	64: 3x3	32: 3x3
ReLU	Max Pool	64: 3x3	Batch Norm	Batch Norm
	Dense 512	Dense 128	MaxPool	MaxPool
			Dense 512	Dense 512

5.5.5 Model Evaluation Metrics

Since the model uses multiplex variables for each score attributes, the output score was normalised to [0 ~ 1] vias sigmoid activation, equation (5.13) and model error was individually evaluated using MAE in equation (5.12), by taking the sum difference between cardiologist's ground truth scores (Q_{GT}) and model's predicted score (Q_P). Minimal error therefore, indicates best fit. Average model's performance in terms of accuracy was computed to allow fair comparison with any similar regression model in equation (5.14).

$$MAE = \frac{\sum_{i=1}^{n} | Q_{GTi} - Q_{pi}|}{n} \qquad (5.12)$$

$$f(x)_{sigmoid} = \frac{1}{1 + e^{-x}} \qquad (5.13)$$

$$Model_{acc} = 1 - \left(\frac{\sum_{i=0}^{n} |Q_{GTi} - Q_{pi}|}{n}\right) * 100 \qquad (5.14)$$

5.6 Results and General Discussion

The results presented here include the findings and summary on echocardiogram dataset quality by classification and regression task using semiNAS derived ultra-lightweight model. While model performance in the classification tasks were compared to the state-of-the-art method on CAMUS and PACS dataset, the results provided technical insight to quality assessment modelling and presents clinical justification as follows:

5.6.1 PACS-1 Image Quality by Regression

Cardiac echo frames are laced with significant complexities among which are patients varying anatomical and pathological differences, these complexities are reflected in each fast-moving echo frame therefore, the model's inference speed is very critical to assessment and real-time operators' guidance. Hence, implementing a customised model that can successfully generalise with high confidence, and speed would be counted a success in objective automated assessment. Our model's measured mean accuracy of 96.20% and 2.52ms inference speed reinforces the viability for real-time feedback on specific quality per frame. The error distribution per quality attribute is depicted in Figure 5.11, while the model performance for visibility, clarity, depth-gain, and foreshortening attributes yields 94.4%, 96.8%, 96.2% and 97.4% of image quality, respectively. Samples of predicted cardiac frames in Figure 5.12 indicate the objective scores for visibility, clarity, depth-gain and foreshortedness automatic score generated by the model.

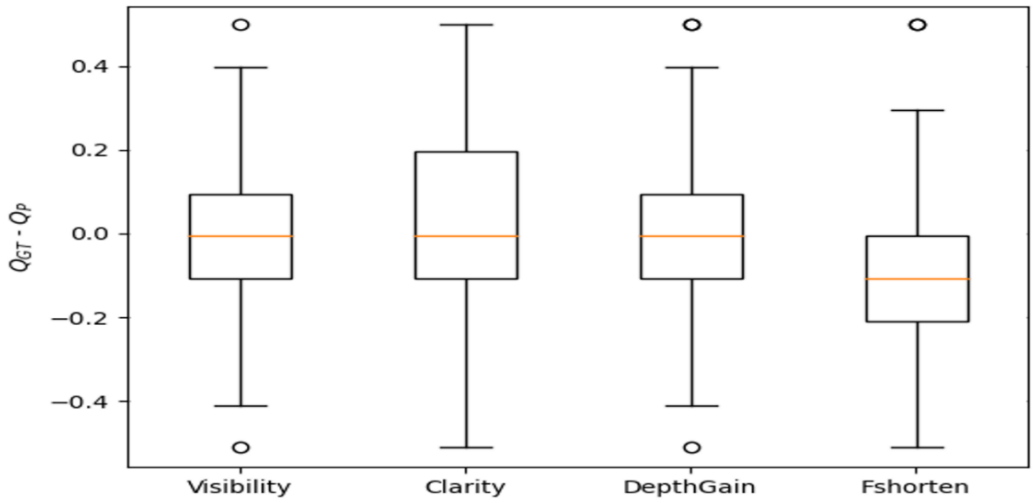


Figure 5.15: Error distribution on the difference between experts ground truth and model predictions per model's quality attributes.

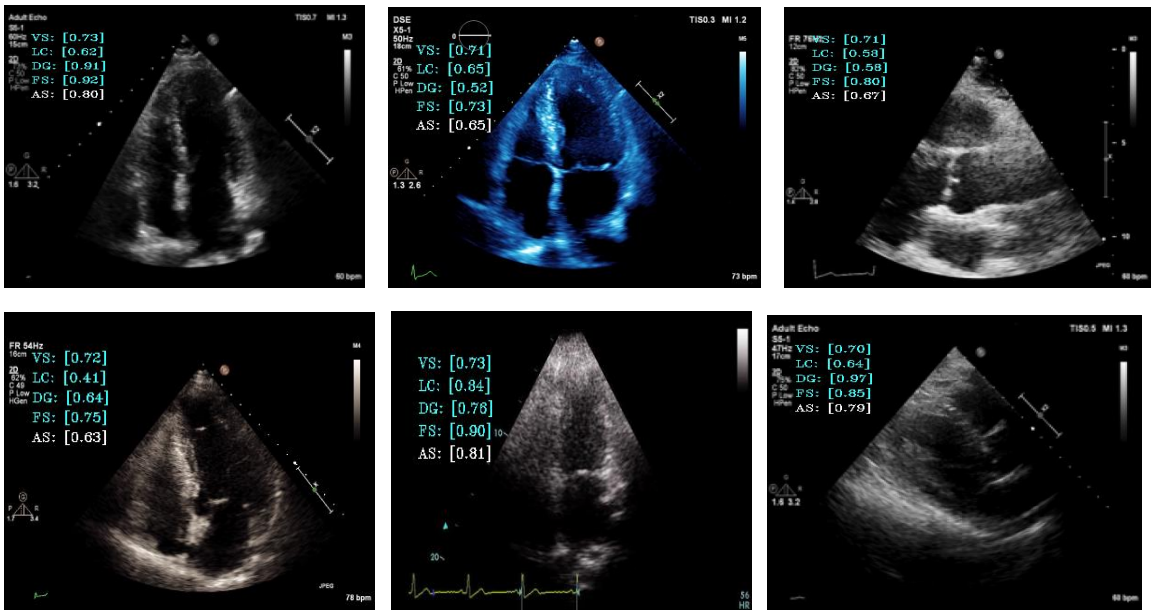


Figure 5.12: Sample of predicted images with respective objective scores. visibility (VS), clarity (CL), depth gain (DG), foreshortening (FS) and overall quality score (AS) are used to assess cardiac image quality during acquisition under clinical conditions.

Table 5.4: Comparison of quality attributes model's accuracies using selected state-of-the-art model and CardioQNet

	REGRESSION ACCURACY (%)					Inference Time (ms)
MODEL DATA	**VISIBILITY**	**CLARITY**	**DEPTHGAIN**	**FORESHORT**	**ACC.**	
VGGNET16	92.30	97.20	98.40	89.20	94.28	30.760
RESNET50	89.45	92.25	87.40	92.20	90.32	19.526
MOBILENETV2	92.35	88.22	95.82	90.64	91.76	15.802
CARDIOQNET	**94.40**	**96.80**	**96.20**	**97.40**	**96.20**	**2.52**

5.6.2 Baseline Results and Discussion

The results of each state-of-the-art model, shown in Table 5.4, vary substantially on each attribute even though each retains its original hyperparameters. This indicator proved that each of the state-of-the-art models is incapable of delivering similar performance in terms of inference speed for real-time deployment. Inference speeds of **30.76ms**, **19.53ms,** and **15.80ms,** was achieved by VggNet16, ResNet50, and MobileNetV2 respectively. The best value being **2.52ms** for CardioQNet model. Making CardioQNet best candidate for real-time deployment.

To put this in perspective, a 60FPS video require 16.67ms inference time. Technically, the maximum achievable frame on VggNet16, ResNet50 and MobileNetV2, worked out as 32FPS,

51FPS, and 63FPS respectively. Although, an excess of 300FPS is possible on CardioQNet, this may be quite low in practice. Furthermore, the results of model performance are consistent across the four quality attributes with CardioQNet. This is not true for all the state-of-the-art models evaluated. This study recorded the lowest performance accuracy on ResNet50 with 90.32% followed by MobileNetV2 at 91.76% while CardioQNet exceeded all by 4.4% in performance accuracy. These figures support the evidence for implementing the ultralightweight model (CardioQNet) for any real-time quality assessment algorithm. Echocardiogram's assessment requires a real-time, fast-moving, and contractive anatomical properties. Therefore, any real-time assessment solution must exceed or match the ultrasound equipment frame rate or speed.

Table 5.5: Summary of comparison on similar quality assessment model's performance found in recent literatures.

Studies by Author	Abdi et al., (2017)	Luong et al., (2020)	Dong et al., (2020)	Current Study (2021)
Ultrasound Source	Philips and GE	Philips iE33 platform Philips S51 frequency 5–1 MHz	Shenzhen Maternal and Child Healthcare Hospital	GE Healthcare (Vivid.i) and Philips Healthcare (iE33 xMATRIX)
Study Population	-	3, 157 patients	-	11,262
Ground Truth annotations	2 expert annotations	1 level 3 echo cardiographer	1 radiological Expert	4 annotations each by AV 2 experts
Defined Quality Criteria	13	12	6	23 (Table I)
Assessment Methods	1 (Weighted Score) Method grossly insufficient of clinical advantage	1 (Weighted Score) Method grossly insufficient of clinical advantage	2 (Zoom, Gain) Focus on foetal ultrasound, adult not implemented	4 Specific Scores (Visibility, Clarity, Depth Gain & Foreshortening) Relevant to TTEs
Std Views considered	5 (A2C, A3C, A4C, $PSAX_A$, $PSAX_{PM}$)	9 (PLAX, A2C, A3C, A4C, PSAX-A, PSAX-M, PSAX-PM, SC4 & IVC)	Limited to (A4C only)	A4C & PLAX (Apical views relevant to TTE standards)
Input Size	200 x 200	Not specified	224 x 224	227 x 227
Sample Size	6,916	14,086	2,032	33, 784
Model Accuracy achieved	85.0%	87.0%	93.52%	**96.2%**

5.7 General Conclusion

This Chapter has presented the clinical significance and feasibility of developing an automated quality assessment in two-dimensional echocardiographic images. Also proved beyond reasonable doubt that weighted-average model for quality scores is insufficient for clinical use cases in a unified TTE workflow. Therefore, a quantitative method that is relevantly defined for cardiac image quality standard can thus provide useful and specific feedback for an operator guidance system as well as a valuable tool for research and clinical practice, first to act as arbiter reference to clinicians, secondly, to accelerate the learning curves for those in training. Also, it can provide specific information on the adequacy of the images obtained in retrospect, which could be universally relevant for a lifesaving procedure during the point of care or emergency services.

In this study, we have considered the prominent apical views (A4C, PLAX) as recommended for clinical use, for cardiac diagnosis, linear measurements, and quantification. Since this is the first time a comprehensive attribute of quality is thus defined and applied to cardiac frames, our work can be compared to existing works on quality assessment, (see Table 5.5.), in terms of use case functionality, accurate feasibility, clinical application, and assessment modality. Consideration should be given also, to domain-specific image quality and clinical advantage rather than model performance on the existing record in quality assessment.

This research has considered four distinctive image quality (attributes) objective standard and a novel method of accessing such attributes. Nevertheless, this study does not claim exhaustiveness in terms of quality criteria and define attributes, because the author is aware that different laboratories are at liberty to adopting what is considered best practice in their region or by legislation. Therefore, we intend to expand on these attributes in cardiac specimen characterization in chapter six.

Finally, the study used the annotation provided by two experts, a cardiologist who provided reference and supervision with an accredited annotator. The intra-observer variability can be examined in future, by obtaining additional annotations from either human experts or through reinforced learning to build more larger quality standard framework.

Chapter 6

Global Framework for Image Quality

6.1 Introduction

Transthoracic cardiac exam usually begins with patients' assessment, followed by ultrasound imaging protocol where high quality images at several views are acquired and quantified for linear and volumetric measurements. Although, A2C, A4C and PLAX have become clinically relevant to quantification and clinical measurements, the focus of this investigation and experiments in this chapter is on apical-four chamber (A4C) cardiac specimen. The methodology can be applied to other standard apical views.

The anatomy of cardiovascular specimens presents enormous complexity to subjective and objective functions in terms of dynamic features identification, measurements, and clinical quantifications. This is due to clinical protocol where cardiologists not only rely on still images but a fast-moving echocardiogram frame in real-time. However, plausible solutions using deep learning models have been widely demonstrated for cardiac view classification (G.N. Balaji, Subashini and Chidambaram, 2015); (Ornstein and Adam, 2021); (Azarmehr *et al.*, 2021), quality assessment (Dong *et al.*, 2020); (Luong *et al.*, 2021); (Yang *et al.*, 2018); (Labs, Zolgharni and Loo, 2021) and now it's been applied to achieve pseudo-labelling and characterization of cardiac samples in a semi-supervised model (Chen *et al.*, 2020); (Shiming Xiang, Feiping Nie, and Changshui Zhang, 2010).

6.1.1 Significant Impact of Experts' Annotation

The annotation of cardiac specimen is one of the significant step and requirement to build an objective model on, either for classification or regression task. The process involves that machine learning researcher specify at least a label (class label) or multiple labels for an action or an image prior to model development. This is true for all classification and segmentation tasks of deep learning algorithm. Unfortunately, annotation process is an expensive venture both in turnkey resources, time, and financial cost. For example, it took 14 months to complete manual annotations for 11,262 samples of PACS dataset. This requirement has inhibited many objective applications and still hampering the development of many echocardiography studies till now. Either for being too expensive, or require specialist input with associated professional costs, monotonous and time-consuming processes are some of the reasons. Therefore, there is a high stake in proposing far reaching quality attributes for global framework without the efforts, available dataset and strong will for extending quality attributes beyond the one discussed in chapter 5 of this research.

In my early research (chapters 4 and 5), the detailed of objective function used for classification and identification of 14 apical views was given. Moreover, the implementation of multivariate quality assessment method where four domain-specific quality attributes were defined, and prediction made per echocardiographic frame was documented. This chapter, seek to build on this and extend quality attributes beyond four domain-specific attributes in order to meet global framework or standard.

6.1.2 Automatic Annotation Possible?

An objective characterization of two-dimensional echocardiographic specimen is one of the potential approaches to define image critical features or extended feature of domain-specific elements of quality. In previous work on classification, we have applied an objective function to classify 14 different apical views and achieved state-of-the-art accuracy. However, that work was based on image label, typically involving human-powered tasks completed over a period of a year. For any larger dataset involving complex annotation and labelling, it would translate to much bigger efforts and higher coast with considerable length of time to implement. Considering the anatomy of cardiovascular specimens' potential complexities, the question is

how can a reinforced learning algorithm be applied to echocardiogram characterization and labelling?

As previously highlighted, object labelling or annotation are predetermined prior to model training, the labelling process is cumbersome and expensive in time and cost. Nevertheless, it's one process you cannot bypass especially with echocardiogram varying anatomical and pathological profiles.

6.2 Clinical Use Case

During transthoracic exam, several images are required to build a complete picture of patients' pathology and the summary of diagnosis; however, several specimens for multiple patients indicate the enormous amount of echo image that must be generated and stored. The task of manually assessing several hundreds of specimens prior to diagnosis constitute bottleneck in the administrative workflow.

An automated system that can objectively allow cardiologist filter-search and group specific specimen down to pathological and anatomical relevance (listed in Table 6.1) would constitute a priced tool, but such facility is certainly beyond the scope of apical view classification. Nevertheless, to the best of the author's knowledge, there exist no automated solution for a comprehensive pathological and anatomical feature detection which is capable of aiding cardiologists in rapid assessment, sorting and offering coherent domain-specific characterization or echocardiographic element-wise search. With this solution, experts can sweep through a class of images using seven (7) different clinical standards defined by domain-21 specific criteria in Table 6.1 as well as the quality class of any images before final diagnosis. This is believed to be a significant advancement in echocardiogram image and quality assessment.

6.2.1 Image's Element-Wise Search

To provide objective aid to cardiologists' workflow, there exist feasible solutions for apical view detection (Huang *et al.*, 2021); (Smistad *et al.*, 2020) which helps sort out different images depending on the image plane and views known as standard views. Nevertheless, this is a limited function and it's not different from searching up an image by textual descriptor

Several researchers have indicated the importance and impact of manually assessing hundreds of cardiac images by experts, to determine clinical relevance and diagnostic values (Nagata *et al.*, 2018); (Sprawls, 2014). Although, the focus in this study encapsulates the apical-four chamber images (A4C), as one of the recommended views for critical quantification and clinical measurement (Lang *et al.*, 2015); (Mitchell *et al.*, 2019). However, the principle can be applied to other apical standard views or any discrete label functions. Global feature characterization could allow rapid assessment and element-wise search of functional tissue, types and quality grading of valves, any anatomic, or pathological (within the scope of criteria) features present in each cardiac specimen.

The outcomes of this method could yield significant impact when expanded on other cardiac standard views. Therefore, defining a comprehensive objective property (characterization) for cardiac images would require a man-machine effort and many hours to implement.

It a common knowledge that machine learning could be applied to the most persistent task which is possible within the confine of human cognitive space. For element-wise search algorithm, objective characterization of cardiac samples needs to be extended beyond four quality attributes earlier implemented in chapter 5. A new measures of image quality need to be developed and evaluated against the defined 'global framework' properties of the two-dimensional echocardiographic images. Therefore, this study investigated semi-supervised learning to build intelligent algorithm for automatic annotation of cardiac frames.

6.3 Related Work

Objective image quality has been a priority of researchers for decades, but for two-dimensional echocardiograms, there are but limited publications achieved. This is due to general challenge in the availability of sufficient cardiac datasets, which unfortunately are highly personalised and require layers of ethical processing and approval for any research study. Some of the prominent earlier research on two-dimensional objective quality measurement using artificial intelligence (AI) models include four plausible works: namely, (Dong *et al.*, 2020); (Labs *et al.*, 2020); (Huang *et al.*, 2021); (Sassaroli *et al.*, 2019).

The authors presented an algorithm based on cardiac view detection and the approach successfully modelled the detection of apical chambers (view classification) of fourteen cardiac

standard views which include A4C echocardiography but admitted that the approach did not guarantee good performance when images consist of significant noise or the unavoidably images with low contrast pathologies. The result yields a contradiction detection result on images with low contrast-gain and high contrast-gain which reinforces an alternative investigation using spatiotemporal regressor for improved performance and prediction accuracy.

To the best of my knowledge, the most recent work on automated quality assessment, is the work by (Luong *et al.*, 2021); (Dong *et al.*, 2020); (Abdi, Luong, Tsang, Jue, *et al.*, 2017a) and (Huang *et al.*, 2021). However, Abdi's work was based on a regression model implemented to elicit automatic quality scoring. A single score was used to determine the strength of quality features in cardiac samples. Although Abdi's and Luong's defined numbers of objective criteria for two-dimensional echocardiograms, they did not include sufficient domain-specific criteria or searchable element-wise characteristics which renders their characterization less useful for cardiologist prioritize in clinical workflow.

Luong had admitted that universal reference standard of evaluation of echocardiographic image does not exist and that the scale of criteria used in many publications do not represent experts' standard assessment and consensus on 2D echocardiographic image quality (Luong *et al.*, 2021). Nevertheless, Luong work was based on average weighted score method (AWS), is the clever way to avoid annotation of cumbersome cardiac samples. Unfortunately, AWS is known to be insufficient of translatory advantages in clinical workflow. Average weighted score method of objective assessment is incapable of assisting ultrasound operators identify the aspect of image quality attributes that require immediate optimization.

Similarly, Dong et al., (2020)., proposed a generic quality control framework where characterization of A4C for foetal ultrasound resulted into zoom and image gain attributes. It was considered the first comprehensive quality control method with possibility for clinical consideration. Unfortunately, its application is limited to foetal ultrasound and lacks significant adequacy for generalisation of quality attributes required for wider use cases.

In this study, the definition of seven (7) domain specific attributes of apical-four image view (A4C) with focus and consideration for clinical and pathological relevance was achieved. It is believed that the defined attributes represent adequate characteristics of apical-four standard views. While other cardiac views would require additional or separate characterization, these

objective attributes could potentially be used to formulate clinical assessment method relevance for different apical views standard as may be required in different clinical lab protocols.

Nevertheless, in medical imaging literature, semi-supervised learning (SSL) methods which are associated with cardiac samples are rare. This is because classification of cardiac samples in the medical domain is subjected to cardiologists' visual assessment and interpretation and real-life scenario dataset require utmost and professional attention for clinical arbitration. Therefore, a shift into automated classification would be considered significant, radical, and historical. Moreover, to implement an automatic classification of cardiac samples would require a consensus of an expert's standard of reference which is non-existent till now. In the author's earlier work (Labs, Zolgharni and Loo, 2021) proposed a coherent standard for quality attributes and assessment methods using domain-specific scores can be further applied for sample characterization to aid clinicians' assessment during clinical exams and diagnostic process.

The implementation of SSL is broadly classified into pseudo-labelling, adversarial learning, consistency-based and graph-based methods. This section provides a review of relevant works from the application of pseudo-labelling on two-dimensional echocardiograms.

To the author's knowledge, there is no currently published work on objective characterization of two-dimensional echocardiograms using global attributes in the characterization of domain-specific criteria and assessment method in echocardiography modelling. This novel approach in echocardiograms image quality characterization is suitable for the implementation of specific quality assessment in retrospect or real-time optimisation.

6.4 Main Contributions

This work is distinguished from (Luong *et al.*, 2021); (Dong *et al.*, 2020); (Abdi, Luong, Tsang, Jue, *et al.*, 2017a) and (Huang *et al.*, 2021) as it's based on domain-specific characterization, featuring three-state objective assessment criteria (Table 6.1). The advantage of this novel method encapsulates the clinical and pathological characteristics of the two-dimensional cardiac image required for objective assessment and quality optimization. In view of the above, this research concludes that earlier works presented plausible contributions, but characterization were too limited for objective consideration for clinical workflow, and assessment criteria are

too limited in scope therefore, earlier works (without exception) are insufficient of translatory advantage in clinical practice. Our main contribution therefore can be summarised as follows:

- Analysis and annotation of independent dataset (EchoLAB) for domain-specific image characteristics covering 21 anatomical and pathological criteria
- Definition of clinical relevance, seven (7) objective characterization for global framework of 2D echocardiographic specimens in A4C standard views.
- Proposal of semi-supervised deep learning pipeline feasibly demonstrated for pseudo-label annotation process for 2D echocardiographic images
- Public release of complete annotated patient dataset to allow future studies and external validation of the new approach or methods

6.5 Methodology

This work is based on elicitation, clinical review, and response from a consortium of experts, cardiologist and researchers who collectively evaluated the independent dataset for objective characterization and prepared initial ground truth annotation for our semi-supervised learning model.

The advantage of this novel characterization method for two-dimensional cardiac image provides details of domain-specific, acceptable standard of objective elements of A4C for high repeatability factor and a non-biased clinical measurement.

This section provides details of the (EchoLAB) independent dataset description, followed by objective characterization of image quality with expert opinion scores. Finally, this chapter details the implementation of an objective model for automatic pseudo-labelling procedure using a robust combination of ensemble and semi-supervised learning model.

6.5.1 Dataset Preparation (EchoLAB)

The study population consisted of a random image sample of 27,230 apical-four echocardiography from earlier studies of patients with age ranges from 17 and 85 years,

recruited from patients who had purposely undergone echocardiography with Imperial College Healthcare NHS Trust. The acquisition of the images had been completed by experienced Echocardiographers using ultrasound equipment from GE and Philips manufacturers according to the standard protocols. Ethical approval was obtained (System identifier 243023) from the Health Regulatory Agency. Patient automated anonymisation was performed to remove the patient-identifiable information. To establish domain-specific evaluation of comprehensive anatomical details and expert ground truth annotation (Figure 6.1), the specimen was randomly selected for clinical characterisation, and referred to as 'EchoLAB' private dataset. Finally, up to 15% of total image specimens were annotated for model development purposes (ensemble semi-supervised pipeline) for a complete ground truth labelling process ready for evaluation and quality assessment method pipeline illustrated in Figure 6.2.

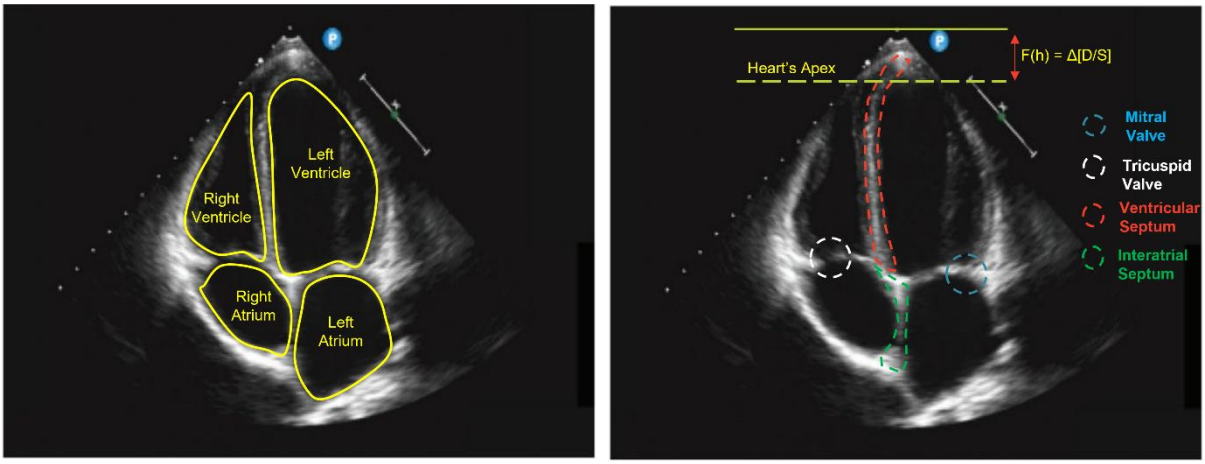


Figure 6.1: Apical four chamber images with annotation for chamber cavities (atrium, ventricles), Mitral valve, Tricuspid valve, interatrial and interventricular Septum.

6.5.2 Objective Characterization and Scoring

As the pace of change in AI and objective quality assessment continue to accelerate towards clinical use case acceptance, integration of such systems continues to pose significant challenges in the scope of acceptance and relevance in a unified workflow. The criteria defined for two-dimensional echocardiograms in many literatures appear to be too limited in scope and use case, consequently, none of the proposed assessment methods found acceptance with healthcare experts. But recent advancement in objective characterization has the potential to

move cardiac image quality assessment from its infancy where limited criteria is applied to characterise cardiac image, which is highly domain specific to a full-fledged, meaningful, and clinically relevant objective standard of assessment. The expertise and experience of member of the clinical consortium: cardiologist, clinician, sonographers, and deep learning researchers, were harnessed for the study which yielded seven (7) domain-specific characterization of two-dimensional echocardiogram in A4C as follows:

(a) Anatomical projection

(b) Intraventricular septum visibility

(c) Valves visibility

(d) Left ventricle clarity

(e) Image sector gain

(f) Foreshortening diastole

(g) Foreshortening systole

Each of characterization attributes could be used separately or in partial combination to satisfy respective requirements in clinical workflow or in the implementation of automated quality assessment pipeline. Details of domain specific characterization are summarised in Table 6.1 for A4C.

6.5.3 Domain-Specific Experts Annotation

Two experts independently provided seven (7) separate annotations for 4,000 echocardiogram specimens representing 15% of total randomly drawn cardiac image samples. To reduce the finite possibility of set of images falling outside the coverage of label space, I adopted additional sampling method by assessing each randomly drawn cardiac specimen and apply stratification to ensure fair representation of pathologies and abnormalities before ground truth scores (Q_{GT}) is assigned using twenty-four (21) opinion scores criteria summarised in Table 6.1. Since each characterization factor has a total score of 10, discretization was applied for score range of 4.6, 7.6, 10 as poor, average, and good quality respectively. This was done to prepare multivariate regression model for the training and inference process described in section 2.5.

Table 6.1: Domain-Specific Characterization of global attributes on Apical Four (A4C) Image Quality Assessment

Domain sp. Characterization		Objective Criteria	EOS	Assessment Description
Q1	CLINICAL PROJECTION	Apex wholly On-axis	6	Domain-specific requirement for on-axis projection of four apical chambers & intraventricular septum
		Apex partially On-Axis	3	
		Apex severe Off-Axis	1	
Q2	INTERATRIAL/ INTERVENTRICULAR SEPTUM VISIBILITY	Wholly visible	6	Clinical requirement for wall assessment, measurement, and strain quantifications
		Partially visible	3	
		Not visible	1	
Q3	AORTIC/MITRA VALVES VISIBILITY	Wholly Visible	6	Clinical requirement for valves assessment, measurement, and regurgitation quantification
		Partially visible	3	
		Neither visible	1	
Q4	LEFT-VENTRICULAR CLARITY	Highly Distinguishable	6	Clinical requirement for cavity assessment, strain measurement, and ejection fraction quantification
		Average Distinguishable	3	
		Poorly Distinguishable	1	
Q5	SECTOR DEPTH-GAIN	Adequate depth-gain	6	Anatomical and pathological requirements for additional decision-making process
		Poor depth-gain	3	
		Excessive depth-gain	1	
Q6	FORESHORTENING DIASTOLE	Minimal	0	Clinical requirement for assessment of ejection fraction, cardiomyopathy, crucial pathologies & apex visibility
		Average	-4	
		Excessive	-6	
Q7	FORESHORTENING SYSTOLE	Minimal	0	Clinical requirement for assessment of ejection fraction, cardiomyopathy, and clinical quantification.
		Average	-4	
		Excessive	-6	

6.5.4 Semi-Supervised Boosting Model

The choice of boosting ensemble strategy implemented in this project incorporate semi-supervise algorithm aimed at performing automatic annotation of unlabelled samples in the characterization of cardiac specimens. Boosting strategy can combine weak classifiers to obtain the effect of a strong classifier model and improve model performance and accuracy.

The implementation consists of four classifiers of a known performance (since they were evaluated in earlier work) to offer a high degree of generalizability for the implementation of cardiac characterization and auto-labelling process. This ensemble architecture illustrated in (Figure 6.3), is a choice for the implementation of pseudo-labelling algorithm for two major reasons: combination of known high performance state-of-the-art models and CardioQNet model which is considered as new ultra-light weight architecture and to ensure high-performance accuracy with comparable convergence and correlation.

The block diagram of the boosting ensemble model illustrated in (Figure 6.2) is the core architecture used for the classification of global attributes on cardiac specimens. Due to the inherent complexity in fast moving cardiac frames which requires both space-time components, individual classifiers may not provide satisfactory results, as the echocardiographic data is highly heterogeneous and noisy (Zhang *et al.*, 2016); (Zhang *et al.*, 2021).

Rather than finding the best single model classifier, consideration for multiple classifiers is applied through boosting ensemble learning methods to achieve a robust layered network architecture, stronger generalizability and obtain better classification accuracy.

Echocardiogram's specimen mostly dominated by noise, (i.e., low signal-to-noise ratio), artefact and hardware induced noise, ensemble architectures provide majority voting which is derived from stacking the optimised model along with two notable state-of-the-art models for semi-supervised ensemble learning.

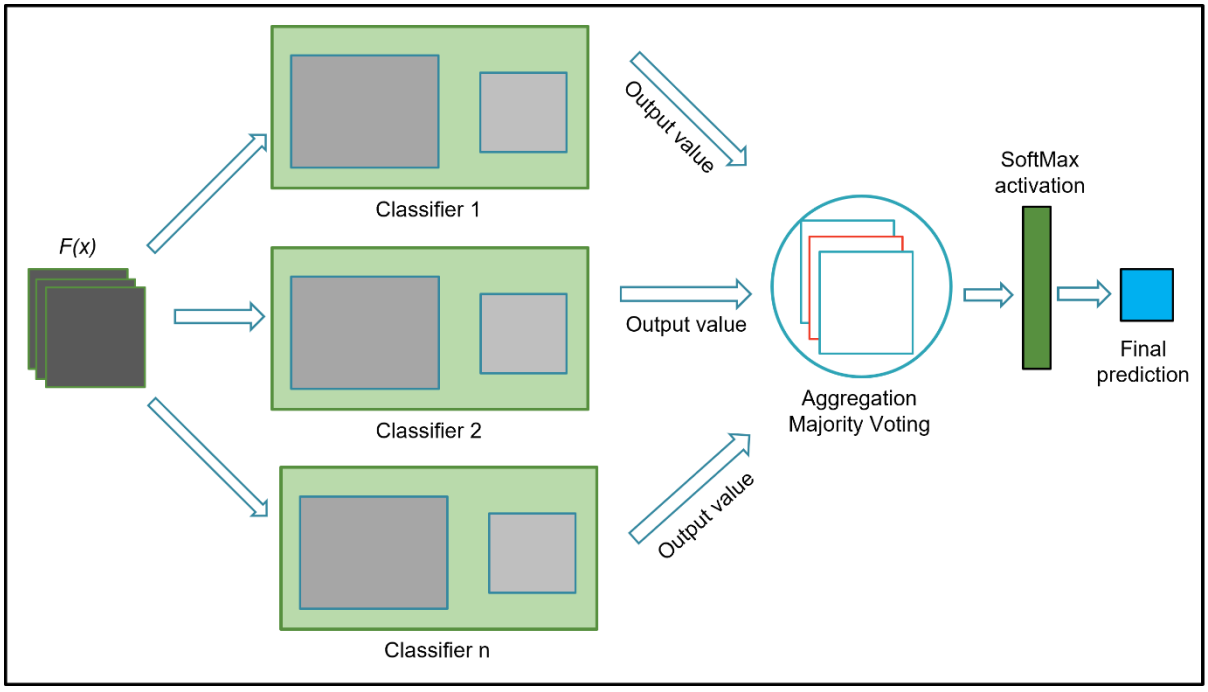


Figure 6.2: Illustrating the semi-supervised learning using ensemble architectures (VGG16, ResNet50, MobileNetV2 and 'CardioQNet') with majority voting algorithm Adapted for self-supervised Architecture on 2D Characterization Pipeline.

The final output value which is a class label Y_k of the ensemble is computed using an averaging voting scheme common with regression models. While in classification mode, priority is given to the majority votes prediction when a specific class labels have more than half of the classifiers votes as illustrated in equation (4.2). There is a possible instance when prediction is

considered void if class votes do not meet the criteria of majority votes, in such case, there is no prediction and the model retry the iteration process until class votes meet the prediction criteria. In a multi-class scenario, where Q is referred to as pre-trained classifiers, while each classifier yields specific prediction per input data, a total of N classifiers, the ensemble decision will be correct if at least [Q/N + 1] classifiers choose a correct label.

Generally, semi-supervised models offer dual advantages in discrete label and performance spaces. It potentially can increase a model's ability to generalise on similar data with exceedingly high performance as well as providing correct predictions for unknown, unlabelled samples. Semi-supervised algorithm involves two major iterations: training of a supervised model using few categorised labelled subsets of the training samples and inferring pseudo-labels for the unlabelled portion of the test data. Consequently, the model re-combines newly auto-labelled samples and initially labelled data and recalculates development samples for subsequent training iterations in a case of k-fold test sets. It is effective and works well for cases with low-density separation of categorical labelled classes. However, it can also reinforce errors learnt by the initial model, which is why I implemented spot checking procedures along the iteration to effectively correct potential errors. Worthy to note that a handful of researchers have applied pseudo-labelling to cardiac samples due to the reason given above.

A. Data Splitting and Data Augmentation

In this experiments, labelled data pool was at minimum 15 percent, (supplied with limited annotated specimens) compared to unlabelled specimens of 23,149 and this may be insufficient to provide adequate representation in data search space required for deep learning models. Medical ultrasound images bear significant pathological and anatomical features which are critical for clinical measurement, classification, and diagnosis. Therefore, we adopted a k-fold validation for the ensemble learning portion and semi-supervised inference on the unlabelled pool while we split the labelled samples L^F into L^T and L^V (training and validation) in 80:20 ratio for the purpose of training. Also, data augmentation was applied to provide stability and improve model learning accuracy. Data augmentation on training samples increases the diversity of the training set using realistic transformations and creating new training samples from original samples without changing the scope of label space. For each round of 5-fold inference, after label propagation had taken place, training, and validation $L^T + L^V$ samples grows until last inference iteration when X^U pools equals zero.

B. Model Training

The pipeline task is to learn with the limited augmented labelled samples L^F, then utilise the abundant pool of unlabelled samples X^U to improve generalisation of 2D_echocardiogram characterization and pseudo-labelling. The entire pipeline process and iteration is illustrated in (Figure 6.3). The classification target associated with X^L is denoted as Y^L, where Y^L is associated with only one of many labels (i.e., maximum configuration of 3 unique classes). For a given training dataset, we have L^F labelled samples $\{(x_i,\ y_i)^L\}$ = $1L^F$ and X^U unlabelled samples $\{x_i\}^U$ = $1X^U$ with X^U not equal to L^F. Model training starts with labelled samples $(X^L,\ Y^L)$ using $(x, y)\ where\ x\ \epsilon\ (X^L,\ Y^L)\ and\ y$ being the corresponding categorical label.

The model training implementation considered a batch of input sample x, where x is taken as a single 2D image or an array of 2D slices and applied random transformation of 10 degrees rotation and translation of 5% each during training phase, yielding an expanded dataset to obtain prediction of category label $\hat{y}$ which is evaluated using the mean absolute error (Grandini, Bagli and Visani, 2020), and error rate (combined error type I and II) metrics (K B and J, 2020) as detailed in Table 6.3.

This process is followed by inference on unlabelled samples X^U, while each prediction is evaluated against the sigmoid activation curve between [>0.95 and < 0.05] and added to the initial pool of labelled samples L^F while deducting $\{x_i\}$i, from the unlabelled samples X^U.

Since k-fold validation method was implemented, subsequent inference iteration can be paused momentarily to allow additional inspection in the process known as spot-checking procedure, where predicted label is evaluated against the label samples to allow addition or prevent addition of label samples. Consequently, the inference iteration process continues yielding a deduction in the unlabelled samples X^U until the last unlabelled batch samples are predicted and added to the pool of annotated samples.

C. Spot Checking

Usually, the assessment of hearts function using 2D echocardiogram requires special attention and so does the opinion on image quality. Most objective models *f(x)* does require large samples to train on, unfortunately, a self-learning model in contrast, usually must consider its training on a pool of limited labelled samples. Consequently, accuracy and subsequent inference iteration are usually not at their best measure.

Indeed, the result of several experimental values obtained and analysed on inference process have proved that propagation error in label space can only be reduced rather than eliminated. This is because, two different images may assume a similar RMS value or yield similar learned features in some dataset like 2D echocardiogram.

One of my experimental investigations had consider the use of basic image dataset containing Bees, Cats and Dogs; this was applied on the semi-supervised learning (SSL) architecture described in this section. The model accuracy was great, but few images still manage to slip through the wrong label classification despite.

However, in a bid to minimise deviation error in the entire self-learning and prevent propagation of classification error in the subsequence process, a new approach that would reinforce confidence in model classification and inference loop must be considered. An example of this is applied in manufacturing plant where each component and system integration are passed through quality assurance using spot-checking methodology.

The author therefore proposes a spot-checking process in the inference loop, where model's iteration is paused momentarily to evaluate inference threshold for all the incoming new labels. This usually can happen after the inference loop is initiated. At this moment, all new incoming samples with respective new target values (pseudo-labelled) are automatically visualised (on-the-fly), for evaluation. The user can either proceed with automatic addition of new label samples into the original label space or re-adjust the threshold values to restart the process as deem fits. The original threshold values for this model encompasses three categories in the label space, ranging from 0 < score <= 0.4500; 0.4501 score <= 0.7500; 0.7501 < score <= 0.9999 for class label poor, average, good quality respectively.

Since the k-fold validation method is implemented for the training loop, the inference loop is consequently grouped accordingly. Each inference group occurs after model training and validation phases and completes the SSL pseudo-labelling cycle. i.e., the auto-annotated samples as illustrated in Figure 6.3.

This means that each pseudo-labelling cycle is isolated before the next training phase is initiated. Therefore, halting the iteration (which can be released without terminating the user session), is essential for on-the-fly inspection of newly labelled samples. Consequently, an erroneously classified samples can then be isolated or new threshold values assigned to accommodate model deviation on the group samples. This method was implemented and has

yielded significant results. Furthermore, the novel method could provide additional layer of quality measure with a view of reducing subsequent error in the propagated label space.

The impact of spot checking on label space was investigated on test dataset (i.e., Bess, Cats, Dogs). The inference accuracy recorded on known label space yielded 82.13% without the spot-checking process and 89.30% with spot checking, respectively. This equates to 7.17% improvement on model generalizability with the implementation of spot-checking on this specific dataset. For 2D echocardiogram, the improvement could be very significant due to the varying frame contents in pathological and anatomical features.

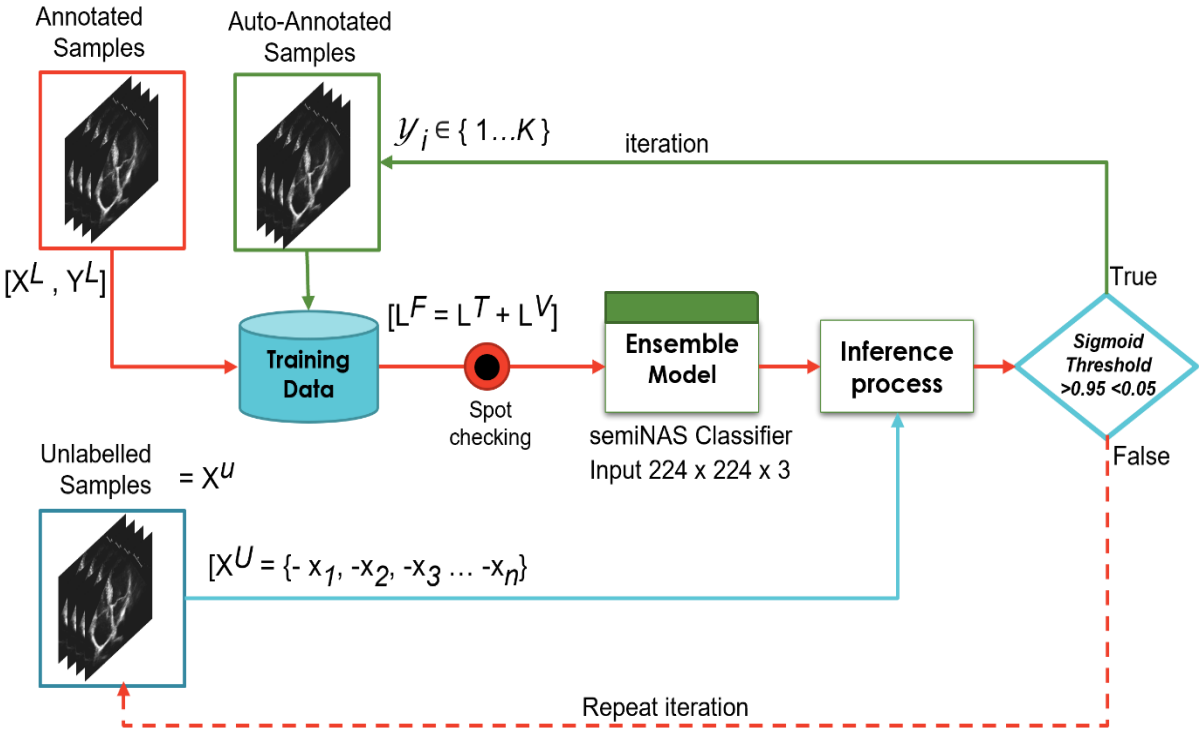


Figure 6.3: Illustration of a robust semi-supervised pipeline implemented for pseudo-labelling process in objective characterization of 2D-echocardiogram samples.

6.6 Evaluation Metrics – Ensemble Model

This study employed several metrics in the evaluation of ensemble model performance; these include average voting method, classification accuracy, inference latency in (ms) and numbers of trainable parameters for the use of system memory.

- **Voting Method:** In practical terms, deep convolution neural classifiers are generally correlated (Zhang *et al.*, 2021) especially when they are trained on the same data, however, it is not out of place to assume for an ensemble model that each classifier's output is statistically independent of each other. Therefore, the probability that each classifier in ensemble stacking technique, makes a correct prediction can thus be summed up using binomial distributions stated in equation (4.1):

$$P_x = \sum_{x=[\frac{T}{n}+1]}^{T} \binom{T}{x} p^x \, (1-q)^{T-x} \qquad (4.1)$$

Where *P* is the binomial probability of mutually exclusive outcome of either success p or failure q, *x* is the number of times for a specific outcome within *T* number of trials within $\binom{T}{x}$ number of combinations. Voting method assesses the collective classifier prediction for class labels is given in equation 4.2.

$$Y_k = \sum_{i=1}^{N} Q_i^k(x) > 0.5 \sum_{k=1}^{L} \sum_{i=1}^{N} Q_i^k(x) \qquad (4.2)$$

Where *N* is the maximum discrete classifier in the list $\{Q_1 \ldots Q_N\}$ are combined to predict Y_k class label from set of *L* possible label space $\{Y_1 \ldots Y_L\}$ yielding a rejection or no prediction unless class label Y_k receives more than half of the total votes.

- **Model Accuracy:** The model performance was evaluated as mean absolute error (MAE) for the collective models in the ensemble classifier. The regressor voting ensemble fits three model regressors on the unlabelled samples and averages each individual prediction to form as final prediction. Model accuracy is evaluated in equation 4.3 as:

$$Acc_{\mu} = 1 - \left(\frac{\sum_{i=0}^{n} |Q_{GTi} - Q_{PDi}|}{n}\right) * 100 \qquad (4.3)$$

- **Training Time:** Training times are affected by the number of convolution layers, number of its training parameters, image spatial size and hardware processing power. Excessive number of trainable parameters in each network can create redundancy and prolong training time per epoch or convergence. All computations were carried out on GPU GeForce GTX 970 chipset's Maxwell architecture and featuring 4GB RAM coupled to 1,664 CUDA cores. A batch of 6 samples were favoured for each training phase for 50 epochs per k-folds. The ensemble model takes a total of 8hrs GPU time for each characterization. This is repeated for each of the seven (7) global characteristics on apical four chamber (A4C) image specimens.

6.7 Results and Discussion

Similarly, to Wen et al. (2019) and Baker et al. (2017), the quality of the data fit via model accuracy per k-fold was assessed on ensemble training in direct sequenced with pseudo-labelling process and spot-checking procedure to minimise error propagation in subsequent training phases. It's apparent that model performance on generalisation improves with increased pool of samples as indicated in Table 6.2, Table 6.3, and Figure 6.4 while model error grows less and less with improved generalization per folds (Table 6.4).

Table 6.2: Summary of ensemble semi-supervised model accuracy on pseudo-labelling inference for A4C cardiac specimens.

Apical Four (A4C) Global Characteristics	**Fold-1 (10%)**	**Fold-2 (18%)**	**Fold-3 (22%)**	**Fold-4 (24%)**	**Fold-5 (26%)**	**Mean Accuracy**
ON-AXIS PROJECTION	98.08	98.37	98.73	99.06	99.24	**98.70**
SEPTUM VISIBILITY	98.81	98.71	98.80	99.10	99.14	**98.91**
VALVES VISIBILITY	98.30	98.17	98.77	98.87	99.09	**98.64**
LV CLARITY	98.48	98.43	98.88	99.04	99.05	**98.78**
DEPTH-GAIN	97.83	98.33	98.49	98.87	99.07	**98.52**
FORESHORT DIASTOLE	97.45	98.29	98.64	99.05	99.08	**98.50**
FORESHORT SYSTOLE	97.53	98.27	98.86	99.06	99.13	**98.60**

Table 6.3: Pseudo-labelling standard error distribution across k-Fold samples expressed in Mean ± SE for sample characterization of Apical-four (A4C) global attributes. SE taken as SD/Sqrt (fold test samples).

[Mean ± SE]	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5
ON-AXIS PROJECTION	0.1105 ± 0.0141	0.0943 ± 0.0092	0.0754 ± 0.0077	0.0625 ± 0.0059	0.0509 ± 0.0046
SEPTUM VISIBILITY	0.0765 ± 0.0124	0.0746 ± 0.0097	0.0823 ± 0.0089	0.0500 ± 0.0054	0.0471 ± 0.0045
VALVES VISIBILITY	0.1105 ± 0.0174	0.0903 ± 0.0108	0.0673 ± 0.0063	0.0676 ± 0.0043	0.0558 ± 0.0042
LEFT-VENT CLARITY	0.0949 ± 0.0134	0.1018 ± 0.0098	0.0721 ± 0.0070	0.0533 ± 0.0046	0.0599 ± 0.0041
SECTOR DEPTH-GAIN	0.1237 ± 0.0103	0.1009 ± 0.0092	0.0876 ± 0.0067	0.0641 ± 0.0052	0.0482 ± 0.0044
FORESHORT DIASTOLE	0.1558 ± 0.0172	0.1130 ± 0.0082	0.0936 ± 0.0086	0.0573 ± 0.0053	0.0483 ± 0.0041
FORESHORT SYSTOLE	0.1691 ± 0.0190	0.0987 ± 0.0088	0.0683 ± 0.0061	0.0502 ± 0.0044	0.0540 ± 0.0044

Table 6.4: Showing the mean and standard distribution on global attributes of apical-four (A4C) samples, across 7 attributes of domain-specific quality, expressed in Mean ± SD for each three-quality class characterization.

	On-Axis Projection	Septum Visibility	Valves Visibility	LV Clarity
Q1	0.0706 ± 0.1095	0.0646 ± 0.1114	0.0801 ± 0.1226	0.0858 ± 0.1158
Q2	0.0866 ± 0.1159	0.0561 ± 0.0973	0.0816 ± 0.1083	0.0667 ± 0.0879
Q3	0.0789 ± 0.1169	0.0776 ± 0.1340	0.0752 ± 0.1126	0.0766 ± 0.1111
	Image DepthGain	**Diastole Foreshorten**	**Systole Foreshorten**	
Q1	0.0775 ± 0.0975	0.0982 ± 0.1333	0.0865 ± 0.1125	
Q2	0.0881 ± 0.1038	0.0915 ± 0.1036	0.0836 ± 0.1037	
Q3	0.0891 ± 0.1019	0.0912 ± 0.1080	0.0940 ± 0.1129	

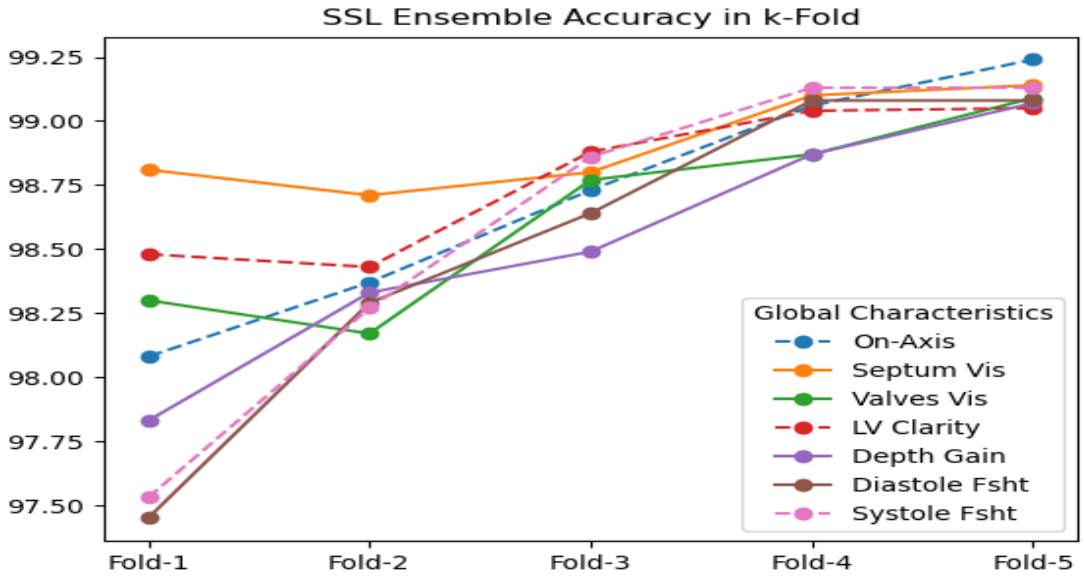


Figure 6.4: Comparison of process accuracy for different global characterization attributes of cardiac specimens in apical-four chamber image using ensemble with semi-supervised learning. Plots indicated consistent model's generalizability with high performance and model accuracy which is congruent in proportion to increased data samples across the k-folds.

It is clear from the result of model accuracy for each respective characterization standard attributes that on-axis attributes top the list in best performance 99.24% on the last k-fold and LV- clarity attributes with 99.05%. The model ability to generalize has not been diminished for any of the considered attributed.

For valve-visibility, has shown a decline in model performance at the second phase of model inference, with 98.17% compared to other attributes on the list. This feature is assumed to be one of the critical aspects of myocardium, they are clinical essential for performance assessment of cardiomyopathy in hospital patients. Model performance was low on limited sample space but improved with additional new labels in the auto label pool. This means that echocardiogram quality classification would require careful and decent volume of data space for any meaningful generalization and results. This is the point where incremental k-fold validation becomes relevant. It assures a minimal error propagation in the adjacent folds and allow the assessment of what volume of data pool could be injurious or impair the classification or regression tasks.

6.8 Conclusion

The inherent constraints in deep learning development and algorithm are the problem with dataset availability and correspondent annotations otherwise known as ground truth. This is especially true in echocardiography domains, that lack of annotated data and the general data restriction on personalised diagnostic medical details are inherent problem facing AI researchers. Saripalli *et al.,* (2020) had been demonstrated that medical image data annotation is a critical feature to objective development and the requirement for experts' skills in echocardiography domain is a priced endeavour. One solution proposed in this research is reinforced learning using SSL. Although this has been widely used in relation extraction of photographic images, little has been demonstrated on 2D echocardiograms. Therefore, healthcare application using reinforced learning are, up till now very sparce. The is specific to medical imaging, because they are complex and require domain expertise for interpretation and annotations. The novel solution proposed, thus adequately provided a lead way, where an unsupervised algorithm latched on boosting learning to generalize on limited label space and produce automatic annotation for majority of unlabelled samples in semi-supervised boosting pipeline.

In contrast to individual classifiers as a semi-supervised model, it is evident that semi-supervised learning in ensemble mode has the logical and technical potential for a robust

pipeline process in the classification of medical image samples and has thus yielded very satisfying results as indicated by the global characteristic curve in Figure 6.4.

Majority of semi-supervised models in literature have provided comparable results but are adapted for classification of high-fidelity or non-medical images than adapted for 2D echocardiograms. This is because medical images are highly personalised and inherently complex for classification models. Nevertheless, the primary aim for semi-supervised models have been to improve and optimise the generalizability of model's performance but this study has extended the application to pseudo-labelling of more complex dataset for apical four chamber images for the purpose of characterization of cardiac global features.

Each fold featured a finite pool of samples which increased when new pseudo-labelled samples were added from unlabelled samples space. The mean standard error reduces as data size increases with each training routine as shown in Table 6.3 and is consistent across the five-fold and the objective global attributes.

While Table 6.4 showing the precision of measurement in the mean standard deviation relating to discretized class category across the seven attributes, low error rates on the label space indicated a good model fit for the unlabelled samples and the ability to distinguish and propagate minimal error to achieve the high-level accuracy. The inference error was drastically minimised by the implementation of spot-checking algorithm. This provided a layer of confidence in auto-labelling process and help reduce error furth in the adjacent k-fold validation process and in label propagation.

Therefore, ensemble learning in a semi-supervised model could be useful in real-time classification of cardiac samples and provide a reliable process for medical image classification.

The significant of this solution provides essential clinical use case for cardiologist and clinicians during echo exams or unified workflow. We find this novel approach to be data efficient, scalable, and generalizable for echocardiogram annotation task, which ordinarily are very costly for huge volume of global quality framework and unified application in the Healthcare sector.

Chapter 7

Feasibility for Real-Time Optimization

7.1 Introduction

A two-dimensional (2D) echocardiography has become a widely accepted standard for assessing cardiac functions (‘Rudski et al. - 2010), (Mitchell *et al.*, 2019) and plays a vital role in detecting heart abnormalities even during medical emergencies. The most important consideration however, is achieving optimum image quality which is essential to a reliable quantification of the left ventricle (LV) systolic functions. In the past decades, several efforts have been geared towards echocardiographic image quality and assessment methods. Unfortunately, many of such solution did not gain translatory advantage in clinical practice, either because the adopted criteria were insignificant or did not meet experts’ expectation in clinical assessment.

In our earlier publication (Labs *et al.*, 2020), we enumerated the significance of domain-specific criteria for grading echocardiographic image quality in apical-four (A4C) and parasternal long axis (PLAX). Also demonstrated the feasibility of a machine learning algorithm to quantify and assess echocardiogram image quality in real-time or in retrospect. The clinical application of such methods employed a domain-specific definition of quality elements that effectively constitute a standard of objectivity when adopted by clinicians. It lays the foundation for the possibility of a real-time echocardiogram optimization of image quality, assessment methods, reliable measurement, repeatability in quantification, objective interpretation, and better diagnosis.

Significance of Real-Time Optimization

Since image quality is inherently subjective with strong dependence on operators' experience level and acquisition skill, this means that a less experience ultrasound operators is prone to acquiring poor quality echo images which can impair diagnostic accuracy (Liao, Z. *et al* 2019). Therefore, to implement an objective system where operators are guided through acquisition process will constitute a tool of clinical significance to cardiac reliability and diagnosis. Such system tool will have to rely on a set of domain-specific quality standard and feedback mechanism to provide objective assessment prior to clinical measurement and quantifications.

In the clinical practices, preference for cardiac image in apical-two (A2C), A4C and PLAX standard views have gained recognition from healthcare experts. Therefore, diagnostic outcomes are derived from the linear measurements and volumetric quantifications of these standard views. Nevertheless, obtaining any of such complex views during medical emergencies or where unwell patients are hardly oriented in dorsal decubitus could present additional layer of complexity. Such instance is inherent in medical practice and potentially exacerbates operators' skill in the acquisition of optimum image quality. The possibility of achieving or acquiring an optimum image quality becomes less and less when dealing with multiple patients in a prolonged TTE exams.

Moreover, echocardiography image quality can be affected during TTE outside the laboratory setup. For example, at point of care or at frontline medical deployment. Research has shown that clinicians are prone to obtaining a sub-optimal image quality of limited value when larger volumes of diagnostic data are generated, especially during medical emergencies or pandemic (Luong, C *et al.,* 2020). However, to overcome these inherent limitations, a real-time optimization algorithm which can provide quantitative information on the adequacy of the images while allowing operator to improve on the specific element of quality is proposed.

The impact of such system on TTE examination could accelerate the learning curve of ultrasound operators and provide significance to accurate diagnostic of pulmonary disease and cardiac infarction (Nagata *et al.*, 2018), (Labs et al., 2021).

Clinical Use Case

A unified transthoracic (TTE) exam workflow can be illustrated in (Figure 7.1), consisting of six (6) major processes of which are classified into three phases namely: Acquisition and Optimization, Functional Analysis, Interpretation and Diagnosis phases.

Typical clinical workflow usually begins with manual echo scan of patient who is undergoing transthoracic exam, standard echocardiogram in either of the 14 apical views are generated by the operator. This is known as the image acquisition process and is regarded as the critical phase in a unified workflow where reliability, variability and accurate quantifications are determined correctly or otherwise.

This phase is followed by view identification since different views are usually required to build an adequate examination profile. Consequently, several images are thus generated for each patient, but each quality is subjected to operators' subjective assessment and are entrusted to produce optimum images. This however does not happen due to the one or combinations of reasons given above.

Before a diagnostic decision is made, the images must pass through phase detection where the start of systole and diastole is accurately identified. The clinical impact of optimum image quality can prevent the propagation of analytical error in the adjacent process. Acquiring optimum image quality is paramount to consistency and reliability of phase detection process, segmentation, measurement quantification and indeed diagnosis of myocardial infarction.

This should be noted that a poor image quality yields poor outcomes in the successive process of the unified workflow. Most importantly, good image quality produces excellent outcome and ease of the processing overhead of subsequent phases in the workflow.

Therefore, to enforce clinical protocol and quality compliance, an objective quality control process is necessary. An objective image quality assessment is thus required during acquisition phase, providing specific feedback for image adequacy, and allowing qualitative optimization.

Optimum image quality is also paramount to the linear measurements of the left ventricle (LV) systolic functions that provide major clues to a healthy heart which triggers a corresponding diagnostic response for patient care.

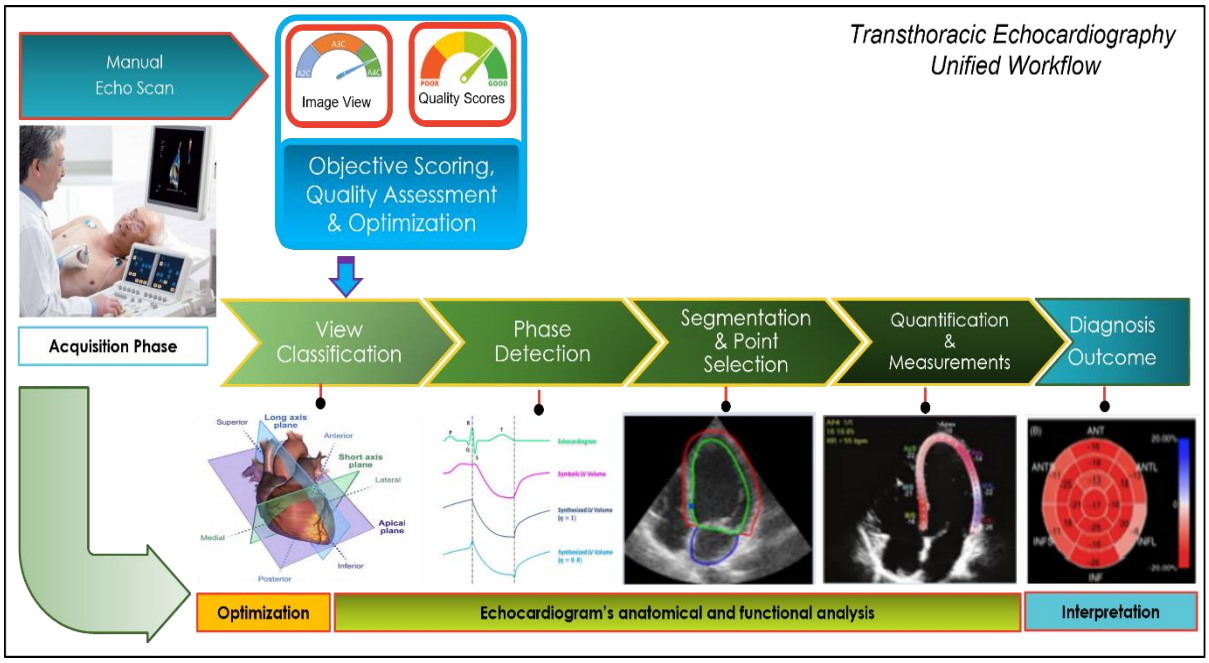


Figure 7.1: Illustrating unified workflow that details significant automated processes from acquisition to diagnosis. Optimization should occur before the analytical stage to guarantee optimal image quality before clinical measurement and quantifications.

Although the inference data reported in the previous studies were adequate for real-time optimization, yet the utility of such systems in the clinical practice would be limited. This is because a single score to denote overall image quality score is provided by the models.

Such scoring system is grossly incapable of guiding operators to which aspect of quality is lacking in the overall assessment and optimization protocol. Hence, practical deployability of such a system is limited to experimental demonstration instead of translatory advantage. If employed as part of a real-time optimization or operator guidance system, the operator is provided with no clues as to why the image is being tagged as low quality, and how to improve on specific element of quality in the acquired image would become time-consuming and guess work.

A practical quality optimization report should contain such information which possibly breaks down the specific attributes of image or frame quality, and updates on the quality while optimization progresses is the novelty in this chapter. Therefore, the optimization utility provides clinical insight on objective standard per quality element of ultrasound's image during acquisition phase. Also, a method of accessing the specific quality attributes for the purpose of optimization of image quality before quantifications.

7.2 Related Work

Cardiac images vary significantly from patient to patients, and it is difficult to define an image with perfect quality compared to non-medical imaging pathology. Consequently, it is considered impracticable to define a reference image with which can be measured by calculating its deviation (Wang and Bovik, 2002); (Wang *et al.*, 2004).

Similarly, Zhang *et al.*, (2016) universal quality index approach which largely focused on distortion of compression, with some implementing machine learning algorithms using a random or structural noise level to evaluate image quality, provided an overall quality score that cannot be implemented for real-time optimization. The specific element of image quality must be explicitly defined, to provide scores visualisation for ultrasound operator' during image acquisition phase.

Furthermore, impact of image quality on echocardiographic measurements and global longitudinal strain has once been demonstrated by Nagata *et al.*, (2018), but failed to address the explicit differentiation on specific elements that constitute quality with respect to image acquisition and quantification. Therefore, optimization cannot be implemented on average weighted value.

Although Abdi *et al.*, (2017) work reported plausible outcomes, the assessment method was technically insufficient for clinical translation. This is because the defined quality features are limited and do not represent experts' global characteristics for cardiac diagnosis using 2D echocardiographic images.

Also was Luong et al. (Luong *et al.*, 2021), defined twelve criteria to grade each of the nine apical standard views, but the assessment method and scores do not represent cardiologists' conventional assessment in practice hence, could not be used for real-time optimization.

The most recent study, to the best of our knowledge, was Dong et al, (Dong *et al.*, 2020) represents the most recent study on objective quality assessment, however the study was limited to apical four-chamber plane (A4C) in fetal cardiology. Dong *et al.*, (2020) did not include A2C and PLAX view, the recommended imaging plane for LV quantification and diagnosis.

Hence, new measures of image quality were proposed by (Labs et al., 2020, 2021) based on the global properties of the echocardiographic images that match the pathological inferences and clinical recommendation. This serves the basis for optimization of ultrasound echo image acquisition and real-time optimization.

7.3 Main Contributions

As noted in the earlier chapter, interpreting the results of the proposed architectures in the literature is not straightforward. This is because a direct comparison of the models' performance and clinical use case would require access to the same patient dataset and clinical application. At present, no echocardiography dataset and the corresponding annotations for the image quality assessment is publicly available or a unified pipeline where real-time optimization of image acquisition has been established.

Therefore, this chapter enumerates the performance of novel deep learning algorithm established for automated image quality optimization of ultrasound images. The solution would integrate an independent echocardiography dataset (PACS-2) which would be made available on request at IntSav repository.

In the view of the above, the main contributions of this research can be summarised as follows:

- Preparation of (40,000) large independent cardiac dataset consisting of A2C, A4C, PLAX apical views for quality assessment and benchmarking standard.
- Release of dataset including experts ground truth annotations for apical visibility, chamber clarity, depth-gain, foreshortening to the public domain.
- Fully optimised deep learning multivariate pipeline that simultaneously predicts four independent scores of quality attributes and view classification from echo sequences.
- Novel method on real-time access of 4 specific quality attributes to aid optimum image acquisition and reliable clinical quantifications.
- Provide evaluation for real-time application pipeline suitable for operator feedback for data acquisition, and real-time optimization for A2C, A4C, and PLAX cardiac standard views.

7.4 Methodology

This section provides new insight to objective attributes of image quality, an overview of the key elements in objective assessment standards. Unlike previous related works, we provide the most comprehensive attributes of image quality as reviewed by cardiologists, and real-time feedback for operators' guides to obtain optimum image quality for quantification and clinical measurements.

7.4.1 Private Dataset PACS-2

For clinical assessment of myocardial functions, cardiologists place a magnitude of importance on A4C and PLAX for volumetric quantification and linear measurements. Even though cardiac data are highly personalised with healthcare legislations, it's rare to have a substantial large number of cardiac datasets in public domain. But for the purpose of this research, an ethical approval (ref. 243023) was sought from the UK's Health Regulatory Agency.

This study is based on randomly selected patients' dataset consisting of 6,216 (A2C frames), 15,476 (A4C frames) and 18,308 (PLAX frames) from patients who had earlier undergone echocardiography TTE with St Mary's Hospital, Private NHS Trust which was purposely acquired for this study. These cardiac images acquired in both standards consist of end systole (ES) and end diastole (ED) frames were completed by experienced echocardiographers using high-end ultrasound GE Healthcare (Vivid.i) and Philips Healthcare (iE33 xMATRIX) equipment. Standard protocol in data protection act (2018) allows for the removal of all patient-identifiable information from DICOM-formatted videos before data analysis and applicable studies. Three frames were randomly drawn from each cine loop video and split into training (32,000 frames), and testing (8,000 frames) sub-datasets in 80:20 ratios. Figure 7.3., summarises model's dataset distributions with three categorical labels derived from expert's quality scores previously assigned to each frame; range of zero (0) to 4.5 as poor quality, 4.6 to 6.5 as average quality, and 6.7 to 9.9 as good (optimum) quality, respectively.

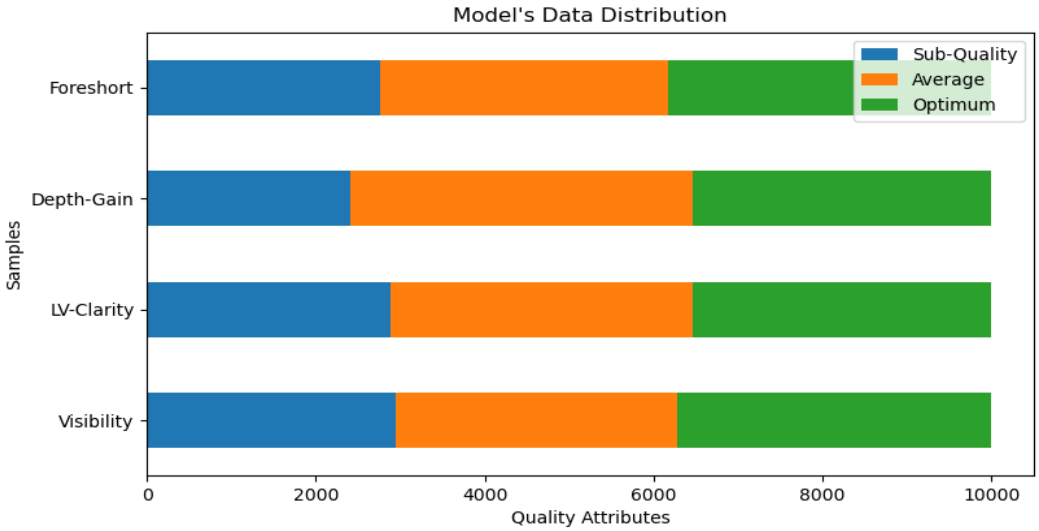


Figure 7.2: Data distribution for total cardiac images used for model development consist of 40,000 extracted frames of A2C, A4C and PLAX images, with three quality-levels: suboptimal quality, average quality, and optimal quality, respectively.

7.4.2 Ground Truth Annotations

Each of the cardiac echo cine (A2C, A4C and PLAX views) were studied for anatomical characteristics that relate to experts' clinical reviews and projection characteristics that relate to 2D echocardiograms. These features were visually analysed and were defined by 23 criteria listed in (Table 7.1) and referred to as 'PACS-1' dataset. Consequently, four specific quality attributes were identified and established for real-time assessment and optimization.

Table 7.1: Score criteria for Apical-Two Chamber (A2C), Apical-Four Chamber (A4C) and Parasternal Long Axis (PLAX) Quality attributes and Ground Truth

A2C - A4C		**PLAX**	
Assessed Element per Attributes	**Maximum Manual Score awarded**	**Assessed Element per Attributes**	**Maximum Manual Score awarded**
APICAL VISIBILITY:		**APICAL VISIBILITY:**	
Correct Axis, Apex visible	6	Left Ventricle (LV) Visible	**5**
Anterior wall / IAS visible	2	Right Ventricle (RV) Visible	3
Inferior wall / IVS visible	2	Full Segment Pericardium Visible	2
ANATOMICAL CLARITY:		**ANATOMICAL CLARITY:**	
LV Cavity clarity, (Endocardial Border)	4	LV Cavity Clarity (distinguishable border)	4
Anterior leaflets (MITRA Valve)	3	LV Anteroseptal Wall Clarity	3
Posterior leaflets (MITRA Valve)	3	LV Inferolateral Wall Clarity	3
DEPTH-GAIN:		**DEPTH-GAIN:**	
Image Sectorial Gain	**4**	Sectorial Gain	**4**
No Excess Gain	3	No Excess Gain	3
Minimum Artefacts	3	Minimum Artefacts	3
LV FORESHORTEN:		**CAVITY FORESHORT:**	
LV Apical Segment present	**4**	No-False Apex present	4
Normal-Shaped Diastole	3	No-Apex Diastole	3
Normal-Shaped Systole	3	No-Apex Systole	3

7.4.3 Quality Optimization Procedure

In all existing ultrasound image quality studies, criteria defined for objective assessment centred on limited features like subjective clarity of image's edges, valves, chambers, and gains. These clearly represent some important features by which anatomical details of the myocardium are analysed. In our studies, we have identified subjective correlation between observer perception of an image's anatomical features and the magnitude of distinguishable features present in the image. Thus, we proposed that objective optimization is based on defined anatomical and pathological features that encompass the echo image. Optimization protocol in this regard, would focus on improving the prediction scores for the following attributes:

1. ***Apical Visibility:*** This is a minimum clinically required feature that involves significant acquisition experience. In A4C plane, apical visibility defines the on-axis projection of the myocardium with a beam cutting through the heart's apex region. This presents a four-chamber view. This view should remain for both diastole and systole frame because of its clinical and pathological significance, paramount to clinical measurement, quantification, and interpretation (Mitchell *et al.*, 2019).

   Assuming the intraventricular septum in Figure 7.2 is oriented along any point P_i on x-axis, this position indicates a spatial distribution with structural deviation from origin P_0 can be corrected by equations (7.1; 7.2), taking moment β from probes' medial (x_p) or lateral translation (y_p) to obtain optimum apical projection (x_1, y_1). In the PLAX view, the left ventricle (LV) apex is not visualised, but emphasis is placed on anatomical orientation of the pericardium, RV at the apex, and LV chamber for linear and volumetric measurement. Pipeline's feedback displays in real-time, the current quality scores while fine tuning probe's translations until the highest score is achieved. Quality score ranges from 0.1 is rejected as unsuitable for quantification to maximum score range of 0.9 as optimum quality, respectively.

$$x_1 = \left(x_p - x_c\right)\ cos\ cos\ \beta\ - \left(y_p - y_c\right) sin\ sin\ \beta\ + x_c \qquad (7.1)$$

$$y_1 = \left(x_p - x_c\right)\ sin\ sin\ \beta\ - \left(y_p - y_c\right) cos\ cos\ \beta\ + y_c \qquad (7.2)$$

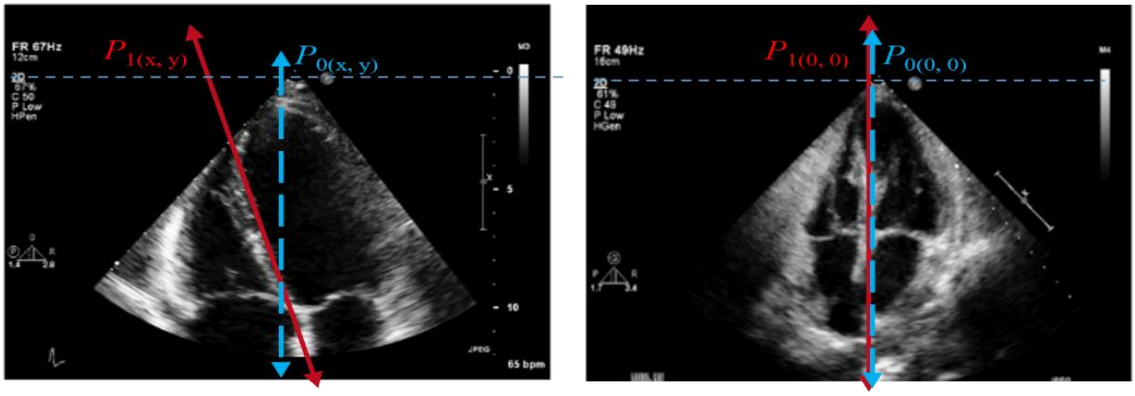


Figure 7.3: Showing the magnitude of (off-axis) projection indicated by red arrow. An optimum (on-axis) projection shows the interventricular septum runs vertically down the middle of the screen indicated by blue arrows in parallel to the red arrow.

2. ***Anatomical Clarity****:* This is a legacy attribute in objective assessment. Unlike non-medical images, chamber cavities, walls and valves are soft tissues that only present rough boundaries and contractive edges. Kurt, et al., (Kurt *et al.*, 2009), have demonstrated the impact of contrast echocardiography (Kurt *et al.*, 2009), however, with respect to quantification, anatomical clarity is visualised by several distinguishable fast-moving pixel's formations during cardiac cycles. This quality element addresses the degree of distinguishable pixel element representing the endocardial border cavities or clear distinction between the intraventricular septum in A4C, pericardium in PLAX, valves, any trabeculated pericardial fluids and endocardial walls. It offers the perceptual distinction between pixels luminance levels which is described by the root means square (RMS) contrast where $f(x, y)$ represents the normalised pixels and μ_f is the mean normalised image in equation (7.3). With RMS contrast, contrast of cardiac frames with dissimilar and similar anatomical features can be computed without the need for angular frequency content. Moreover, cardiac frames with very high or very low contrast represent the extreme ends of the spectrum and pose significant challenges (Nagata *et al.*, 2018), (Kurt *et al.*, 2009) where little expertise is found. A very high contrast could generate artefacts and potentially obscured essential pathological details while a very low-contrast image could bear significant anatomical details required for clinical measurement. For optimization on clarity attributes, operators can rely on real-time scores while rocking, fanning or use probe cardinal compression until optimum score is achieved.

$$f_c = [\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\{f(x,y) - \mu_f\}^2]^{\frac{1}{2}} \qquad (7.3)$$

3. ***Image Depth Gain*****:** Depth-gain quality features present a measure of intensity of discrete signal samples of a specific region of interest. Echocardiograms' acoustic beams, some of which pass through trabeculated tissues, yields subtle impedance which influence the intensity of the image signals. This explains why anatomical visibility are susceptible to depth changes, sector width and frame speed. Selection of appropriate transducer probes can provide improved intensity at near field of tissue penetration (Sassaroli *et al.*, 2019), while low frequency probes yield the opposite. Furthermore, an excessive gain at the near field usually possesses strong intensity or high amplitude and may become excessively high or excessively low at the far field region. In the same vein, excessive gain can present as pulmonary fluid in some cases (Dong *et al.*, 2020) and images with very low gain attributes but bear significant anatomical details or noticeable artefact are not ignored in clinical practice. Nevertheless, potential introduction of artefact from excessive gain would equally exacerbate visibility issues, yield an incorrect depiction of true anatomical tissues or obscure relevant anatomical details. Therefore, improper depth-gain can induce significant dis-uniformity in the pixel intensities across the image, most often at the lower part of the image sector. For real-time optimization, time gain compensation (TGC) which compensate the unequal attenuation and spreading of the received signals, are often used to optimise images' near-field or far-field attenuation quality, along with appropriate probe choices. This is where the optimization algorithm becomes relevant, by effectively guiding the ultrasound operator to choose the right setting while **scoring the effect** of the choices in real-time. TGC setting is repeated until optimum depth-gain is achieved.

4. ***Apical Foreshortening****:* This is inherently common in echocardiogram acquisition as a type of deformation of the ventricular apical region. Smistad et al., (Smistad *et al.*, 2020), have described the importance of real-time detection of apical foreshortening with deep learning pipeline. Foreshortening is a non-linear structural deformation where changes in size of the areas and volumes become geometrically incongruent (Ünlü *et al.*, 2019). It accounts for inaccurate quantification of ejection fraction (EF) (Labs *et al.*, 2020), prevents the detection of crucial pathologies in the apical region and exacerbates

clinical measurements. However, in a PLAX view where LV apex visibility is not required, apex visibility could be taken as 'false-apex' (Ünlü *et al.*, 2019) and counts as LV foreshortening. From a clinical standpoint, eliminating foreshortedness is paramount for optimum anatomical assessment and diagnosis. For real-time optimization, pipeline displays objective score relating to the magnitude of foreshortening in the current frame, this can be improved using one or the combination of probe manipulations until foreshortening score is achieved.

7.4.4 Model Architecture & Training

Prior to the implementation of real-time streaming of ultrasound protocol, this research has established an efficient light-weight spatiotemporal model (illustrated in Figure 7.4 process 2 and 3) based on differentiable neural architecture search (NAS) approach (Azarmehr *et al.*, 2021). The predictive model was based on earlier work; multi-stream time series regression architecture (Labs, Zolgharni and Loo, 2021) implemented via model subclassing for greater control, each independent stream predicts specific quality attribute proposed in section 2.2 and included a corresponding prediction for view classification simultaneously. The architecture is logically divided into two parts; the first shared layer allows weight sharing through TensorFlow API module, while extracting the hierarchical spatial feature in the frame sequence. The resultant vector is flattened and fed into the second part of the network, a single layered Long Short-Term Memory (LSTM) (Donahue *et al.*, 2016) for temporal extraction. The spatiotemporal architecture is trained on 24,000 frames of 224 x 224 x 3 spatial size, with 8,000 validation samples in 80:20 ratio. Predictions are made via fully connected layers which compute specific quality scores and the probability for discrete labels via logistic regression module, simultaneously. Each layer employs Rectifier Linear Units (ReLU) for its internal activation function while the output layer employs sigmoid function to provide boundary for normalised scores on each model output. The model incorporated dual loss functions; mean squared error (MSE) for regression and binary cross entropy for view classification are optimised via adaptive moment estimation (ADAM). The resultant output scores are bound normalised in the range of (0 to 1), yielding four quality attributes per frame.

7.4.5 Proposed Optimization Pipeline

The objective quality assessment pipeline illustrated in Figure 7.4, is intended for real-time operators' feedback and optimization of cardiac image quality before clinical measurement and quantifications. The experiment was carried out on Z600 Mini server with GeForce GTX 970 chipset's Maxwell GPU architecture and featuring 4GB RAM coupled to 1,664 CUDA cores. The pipeline framework accepts high-speed, streaming (frame) data of any varying length for either A2C, A4C or PLAX from GE Vivid.i ultrasound source equipment, composite video input or echo cine loop video (in retrospect) while observing all clinical protocols in transthoracic workflow. Each user session is divided into four sequential processes, each process with its respective varying threads. The heart of process step 1 is an external frame grabber with capability for high data rates, frame buffer, and low latency. During the clinical ultrasound acquisition phase, two-dimensional cardiac frame data is sequentially fed into the encoder module where real-time feature extraction takes place. While maintaining active connection with ultrasound equipment, process step 2 establishes two major threads of spatiotemporal convolution to predict four specific quality attributes scores and logistic probability module handles the prediction class of currently generated image as illustrated in process step 3. The objective scores are then visualised on the fast-moving frames, in real-time, providing specific feedback to operators who then make necessary probe's adjustment until optimum quality is achieved on the specific attribute. The final process step 4 allows the operator to record the optimised cine loop in Microsoft's audio video interleave (.AVI) format, or a still image from the sequenced frames in Joint Photographic Experts Group (.JPG) format. Each of the sessions can thus be recorded and sequentially stored for further future assessment.

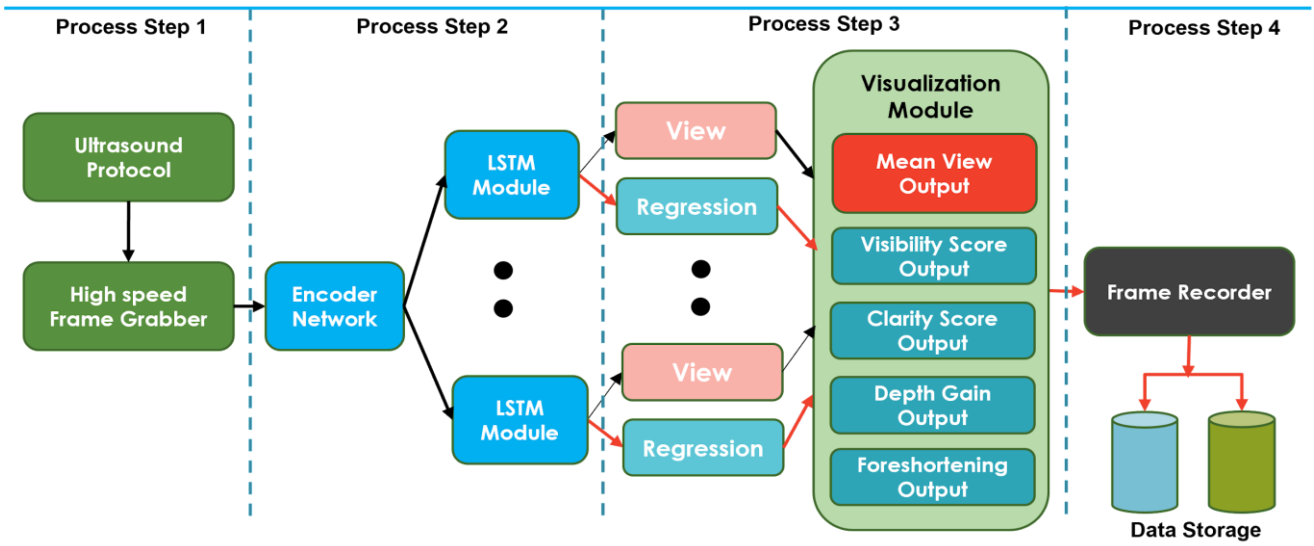


Figure 7.4: Block diagram of a real-time quality assessment and optimization pipeline showing essential processing steps, threads for user session. Features embedded 4 streams deep learning architecture dedicated to assessment and operators' feedback on apical visibility, anatomical clarity, depth gain and apical foreshortening attributes of image quality.

7.5 Evaluation Metrics

Evaluation metrics used encompass the proposed architecture and the integration of model into optimisation pipeline and are enumerated as follows:

A. ***Model Performance:*** Since the model uses multiplex variables for each score attribute, performance was evaluated via objective function using linear correlation coefficient (LCC) in equation (7.4); measures linear difference between cardiologist's score (Q_{MS}) and algorithm's predicted score (Q_{PS}). Minimal LCC error indicates the best fit model, hence better predictions. Figure 7.5, indicate LCC error distribution per selected quality attributes. Model's accuracy was determined in (6) MAE (equation 7.5), while computational inference speed was found at 4.24ms per frame as detailed in Table 14. Results reinforce possibility for real-time feasibility and clinical deployability.

$$Err_{LCC} = \frac{\sum_{i=1}^{N}(\bar{x}_i - \hat{x}_i)(\bar{y}_i - \hat{y}_i)}{\sqrt{\sum_{i=1}^{N}(\bar{x}_i - \hat{x}_i)^2}\sqrt{\sum_{i=1}^{M}\bar{y}_i - \hat{y}_i)^2}} \qquad (7.4)$$

$$Acc_m = 1 - \left(\frac{1}{n}\sum_{i=0}^{n}|Q_{MSi} - Q_{ASi}|\right) * 100 \qquad (7.5)$$

B. ***Pipeline Performance and Validation:*** Acquisition of cardiac frames in PLAX, A4C and A2C were performed using the phased array probe with GE Healthcare (Vivid.i) ultrasound equipment known as the source equipment. This acquisition was done by an experienced clinician under transthoracic laboratory conditions. The frames were stored in DICOM and AVI formats to allow retrospective assessment for the purposes of performance evaluation and analysis. Our source equipment features relevant hardware interface ports for possible external connectivity. In the setup, an off-the-shelf but a high-end, high speed frame grabber with high-definition media interface (HDMI) (Xiong *et al.*, 2013) input port and universal serial bus (USB) 3.0 (Ghetia, Tripathi and Gupta, 2013) output port was considered. USB 3.0 boasting of 5Gbps data rates with short cable connection (one metre length) was selected to avoid excessive transmission delay. Our host equipment, an Intel i7 dual core (Khan *et al.*, 2019) laptop running deep learning algorithm for real-time quality assessment as described in our methodology. Hereafter, each cardiac cine loop was reproduced and visualised on the source and destination screens where quality scoring is performed in real-time. In practice, frame length and frame speed are expected to vary significantly while providing real-time feedback to the operator during acquisition phase, pipeline performance was therefore, estimated using aggregated values for end-to-end classic characteristic delay in transmission D_{tx}, propagation D_{gt}, processing D_{pt}, and queuing D_{qt}; given in equation (7.6). A standard synchronisation clock is set up for both source and destination, to measure bits transmission and delivery time. In a one-way delay scenario, which involves streaming video exhibits the properties of probability density function, a randomized pixelating variable in a continuum as opposed to discrete random variable.

$$D_{e2e} = \frac{l}{q} + \frac{d}{s} + \left[\frac{i_a(l)}{f} + (b_a(l) * t_m)\right] + \frac{1}{\mu} ln\, ln\, \frac{\beta}{\mu} \qquad (7.6)$$

Where delay components are expressed as sum of all the delays; $\{d_1, d_2, d_3, \ldots d_N\}$, average values taken over a series of measurement is thus calculated:

$$\Delta_\mu = \frac{\sum_{i=1}^{N} d_i}{N} \qquad (7.7)$$

Each term in the equation, except D_{gt} and D_{pt} which could be microscopically negligible due to latest advancement in processing power, constitutes significant impact to the overall delay mechanism; where l is the data packet length, q for rate of data transmission, d for

distance using cable connection, i_a pipeline embedded instructions, f processor's clock frequency, b_a buffer delay, t_m memory access time and D_{qt} which details the queue waiting time using β as arrival and μ departure rate. The overall delay (in milliseconds) must satisfy real-time feedback support for cardiac frames between 40fps and 60fps.

7.6 Results and Analysis

In this study, we have presented the most comprehensive attributes and large dataset for domain-specific evaluation, assessment, and optimization than any known previous studies.

Both the multivariate model and the pipeline were evaluated using metrics detailed in section 7.5 while (Table 7.2) list the mean accuracy and error distribution per quality attribute is also depicted (Figure 7.5) by box plot, showing the error distribution for apical visibility, anatomical clarity, depth gain and foreshortening properties, respectively. The model prediction speed was found to be at 4.24ms per frame for input pixel size of 128 x 128 x 3, which is sufficient for real-time deployment. However, this speed was found to be much higher considering the end-to-end delay as detailed in Table 7.2. Displayed samples indicated a real-time quality scores evaluation per frame (Figure 7.6) showing view classification type (ID), score for Apical visibility (VS), LV clarity, Depth Gain (DG), Apical foreshortening (FS), and aggregated quality scores respectively.

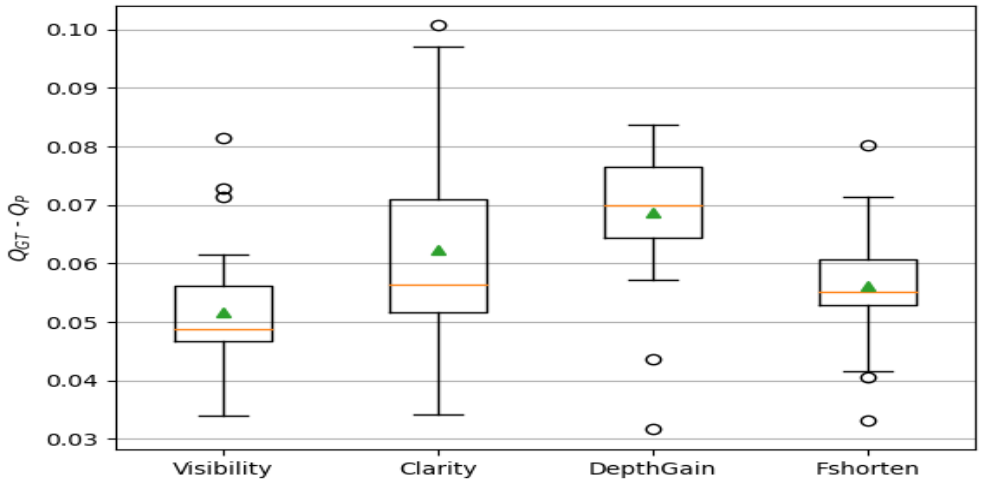


Figure 7.5: Error distribution per model outputs, computed as the absolute difference between predicted score per attributes and experts scores. Very minimal (0.032%) percentage of error distribution and high accuracy are paramount indicators to reliable quantifications.

Table 7.2: Computed errors per quality attributes and end-to-end delay measured in milliseconds.

	Visibility	**Clarity**	**Depth Gain**	**Foreshortening**
Mean +/- SD	0.052 0.011	0.062 0.017	0.069 0.011	0.056 0.010
Average End-to-End Delay D_{e2e} (ms)	**127.90* (* stereo process)**			

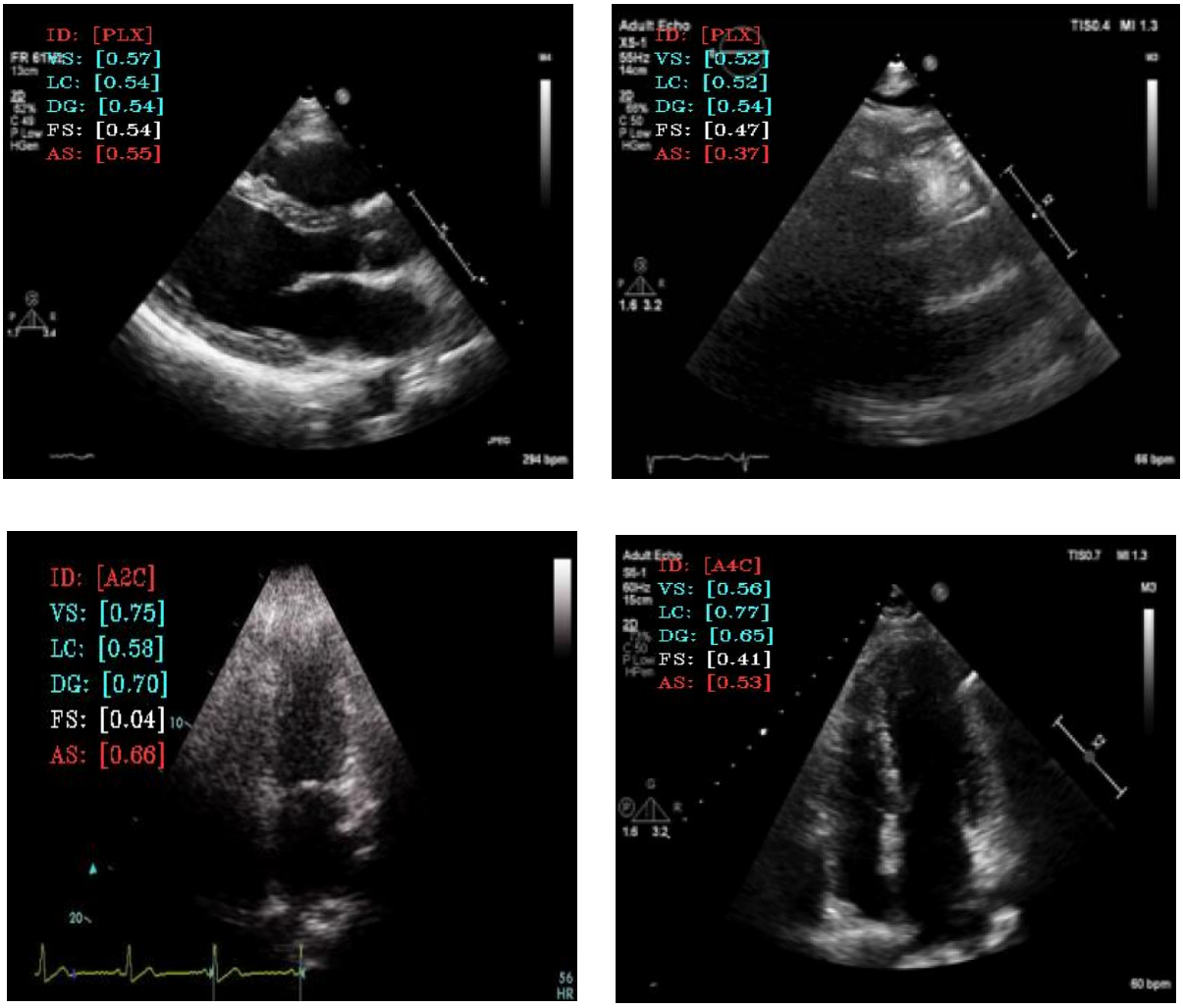


Figure 7.6: Showing samples with retrospective quality grading for visibility (VS), clarity (LC), depth-gain (DG) and apical foreshortening (FS). Pipeline model also shows image view classification and overall quality score (AS). Each quality grading varies from 0.0 to 1 and reflects the aspect of image quality that must be optimised. Pipeline allows real-time assessment and optimization simultaneously.

Although, sample results are self-explanatory with the prediction scores clearly visualised on each specimen. First, the model accurately identifies the image view types, and provides the quality rating for each quality element. Also, samples quality are better than the other in terms of clarity, visibility, depth-gain and foreshortedness. High score range indicate better quality while low score range indicate poor quality.

7.7 Conclusion

In this study, I have presented the feasibility for the most comprehensive objective attributes of cardiac images with provides the underlining principle of real-time optimization of ultrasound image quality.

Since ultrasound image quality is paramount to accurate diagnosis and the acquisition of optimum image quality involves significant skill, an effective optimization utility would be termed a priced tool. Therefore, real-time optimization utility is clinically relevant to effective LV diagnosis, offset inherent variability, and save life during clinical emergencies. The possibility of accelerating clinicians' acquisition skill while providing arbitration for standard reproducibility in clinical practice is also a huge benefit.

Consequently, this thesis proposes a combined methods of multivariate assessment and real-time optimization of echocardiographic images. This chapter have demonstrated, by experiment the performance of deep convolutional neural pipeline suitable for operator feedback. This can accelerate clinicians acquisition skills and save lives during medical emergencies. Moreover, the author has considered imaging planes in A2C, A4C and PLAX, since these are de facto requirement for clinical measurement and quantification, as recommended by the Association of American Cardiologists.

A future study may include wider population and intensive clinical trials with pipeline exhibiting support for different image compression and selective quality attributes that would satisfy individual laboratory requirements.

The author has considered the optimization feasibility on four clinical element of image quality and believes it's significant for comprehensive clinical standard framework that can be implemented in various laboratories for translatory advantages.

A more comprehensive study would include additional criteria for 3D image quality assessments and other point of care systems.

The author admit that ground truth annotation process was laboriously expensive, hence the annotations made by a single cardiologist and an accredited annotator is an indicator that an intra-observer variability may influence model' performance if not included in the analysis. Nevertheless, an intra-observer variability could be taken as different standards that exist in clinical laboratories dues to legislation or expert preferences.

Due to cost and availability of expert resources, the requirement to obtaining additional annotation of any large echocardiogram dataset may hinder further studies on intra/inter-observer variations on objective model development. My work on ‘Global Framework for Image Quality’ plays out here. Researchers and cardiologists can rely on reinforced learning algorithm rather than using human experts to automatically generate annotation and labels for future studies.

Here, we have optimised on four quality standard benchmark of image quality and believe it would eliminate the limitations imposed by arbitrary optimization techniques currently common in clinical lab or arbitrate to professional efforts in the healthcare sector. Therefore, future study may include other critical considerations required for selective standards that could satisfy different laboratory requirements, regionally or globally.

Chapter 8

Conclusion and Future Work

8.1 General Conclusion

This research focused on contributing to the automation of image quality assessment that exists within the unified clinical workflow of cardiac functional assessment, diagnosis of cardiovascular diseases and decision making in cardiovascular interventions.

Since cardiologist reliance on echocardiography is expected to grow due to its numerous advantages over other imaging systems, there is a concerted effort to provide a reliable automated system for quality assessment, quality standard arbiter and great influence on accurate diagnosis.

Chapter 1 provided the clinical context of the research, the objectives, motivation, and laid foundation for the inherent issues with image quality, the assessment method, limitations, and challenges to ultrasound imaging prior to having a reliable diagnostic view of echocardiography.

Additionally, **Chapter 2** details the clinical background to the issues of obtaining optimum image quality, the hardware throughput and clinical image acquisition.

In **Chapter 3**, technical background was detailed with introduction to several artificial intelligence models in the public domain. Also details of the approach implemented in this thesis are semi-automatic and automatic search for neural networks that is considered suitable for assessing image quality in 2d echocardiograms. While in Chapter 4, the view classification approach was implemented using the PACS dataset. The aim was to provide automatic identification of any standard apical views that existed among fourteen (14) different views in the private dataset PACS. Chapter 4 also detailed the method of semi-automatic neural architecture search for most efficient lightweight model architecture to suit classification and

provided justification for selecting semi-automatic neural architecture search along with model ensemble learning and the contributions made on NAS-DARTS-based architectures in earlier publication.

Chapter 5 details how the absence of 2D image quality standard and cardiac dataset have limited research effort in automated assessment within recent years. Furthermore, it proposed a novel standard that specifies clinical and anatomical constituents of quality features using domain-specific attributes which are reviewed by expert and cardiologists. It details why a single weighted-average score is grossly insufficient for 2D echocardiogram quality evaluation and provides justification for defining 4 image quality attributes and the novel method of assessing each specific attribute.

In **chapter 6** further expatiate on quality attributes of echo image by taking a comprehensive annotation and criteria for explicit assessment of apical four chamber image. Using apical four chamber view (A4C) the method is although applicable to other apical standard views of echocardiograms, each cardiac specimens were analysed using 21 clinical criteria and characterised based on 7 class group namely: anatomical projection, Intraventricular septum visibility, valves visibility, left ventricle clarity, image sector gain, Foreshortening diastole, and Foreshortening systole. This research acknowledges how it's practically impossible to have multiple annotations per frame or video because it's a costly venture and resorted to semi-supervised learning instead with an additional method of novel spot-checking process which is aimed at reducing trajectory errors associated with semi-supervised learning. Semi supervised learning latches on the fraction of labelled samples to generalise on the unlabelled samples and thus provide discrete labels from the samples label space.

Chapter 7 latches on to the established model used in **chapter 5** and details the implementation of real-time optimization process for acquiring optimum echocardiogram images. It provided the architecture for on-the-job optimization feedback, techniques, and suggestions to users for the onward clinical processes within the unified workflow. In the same vein, it provides the analysis of a model's performance, pipeline performance, and results of a system evaluated against end-to-end data propagation and suitability for clinical workflow.

8.2 Future Works

This proposed approach used in the research work, has yielded superior performance in terms of deployability and use case compared to any existing approach in quality assessment which indicates the specific element of image quality that must be optimised in real-time and provides objectivity on quality assessment score for operators' feedback and guidance.

Since this is a novel approach in objective quality standard and method of assessment of such quality, a comparison with any existing approach would make a fair judgement. Unfortunately, the existing assessment was only based on a weighted average method of quality scoring and did not provide any parallel approach to measure by. Hence, its near impossibility to determine equivalence while different dataset of corresponding annotation was utilised. Therefore, we make our dataset with expert annotations on four quality attributes public at IntSav repository to allow external validation by other researchers and equipment manufactures.

8.2.1 Standard View Classification

This study has focused on the identification of 14 standard apical view using ensemble learning classification method which combined the spatial and temporal element of echocardiograms.

Such a task will be crucial for a real-time view detection system in clinical scenarios where images need to be processed while they are acquired from the patient and/or where the system is to be used for operator guidance. Also, this is useful as a post processing tool to assess cardiac specimens in retrospect or offline assessment.

Unlike some earlier studies where a similar ensemble method was implemented (Østvik, Smistad, Aase et al., 2019); their majority votes ensemble method was based majorly on spatial element of the images rather than combining both spatial and temporal components associated with 2D echocardiogram frames or specimen. (Madani, Ong et al., 2018).

Although 2D echocardiography is widely accepted for preliminary exams, the images produced vary in terms of sizes, look, feel, and protocol. Vendor dependent images still constitute considerable issues that may be resolved by future solutions in cardiac classification. Future

work could detect the apical views as well as vendor signature in the image. This is essential for comparison in terms of performance and informs user preference.

Other studies have revealed the impact of 3D echocardiography but admit that it suffers from a significant reduction in frame rate and image quality in contrast to 2D echocardiogram. Consequently, has limited usage in routine transthoracic exams (Cheng et al., 2018).

Looking at other modality in echocardiogram is essential to technological advancement. For example, recent 3D echocardiography technology has precipitated research interests in recent times however, 3D echo suffers significant reduction in frame rate and image quality in contrast to 2D echocardiogram (Cheng et al., 2018). Consequently, has limited usage in transthoracic routine examination workflow. Future work would consider the inclusion of 3D echo with the view of generating accurate anatomical and pathological classification.

8.2.2 Objective Quality Standards

Echocardiogram objective image quality is not a trivial issue in transthoracic exam workflow since the accuracy and reliability of diagnosis is hugely dependent on optimum quality of the acquired cardiac specimen. Unfortunately, it's not clear to researchers and experts what constitutes the element of 'good quality' which pushes the idea of optimum image quality to mere relative terms.

Plausible research efforts have attempted this issue by defining a reference image and calculating the deviation of any given image to this reference, specifying an image farther from the reference image the less the objective quality score. This may be applicable to non-medical images, however, in echocardiography, this method is not practical since images formation do not follow a hi-fi photography but vary significantly in amplitude and based on patient anatomical and pathological profiles. Therefore, it is difficult to define an image with perfect quality as a reference.

Some studies have been carried out on blind image quality standard; largely focusing on the distortion of images due to compression, with some implementing machine learning algorithms using edge sharpness and random/structural noise level to evaluate image quality. Unfortunately, this approach is difficult to apply to echocardiography because cardiac ultrasound does not present well defined edges due to two facts: 2D cardiac images are formed by interference pattern of scattering centres presenting an inherent poor resolution; and

anatomical features do not present crisp edges because the endocardium is trabeculated there are papillary muscles, the external purkinje network. So there exist relatively subtle acoustic impedance transitions next to larger ones. Prior to consideration for automatic image quality assessment, are the needs to specify and understand what is meant by quality standard and how to represent such standard in clinical practice

Consequently, it is incumbent upon the research to develop a blind image quality standard which does not depend on a reference image but on domain-specific features, clinically relevant to patients' pathological and anatomical profile.

Training a deep neural network requires large amounts of annotated data, which is often very difficult to obtain because cardiac dataset is highly personalised and protected by data protection act in the United Kingdom. Therefore, researcher frantic effort gravitates towards reducing or eliminating the cost associated with experts' annotation by implementing self-supervised learning as demonstrated in chapter 6, characterization of cardiac specimen global properties. Semi-supervised learning provides other generalisation techniques for neural models while obtaining discrete predictions on the unlabelled dataset. Danu et. al., (2020) had recently demonstrated the potential of self-supervision learning to yield significant advantage in classification tasks (Danu, Ciușdel and Itu, 2020).

Therefore, this study has focused on defining and establishing for the first time, new measures of image quality standard using domain-specific, comprehensive anatomical features relevant for linear measurement and quantification. These are reviewed and considered by clinicians and cardiologists. The research also expanded on the objective standard feature to include global characterization of apical four chamber image. Future work therefore would consider global properties on other apical standard views in the private dataset.

Moreover, this study examined the proposed neural network using only one manual expert annotation. Future study can consider examining the performance of the models using more than one manual expert judgement to study the discrepancies of the variability of annotations.

8.2.3 Quality Assessment Method

Image quality assessment is generally approached by defining a reference standard and calculating the deviation of any given image to this reference. Several researchers have tried to solve this using different techniques like universal image quality index (UIQ) (Wang, Z et al 2002) blind image quality assessment (Zhang, Y., et al 2016), These methods may be valid for non-medical images but certainly not practicable in echocardiography, since images vary significantly from patient to patient, and it is difficult to define an image with perfect quality. Therefore, assessment is a limited subjective domain where experts manually sweep through each cardiac specimen and determine its clinical and pathological relevance. The subjective assessment is subject to variability error and potentially cascade further error in quantification and measurement analysis.

Earlier research efforts by Abdi., H et al., (2017; 2018) had demonstrated the feasibility of automated assessment methods of echocardiogram image quality. The method of assessment was based on a weighted average score which was predicted for each cardiac frame. his approach was supported by Luong., et al., (2020) and Dong., et al., (2020). Unfortunately, predicting weighted average score did not culminate into translatory advantage because single prediction score does not provide enough information to the operator as to why the image is tagged as suboptimal quality. Hence, this research focused on defining novel measures and methods of image quality assessment and implemented an artificial intelligence pipeline to access four quality properties of image quality as defined in chapter 5 using the domain-specific, multi-stream, multi-output prediction method.

The research has presented the clinical significance and feasibility of developing an automated quality assessment in 2D echocardiographic images that contribute to automated diagnosis and quantification in echocardiology. An automated image quality assessment technique would be significant as part of a system that could accelerate the learning curve for those training in echocardiography and automated quality control process which is required for both clinical and research purposes. This would provide real-time guidance to less experienced operators, increase their chances of acquiring optimum quality images and enhance diagnostic accuracy of cardiac functions.

Similarly, this study has considered quality assessment for A2C, A4C and PLAX frames as the primary apical view standards to demonstrate the feasibility of clinical application for quality assessment, however, A2C quantifications may not be a focus under clinical practice suitable

for unified quantification, therefore, future study would include other relevant apical view standards like PSAX, A5C, A3C, and other standard that is consider useful for research and diagnostic purposes.

Future work would consider areas of improvement to include assessment of selectable global properties of echocardiogram quality features, this may be significant for laboratory consideration and account for differences in clinical protocol that may exist in different countries or laboratories.

Finally, the future work would consider multiple annotations rather than the annotation provided by one expert cardiologist and once accredited annotator. Intra-observer variability can be examined by obtaining additional annotations from human experts and compared with the error in the predicted scores.

.

8.2.4 Real-Time Optimization

An automated image quality assessment and real-time optimization technique is a high-priced solution in echocardiography workflow for so many reasons: (i) such system could integrate individual steps under cardiac specimen analysis into a unified workflow, (ii) could accelerate the learning curve for those training in echocardiography, (iii) assisting as part of an automated quality control process (for both clinical and research purposes) while providing real-time guidance to operators with a view of increasing chances of acquiring optimum quality images, thereby enhance diagnostic accuracy of cardiovascular diseases.

This research has focused on implementing such a system incorporating what is considered the most relevant standard apical views in A2C, A4C and PLAX cardiac specimens. For example, the study has considered 14 different apical standard views under images classification, while another experiment in this research has successfully implemented pipeline build on deep convolutional neural network to assess 3 different apical standard views with feasibility for global properties characterization. Other apical standards would be considered in future work as they could find relevance with real-time optimization and translatory advantages in clinical practice. This study has therefore, implemented a pipeline that potentially be used for real-time optimization pipeline for A2C, A4C and PLAX images. Although, testing and verification with echo equipment has been hampered by Covid-19 social distance restrictions and lockdown policy which was in effect at the time of this experiment, the offline verification experiment,

which represent a prove of concept and feasibility, indicated a significant success that should be applied in future study to assess different standard echocardiographic views and more quality attributes that would satisfy wider requirements.

A future possible idea would integrate user selectable criteria that would be suitable for point of care deployment and encompass major laboratory assessments criteria.

Although, two annotations or ground truth labels provided by two expert cardiologist and one accredited annotator were used in the study, the properties that encompass intra-observer and inter-observer variability could be examined in future work, by obtaining additional annotations from human experts and compared with the error in the predicted scores.

The implemented pipeline yielded a superior performance in terms of deployability and use case compared to any existing approach in real-time quality optimization which present and indicate the specific element of image quality that must be optimised in real-time and provides objectivity on assessment score for operators' feedback and guidance.

Appendix A

The following are the list of publications made during this PhD study:

(i) **Journal Articles:**

- Azarmehr N, Ye X, Howard P, Lane E, **Labs R**, Shun-shin M, Cole G, Bidaut L, Francis D, and Zolgharni M, 2020. Neural Architecture Search of Echocardiography View Classifiers, Journal of Medical Imaging.
- **Labs RB**, Vrettos A, Loo J, Zolgharni M, 2021. Automated Assessment of Transthoracic Echocardiogram Image Quality Using Deep Neural Networks (under review with Journal of Intelligent Medicine).
- **Labs RB**, Vrettos A, Loo J, Zolgharni M, 2021, Objective characterization of 2D echocardiographic Image Quality Using Semi-Supervised Deep Learning, (Ready for submission for peer review with SPI Journal).
- **Labs RB**, Vrettos A, Loo J, Zolgharni M, 2021, Real-time Optimization of 2D echocardiographic Image Quality for Accurate Clinical Quantifications, (Under final review for submission to IEEE Journal)

(ii) **Conference Proceedings:**

- Ogbuabor, G., **Labs, R. B**., (2018) Human Activity Recognition for Healthcare using Smartphones. In: International Conference on Machine Learning and Computing. ICMLC 2018. Association for Computing Machinery, New York, NY, USA, (pp. 41-46) DOI: https://doi.org/10.1145/3195106.3195157.
- **Labs, R. B**., Vrettos, A., Azarmehr, N., Howard, J.P., Shun-shin, M. J., Francis, D.P. and Zolgharni, M., (2020) Automated Assessment of Image Quality in 2D Echocardiography Using Deep Learning. In: International Conference on Radiology, Medical Imaging and Radiation Oncology. ICRMIRO 2020, Part XVII (pp. 2160 - 2165).

- **Labs R.B**., Zolgharni M., Loo J.P. (2021) Echocardiographic Image Quality Assessment Using Deep Neural Networks. In: Papież B.W., Yaqub M., Jiao J., Namburete A.I.L., Noble J.A. (eds) Medical Image Understanding and Analysis. MIUA 2021. Lecture Notes in Computer Science, vol 12722. Springer, Cham. https://doi.org/10.1007/978-3-030-80432-9_36

References

Zhou, Y. et al. (2018) ‘A Framework for the Generation of Realistic Synthetic Cardiac Ultrasound and Magnetic Resonance Imaging Sequences from the Same Virtual Patients’, IEEE Transactions on Medical Imaging, 37(3), pp. 741–754. doi: 10.1109/TMI.2017.2708159.

Benacerraf, B. R. et al. (2018) ‘Proceedings: Beyond Ultrasound First Forum on improving the quality of ultrasound imaging in obstetrics and gynecology’, American Journal of Obstetrics and Gynecology. Elsevier Inc., 218(1), pp. 19– 28. doi: 10.1016/j.ajog.2017.06.033.

Khamis, Hanan, Grigoriy Zurakhov et al. (2017). ‘Automatic apical view classification of echocardiograms using a discriminative learning dictionary’. In: Medical image analysis 36, pp. 15–21

Hendee, W.R. et al (2002) ‘Medical Imaging Physics’, Fourth Edition, ISBN: 0-471-38226-4 2002 Wiley-Liss, Inc.

Atzeni, Fabiola et al. (2017). The Heart in Systemic Autoimmune Diseases. Elsevier.

Kadappu, Krishna K and Liza Thomas (2015). ‘Tissue Doppler imaging in echocardiography: value and limitations”. In: Heart, Lung and Circulation 24.3, pp. 224– 233.

Prada, Francesco et al. (2015). ‘From grey scale B-mode to elastosonography: multimodal ultrasound imaging in meningioma surgerypictorial essay and literature review’. In: BioMed research international 2015.

Marwick, Thomas H (2006). ‘Measurement of strain and strain rate by echocardiography: ready for prime time’ In: Journal of the American College of cardiology 47.7, pp. 1313–1327.

eCun, Yann, Yoshua Bengio and Geoffrey Hinton (2015). ‘Deep learning’. In: nature 521.7553, pp. 436–444.

Shen, Dinggang, Guorong Wu and Heung-Il Suk (2017). ‘Deep learning in medical image analysis’ In: Annual review of biomedical engineering 19, pp. 221–248.

Suzuki, Kenji (2017). ‘Machine learning in medical imaging before and after introduction of deep learning’. In: Medical Imaging and Information Sciences 34.2, pp. 14–24.

Shrestha, Sirish and Partho P Sengupta (2018). ‘Imaging heart failure with artificial intelligence’. In

Dean, Jeffrey (2016). ‘Large-scale deep learning for building intelligent computer systems’ In:

Feurer, Matthias and Frank Hutter (2019). ‘Hyperparameter optimization’. In: Automated Machine Learning. Springer, Cham, pp. 3–33.

Elsken, Thomas, Jan-Hendrik Metzen and Frank Hutter (2017). ‘Simple and efficient architecture search for convolutional neural networks’ In: arXiv preprint arXiv:1711.04528.

Chen, Liang-Chieh, Maxwell Collins et al. (2018). ‘Searching for efficient multi-scale architectures for dense image prediction’. In: Advances in neural information processing systems, pp. 8699–8710.

Vanschoren, Joaquin (2019). ‘Meta-learning’. In: Automated Machine Learning. Springer, Cham, pp. 35–61.

Zoph, Barret and Quoc V Le (2016). ‘Neural architecture search with reinforcement learning’. In: arXiv preprint arXiv:1611.01578 (cit. on pp. 60, 62, 63).

Yu, Fisher and Vladlen Koltun (2015). ‘Multi-scale context aggregation by dilated convolutions’ In: arXiv preprint arXiv:1511.07122.

Chollet, François (2017). ‘Xception: Deep learning with depth wise separable convolutions’ In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251–1258.

Baker, Bowen et al. (2016). ‘Designing neural network architectures using reinforcement learning’. In: arXiv preprint arXiv:1611.02167.

Hui, Liu and Ye Xinhua (2020). ‘A Novel Improved Multi-Point Matching Based Coronary Disease Quantitative Analysis for Speckle Tracking in Ultrasound Image’. In: Journal of Medical Imaging and Health Informatics 10.2, pp. 489–495.

Liu, Chenxi, Barret Zoph et al. (2018). ‘Progressive neural architecture search’. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 19–34.

Bergstra, James, Daniel Yamins and David Cox (2013). ‘Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures’ In: International conference on machine learning, pp. 115–123.

Miller, Geoffrey F, Peter M Todd and Shailesh U Hegde (1989). ‘Designing Neural Networks using Genetic Algorithms.’ In: ICGA. Vol. 89, pp. 379–384.

Williams, Ronald J (1992). ‘Simple statistical gradient-following algorithms for connectionist reinforcement learning’. In: Machine learning 8.3-4, pp. 229–256.

Klein, Aaron et al. (2017). ‘Fast bayesian optimization of machine learning hyper-parameters on large datasets’ In: Artificial Intelligence and Statistics. PMLR, pp. 528–536.

Chrabaszcz, Patryk, Ilya Loshchilov and Frank Hutter (2017). ‘A downsampled variant of imagenet as an alternative to the cifar datasets’ In: arXiv preprint arXiv:1707.08819.

Domhan, Tobias, Jost Tobias Springenberg and Frank Hutter (2015). ‘Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves’ In: Twenty-Fourth International Joint Conference on Artificial Intelligence.

Wei, Tao et al. (2016). ‘Network morphism’. In: International Conference on Machine

Learning, pp. 564–572.

Pham, Hieu et al. (2019). ‘Efficient neural architecture search via parameter sharing’. In: arXiv preprint arXiv:1802.03268.

Zoph, Barret, Vijay Vasudevan et al. (2018). ‘Learning transferable architectures for scalable image recognition’. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8697–8710.

Liu, Hanxiao, Karen Simonyan and Yiming Yang (2018). ‘Darts: Differentiable architecture search’. In: arXiv preprint arXiv:1806.09055.

Khamis, Hanan, Grigoriy Zurakhov et al. (2017). ‘Automatic apical view classification of echocardiograms using a discriminative learning dictionary’. In: Medical image analysis 36, pp. 15–21 (cit. on pp. 4, 67).

Stoitsis, John et al. (2006). ‘Computer aided diagnosis based on medical image processing and artificial intelligence methods’ In: Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 569.2, pp. 591–595

Coates, Adam et al. (2013). ‘Deep learning with COTS HPC systems’ In: International conference on machine learning, pp. 1337–1345.

Deo, Rahul C et al. (2017). ‘An end-to-end computer vision pipeline for automated cardiac function assessment by echocardiography’. In: CoRR (cit. on pp. 5, 52, 66).

Ronneberger, Olaf, Philipp Fischer and Thomas Brox (2015). ‘U-net: Convolutional networks for biomedical image segmentation’. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234– 241.

Vaseli, Hooman et al. (2019). ‘Designing lightweight deep learning models for echocardiography view classification’. In: Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modelling. Vol. 10951. International Society for Optics and Photonics, 109510F (cit. on p. 70).

Abdi, A.H., Luong, C., Tsang, T., Allan, G., *et al.* (2017) ‘Automatic Quality Assessment of Echocardiograms Using Convolutional Neural Networks: Feasibility on the Apical Four-Chamber View’, *IEEE Transactions on Medical Imaging*, 36(6), pp. 1221–1230. doi:10.1109/TMI.2017.2690836.

Abdi, A.H., Luong, C., Tsang, T., Jue, J., *et al.* (2017a) ‘Quality Assessment of Echocardiographic Cine Using Recurrent Neural Networks: Feasibility on Five Standard View Planes’, in Descoteaux, M. et al. (eds) *Medical Image Computing and Computer Assisted Intervention − MICCAI 2017*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 302–310. doi:10.1007/978-3-319-66179-7_35.

Abdi, A.H., Luong, C., Tsang, T., Jue, J., *et al.* (2017b) 'Quality Assessment of Echocardiographic Cine Using Recurrent Neural Networks: Feasibility on Five Standard View Planes', in Descoteaux, M. et al. (eds) *Medical Image Computing and Computer Assisted Intervention − MICCAI 2017*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 302–310. doi:10.1007/978-3-319-66179-7_35.

Aschkenasy, S.V. *et al.* (2006) 'Unsupervised image classification of medical ultrasound data by multiresolution elastic registration', *Ultrasound in Medicine & Biology*, 32(7), pp. 1047–1054. doi: 10.1016/j.ultrasmedbio.2006.03.010.

Atzeni, F. *et al.* (2017) 'From Old to New Cardiovascular Complications in Ankylosing Spondylitis', p. 4.

Azarmehr, N. *et al.* (2021) 'Neural architecture search of echocardiography view classifiers', *Journal of Medical Imaging*, 8(03). doi: 10.1117/1.JMI.8.3.034002.

Balaji, G.N., Subashini, T.S. and Chidambaram, N. (2015) 'Automatic Classification of Cardiac Views in Echocardiogram Using Histogram and Statistical Features', *Procedia Computer Science*, 46, pp. 1569–1576. doi: 10.1016/j.procs.2015.02.084.

Balaji, G. N., Subashini, T.S. and Chidambaram, N. (2015) 'Cardiac View Classification Using Speed Up Robust Features', *Indian Journal of Science and Technology*, 8(S7), p. 1. doi:10.17485/ijst/2015/v8iS7/62245.

Barros, B. *et al.* (2021) 'Pulmonary COVID-19: Learning Spatiotemporal Features Combining CNN and LSTM Networks for Lung Ultrasound Video Classification', *Sensors*, 21(16), p. 5486. doi:10.3390/s21165486.

Chen, Q. *et al.* (2021) 'Current Development and Applications of Super-Resolution Ultrasound Imaging', *Sensors*, 21(7), p. 2417. doi:10.3390/s21072417.

Chen, T. *et al.* (2020) 'Big Self-Supervised Models are Strong Semi-Supervised Learners', *arXiv:2006.10029 [cs, stat]* [Preprint]. Available at: http://arxiv.org/abs/2006.10029 (Accessed: 20 October 2021).

Cheng, J. *et al.* (2011) 'A new algorithm for spatial impulse response of rectangular planar transducers', *Ultrasonics*, 51(2), pp. 229–237. doi: 10.1016/j.ultras.2010.08.007.

Cozman, F.G., Cohen, I. and Cirelo, M.C. (2003) 'Semi-Supervised Learning of Mixture Models', In: Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC, 2003 p. 8.

Díez, J. *et al.* (2015) 'Optimizing different loss functions in multilabel classifications', *Progress in Artificial Intelligence*, 3(2), pp. 107–118. doi:10.1007/s13748-014-0060-7.

Domhan, T., Springenberg, J.T. and Hutter, F. (2015) 'Speeding up Automatic Hyperparameter Optimization of Deep Neural Networks by Extrapolation of Learning Curves', p. 9.

Donahue, J. *et al.* (2016) 'Long-term Recurrent Convolutional Networks for Visual Recognition and Description', *arXiv:1411.4389 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1411.4389 (Accessed: 3 July 2020).

Dong, J. *et al.* (2020) 'A Generic Quality Control Framework for Fetal Ultrasound Cardiac Four-Chamber Planes', *IEEE Journal of Biomedical and Health Informatics*, 24(4), pp. 931–942. doi:10.1109/JBHI.2019.2948316.

Ebadollahi, S., Shih-Fu Chang and Wu, H. (2004) 'Automatic view recognition in echocardiogram videos using parts-based representation', in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, Washington, DC, USA: IEEE, pp. 2–9. doi:10.1109/CVPR.2004.1315137.

Elsken, T., Hutter, F. and Metzen, J.H. (2018) 'SIMPLE AND EFFICIENT ARCHITECTURE SEARCH FOR CONVOLUTIONAL NEURAL NETWORKS', p. 15.

Gao, X. *et al.* (2017) 'A fused deep learning architecture for viewpoint classification of echocardiography', *Information Fusion*, 36, pp. 103–113. doi: 10.1016/j.inffus.2016.11.007.

Gaudet, J. *et al.* (2016) 'Focused Critical Care Echocardiography: Development and Evaluation of an Image Acquisition Assessment Tool', *Critical Care Medicine*, 44(6), pp. e329-335. doi:10.1097/CCM.0000000000001620.

Ghetia, S., Tripathi, S. and Gupta, A. (2013) 'PS2 and USB Data Transmission Protocols', *International Journal of Scientific and Engineering Research*, 4(7), p. 5.

Grandini, M., Bagli, E. and Visani, G. (2020) 'Metrics for Multi-Class Classification: an Overview', *arXiv:2008.05756 [cs, stat]* [Preprint]. Available at: http://arxiv.org/abs/2008.05756 (Accessed: 9 December 2021).

Greff, K. *et al.* (2017) 'LSTM: A Search Space Odyssey', *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), pp. 2222–2232. doi:10.1109/TNNLS.2016.2582924.

Hanft, L.M., Korte, F.S. and McDonald, K.S. (2007) 'Cardiac function and modulation of sarcomeric function by length', *Cardiovascular Research*, 77(4), pp. 627–636. doi:10.1093/cvr/cvm099.

He, K. *et al.* (2015) 'Deep Residual Learning for Image Recognition', *arXiv:1512.03385 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1512.03385 (Accessed: 3 July 2020).

Hernández-Vicente, A. *et al.* (2021) 'ECG Ventricular Repolarization Dynamics during Exercise: Temporal Profile, Relation to Heart Rate Variability and Effects of Age and Physical Health', *International Journal of Environmental Research and Public Health*, 18(18), p. 9497. doi:10.3390/ijerph18189497.

Hinton, R.B. and Yutzey, K.E. (2011) 'Heart Valve Structure and Function in Development and Disease', *Annual Review of Physiology*, 73(1), pp. 29–46. doi:10.1146/annurev-physiol-012110-142145.

Howard, A.G. *et al.* (2017) 'MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications', p. 9.

Hu, H. *et al.* (2019) 'Efficient Forward Architecture Search', In: 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada p. 10.

Huang, G. *et al.* (2017) ‘Densely Connected Convolutional Networks’, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, pp. 2261–2269. doi:10.1109/CVPR.2017.243.

Huang, K.-C. *et al.* (2021) ‘Artificial Intelligence Aids Cardiac Image Quality Assessment for Improving Precision in Strain Measurements’, *JACC: Cardiovascular Imaging*, 14(2), pp. 335–345. doi: 10.1016/j.jcmg.2020.08.034.

Ioffe, S. and Szegedy, C. (2015) ‘Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift’, In: Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 2015. JMLR: W&CP volume 37 p. 9.

Jensen, J.A. (1991) ‘A model for the propagation and scattering of ultrasound in tissue’, *The Journal of the Acoustical Society of America*, 89(1), pp. 182–190. doi:10.1121/1.400497.

Jensen, J.A. and Munk, P. (1997) ‘Computer Phantoms for Simulating Ultrasound B-Mode and CFM Images’, in Lees, S. and Ferrari, L.A. (eds) *Acoustical Imaging*. Boston, MA: Springer US (Acoustical Imaging), pp. 75–80. doi:10.1007/978-1-4419-8588-0_12.

K B, P. and J, M. (2020) ‘Design and Evaluation of a Real-Time Face Recognition System using Convolutional Neural Networks’, *Procedia Computer Science*, 171, pp. 1651–1659. doi: 10.1016/j.procs.2020.04.177.

Khan, F. *et al.* (2019) ‘Intel’s core i7 10th generation (1065G7) Review’, *International Journal of Computer Sciences and Engineering*, 7(10), pp. 174–176. doi:10.26438/ijcse/v7i10.174176.

Kurt, M. *et al.* (2009) ‘Impact of Contrast Echocardiography on Evaluation of Ventricular Function and Clinical Management in a Large Prospective Cohort’, *Journal of the American College of Cardiology*, 53(9), pp. 802–810. doi: 10.1016/j.jacc.2009.01.005.

Labs, R.B. *et al.* (2020) ‘Automated Assessment of Image Quality in 2D Echocardiography Using Deep Learning’, In: *ICRMIRO 2020: International Conference on Radiology, Medical Imaging and Radiation Oncology At: Paris*, Part XVII (Paris France June 25-26, 2020, Part XVII), pp. 2160–2165.

Labs, R.B., Zolgharni, M. and Loo, J.P. (2021) ‘Echocardiographic Image Quality Assessment Using Deep Neural Networks’, in Papież, B.W. et al. (eds) *Medical Image Understanding and Analysis*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 488–502. doi:10.1007/978-3-030-80432-9_36.

Lang, R.M. *et al.* (2015) ‘Recommendations for Cardiac Chamber Quantification by Echocardiography in Adults: An Update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging’, *European Heart Journal – Cardiovascular Imaging*, 16(3), pp. 233–271. doi:10.1093/ehjci/jev014.

Lathuiliere, S. *et al.* (2020) ‘A Comprehensive Analysis of Deep Regression’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(9), pp. 2065–2081. doi:10.1109/TPAMI.2019.2910523.

Leclerc, S. *et al.* (2019) 'Deep Learning for Segmentation Using an Open Large-Scale Dataset in 2D Echocardiography', *IEEE Transactions on Medical Imaging*, 38(9), pp. 2198–2210. doi:10.1109/TMI.2019.2900516.

Liao, Z. *et al.* (2019) 'On Modelling Label Uncertainty in Deep Neural Networks: Automatic Estimation of Intra-observer Variability in 2D Echocardiography Quality Assessment', *arXiv:1911.00674 [cs, eess, stat]* [Preprint]. Available at: http://arxiv.org/abs/1911.00674 (Accessed: 2 January 2020).

Liu, C. *et al.* (2018) 'Progressive Neural Architecture Search', in Ferrari, V. et al. (eds) *Computer Vision – ECCV 2018*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 19–35. doi:10.1007/978-3-030-01246-5_2.

Luong, C. *et al.* (2021) 'Automated estimation of echocardiogram image quality in hospitalized patients', *The International Journal of Cardiovascular Imaging*, 37(1), pp. 229–239. doi:10.1007/s10554-020-01981-8.

Maznyczka, A.M. *et al.* (2021) 'Risk Stratification Guided by the Index of Microcirculatory Resistance and Left Ventricular End-Diastolic Pressure in Acute Myocardial Infarction', p. 10.

Merritt, C.R.B. (1987) 'Chapter 1 - Physics of Ultrasound', J Clin Ultrasound. 1987;15(9):591-597. p. 33.

Mesmakhosroshahi, M. and Kim, J. (2012) 'Improving spatio-temporal feature extraction techniques and their applications in action classification', in *2012 Visual Communications and Image Processing*. *2012 Visual Communications and Image Processing (VCIP)*, San Diego, CA, USA: IEEE, pp. 1–6. doi:10.1109/VCIP.2012.6410811.

Mézière, F. *et al.* (2014) 'Measurements of ultrasound velocity and attenuation in numerical anisotropic porous media compared to Biot's and multiple scattering models', *Ultrasonics*, 54(5), pp. 1146–1154. doi: 10.1016/j.ultras.2013.09.013.

Mitchell, C. *et al.* (2019) 'Guidelines for Performing a Comprehensive Transthoracic Echocardiographic Examination in Adults: Recommendations from the American Society of Echocardiography', *Journal of the American Society of Echocardiography*, 32(1), pp. 1–64. doi: 10.1016/j.echo.2018.06.004.

Nafchi, H.Z. and Cheriet, M. (2018) 'Efficient No-Reference Quality Assessment and Classification Model for Contrast Distorted Images', *arXiv:1804.02554 [cs]* [Preprint]. doi:10.1109/TBC.2018.2818402.

Nagata, Y. *et al.* (2018) 'Impact of image quality on reliability of the measurements of left ventricular systolic function and global longitudinal strain in 2D echocardiography', *Echo Research and Practice*, 5(1), pp. 27–39. doi:10.1530/ERP-17-0047.

Nair, V. and Hinton, G.E. (2010) 'Rectified Linear Units Improve Restricted Boltzmann Machines', In: Appearing in Proceedings of the 27 th International Conference on Machine Learning, Haifa, Israel, p. 8.

Ornstein, H. and Adam, D. (2021) 'Classification of Echocardiogram View using A Convolutional Neural Network', *Artificial Intelligence Research*, 11(1), p. 1. doi:10.5430/air.v11n1p1.

Ouali, Y. *et al.* (2020) 'An Overview of Deep Semi-Supervised Learning', arXiv:2006.05278v2 [cs.LG], https://doi.org/10.48550/arXiv.2006.05278 p. 43.

Potter, E. and Marwick, T.H. (2018) 'Assessment of Left Ventricular Function by Echocardiography', *JACC: Cardiovascular Imaging*, 11(2), pp. 260–274. doi: 10.1016/j.jcmg.2017.11.017.

Ren, P. *et al.* (2021) 'NAS-TC: Neural Architecture Search on Temporal Convolutions for Complex Action Recognition', *arXiv:2104.01110 [cs]* [Preprint]. Available at: http://arxiv.org/abs/2104.01110 (Accessed: 27 October 2021).

Ruddox V, *et al.* (2013) 'Is 3D echocardiography superior to 2D echocardiography in general practice? A systematic review of studies published between 2007 and 2012', *Int J Cardiol*. 2013 Sep 30;168(2):1306-15. doi: 10.1016/j.ijcard.2012.12.002. Epub 2013 Jan 5. PMID: 23295040.

Rudski, L.G. *et al.* (2010) 'Guidelines for the Echocardiographic Assessment of the Right Heart in Adults: A Report from the American Society of Echocardiography', *Journal of the American Society of Echocardiography*, 23(7), pp. 685–713. doi: 10.1016/j.echo.2010.05.010.

Sarvazyan, A.P., Urban, M.W. and Greenleaf, J.F. (2013) 'Acoustic Waves in Medical Imaging and Diagnostics', *Ultrasound in Medicine & Biology*, 39(7), pp. 1133–1146. doi: 10.1016/j.ultrasmedbio.2013.02.006.

Sassaroli, E. *et al.* (2019) 'Image quality evaluation of ultrasound imaging systems: advanced B-modes', *Journal of Applied Clinical Medical Physics*, 20(3), pp. 115–124. doi:10.1002/acm2.12544.

'SCoR & BMUS Guidelines 2020' (2020), p. 153.

Shiming Xiang, Feiping Nie, and Changshui Zhang (2010) 'Semi-Supervised Classification via Local Spline Regression', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11), pp. 2039–2053. doi:10.1109/TPAMI.2010.35.

Siems, J. *et al.* (2020) 'NAS-Bench-301 and the Case for Surrogate Benchmarks for Neural Architecture Search', *arXiv:2008.09777 [cs]* [Preprint]. Available at: http://arxiv.org/abs/2008.09777 (Accessed: 22 October 2021).

Simonyan, K. and Zisserman, A. (2015) 'Very Deep Convolutional Networks for Large-Scale Image Recognition', *arXiv:1409.1556 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1409.1556 (Accessed: 3 July 2020).

Smistad, E. *et al.* (2020) 'Real-time automatic ejection fraction and foreshortening detection using deep learning', *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, pp. 1–1. doi:10.1109/TUFFC.2020.2981037.

Sprawls, P. (2014) 'OPTIMIZING MEDICAL IMAGE CONTRAST, DETAIL AND NOISE IN THE DIGITAL ERA', p. 8.

Srivastava, N. *et al.* (2018) 'Dropout: A Simple Way to Prevent Neural Networks from Overfitting', p. 30.

Saripalli et al 2020 ‘AI Assisted Annotator using Reinforcement Learning’, *SN Computer Science* (6):327. doi:10.1007/s42979-020-00356-z.

Talman, V. and Kivelä, R. (2018) ‘Cardiomyocyte—Endothelial Cell Interactions in Cardiac Remodelling and Regeneration’, *Frontiers in Cardiovascular Medicine*, 5, p. 101. doi:10.3389/fcvm.2018.00101.

Thomas, J.D. *et al.* (2005) ‘Guidelines and recommendations for digital echocardiography’, *Journal of the American Society of Echocardiography*, 18(3), pp. 287–297. doi: 10.1016/j.echo.2005.01.010.

Training | BMUS (2020). Available at: https://www.bmus.org/education-and-cpd/careers-training/training/ (Accessed: 2 December 2021).

Ullah, A. *et al.* (2018) ‘Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features’, *IEEE Access*, 6, pp. 1155–1166. doi:10.1109/ACCESS.2017.2778011.

Ungru, K. *et al.* (2014) ‘Automatic classification of left ventricular wall segments in small animal ultrasound imaging’, *Computer Methods and Programs in Biomedicine*, 117(1), pp. 2–12. doi: 10.1016/j.cmpb.2014.06.015.

Ünlü, S. *et al.* (2019) ‘Impact of apical foreshortening on deformation measurements: a report from the EACVI-ASE Strain Standardization Task Force’, *European Heart Journal - Cardiovascular Imaging*, p. jez189. doi:10.1093/ehjci/jez189.

Wang, Y. *et al.* (2018) ‘An adaptive beamforming method for ultrasound imaging based on the mean-to-standard-deviation factor’, *Ultrasonics*, 90, pp. 32–41. doi: 10.1016/j.ultras.2018.06.006.

Wang, Z. *et al.* (2004) ‘Image Quality Assessment: From Error Visibility to Structural Similarity’, *IEEE Transactions on Image Processing*, 13(4), pp. 600–612. doi:10.1109/TIP.2003.819861.

Wang, Z. and Bovik, A.C. (2002) ‘A universal image quality index’, *IEEE Signal Processing Letters*, Vol. 9(3, March 2002), pp. 81–84. doi:10.1109/97.995823.

WHO reveals leading causes of death and disability worldwide: (2000-2019)? Available at: https://www.who.int/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019 (Accessed: 2 December 2021).

Wu, C.W. (2019) ‘ProdSumNet: reducing model parameters in deep neural networks via product-of-sums matrix decompositions’, *arXiv:1809.02209 [cs, stat]* [Preprint]. Available at: http://arxiv.org/abs/1809.02209 (Accessed: 18 November 2021).

Wu, H. and Gu, X. (2015) ‘Max-Pooling Dropout for Regularization of Convolutional Neural Networks’, in Arik, S. et al. (eds) *Neural Information Processing*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 46–54. doi:10.1007/978-3-319-26532-2_6.

Xie, S. *et al.* (2017) ‘Aggregated Residual Transformations for Deep Neural Networks’, *arXiv:1611.05431 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1611.05431 (Accessed: 22 October 2021).

Xiong, J. *et al.* (2013) ‘Research and Design of Data Transmission System Based on HDMI’, in *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*. *2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*, China: Atlantis Press. doi:10.2991/iccsee.2013.88.

Yang, J. *et al.* (2018) ‘Stereoscopic video quality assessment based on 3D convolutional neural networks’, *Neurocomputing*, 309, pp. 83–93. doi: 10.1016/j.neucom.2018.04.072.

Yoon, Y.E., Kim, S. and Chang, H.-J. (2021) ‘Artificial Intelligence and Echocardiography’, *Journal of Cardiovascular Imaging*, 29 (3), p. 193. doi:10.4250/jcvi.2021.0039.

Zahedi, L. *et al.* (2021) ‘Search Algorithms for Automated Hyper-Parameter Tuning’, *arXiv:2104.14677 [cs]* [Preprint]. Available at: http://arxiv.org/abs/2104.14677 (Accessed: 16 November 2021).

Zahedi, L., Mohammadi, F.G. and Amini, M.H. (2021) ‘HyP-ABC: A Novel Automated Hyper-Parameter Tuning Algorithm Using Evolutionary Optimization’, 00(0), p. 10.

Zeng, X. and McGough, R.J. (2008) ‘Evaluation of the angular spectrum approach for simulations of near-field pressures’, *The Journal of the Acoustical Society of America*, 123(1), pp. 68–76. doi:10.1121/1.2812579.

Zhang, J. *et al.* (2018) ‘Fully Automated Echocardiogram Interpretation in Clinical Practice: Feasibility and Diagnostic Accuracy’, *Circulation*, 138(16), pp. 1623–1635. doi:10.1161/CIRCULATIONAHA.118.034338.

Zhang, J. *et al.* (2021) ‘Ensemble machine learning approach for screening of coronary heart disease based on echocardiography and risk factors’, *arXiv:2105.09670 [cs, stat]* [Preprint]. Available at: http://arxiv.org/abs/2105.09670 (Accessed: 26 October 2021).

Zhang, Y. *et al.* (2016) ‘Blind image quality assessment with improved natural scene statistics model’, *Digital Signal Processing*, 57, pp. 56–65. doi: 10.1016/j.dsp.2016.05.012.