Translation of Computer-Assisted Point-of-care Ultrasound Imaging Methods in a Resource Limited Intensive Care Unit

Phung, Nhat

Awarding institution:
King's College London

School of Biomedical Engineering & Imaging Sciences

Translation of Computer-Assisted Point-of-care Ultrasound Imaging Methods in a Resource Limited Intensive Care Unit

Phung Tran Huy Nhat

This dissertation is submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

in the School of Biomedical Engineering & Imaging Sciences

Supervisors:

King's College London: Andrew King, Alberto Gomez, Reza Razavi

Oxford University Clinical Research Unit: Louise Thwaites, Sophie Yacoub

Abstract

Intensive care units (ICUs) in low- and middle-income countries (LMICs) typically suffer from insufficient staff expertise and lack of resources. These ICUs normally manage different patient cohorts to those in high income countries, for example dengue, tetanus, tuberculosis and HIV patients. The Vietnam ICU Translational Applications Laboratory (VITAL) project, which hosts this PhD project, aims at developing, and testing the utility of, affordable technology including wearable devices, artificial intelligence (AI)-enabled imaging and smart usage of patient data to support critical care clinical decision making in Vietnamese ICUs.

Ultrasound (US) imaging is affordable, portable, and safe, and can be used to investigate many body organs. As a result, US can be an invaluable tool in a resource limited ICU setting. However, US requires extensive operator experience to be carried out effectively. Such expertise is scarce in LMICs, where there are few specialists and formal US training is uncommon.

AI in US is an exciting prospect and has the potential to optimize existing resources and help overcome workforce shortages by assisting US system operation, measurement of biometric parameters from images, image interpretation, and providing insights that can help patient management. As a result, AI can make US accessible and ultimately improve patient outcomes in resource-limited settings. Although there are barriers to deploying AI-enabled US at scale in LMICs, a strategy of promoting local innovation and initiative can accelerate progress towards sustainable AI-enabled US implementation.

A challenge to accelerate the use of AI in US in LMICs relates to the quantity and quality of the available data. Most algorithms are trained with data from high-income countries (HICs), which may not be representative of LMIC populations (both in terms of patients and diseases). This means that even if AI algorithms are approved commercially, their efficacy and translatability may still be inhibited by the lack of data from LMICs.

This thesis evaluates the clinical usability of AI-enabled ultrasound tools in the ICU in a LMIC. To this end, my work has focused on i) identifying the main challenges and opportunities for AI-enabled US in the ICU of a LMIC, ii) collecting and curating a comprehensive US dataset, iii) adapting pre-existing computational methods for real-time classification and quantification of

lung, heart and muscle US videos, and iv) clinical translation of such methods to enable non-experts to obtain expert-quality scans in ICU in LMICs.

Contents

List of figures

List of tables

List of abbreviations

A4C	Apical 4-Chamber
AI	Artificial Intelligence
ARDS	Acute Respiratory Distress Syndrome
AUROC	Area Under the Receiver Operating Characteristics
BIRADs	Breast Imaging Reporting & Data System
CAMUS	Cardiac Acquisitions for Multi-structure Ultrasound Segmentation
CNN	Convolutional Neural Network
CO	Cardiac Output
CRF	Case Report From
CT	Computed Tomography
CXR	Chest X-Ray
DBP	Diastolic Blood Pressure
DICOM	Digital Imaging and Communications in Medicine
ED	End Diastolic
EDV	End Diastolic Volume
ES	End Systolic
ESV	End Systolic Volume
FDA	U.S. Food and Drug Administration
GAN	Generative Adversarial Network
HICs	High-Income Countries
HR	Heart Rate
HTD	Hospital for Tropical Diseases at Ho Chi Minh city
ICC	The Intraclass Correlation Coefficient
ICU	Intensive Care Unit
ICUAW	Intensive Care Unit Acquired Weakness
JSON	Java Script Object Notation
LMICs	Low- and Middle-Income Countries
LoA	Limits of Agreement

LSTM	Long Short-Term Memory
LUS	Lung ultrasound
LV	Left Ventricular
LVCO	Left Ventricular Cardiac Output
LVEF	Left Ventricular Ejection Fraction
$LVEF_{AI}$	Left Ventricular Ejection Fraction by AI-assisted acquisition group
$LVEF_{expert}$	Left Ventricular Ejection Fraction of expert manual measurement
$LVEF_{nonAI}$	Left Ventricular Ejection Fraction by non-AI acquisition group
LVOT-VTI	Left Ventricular Outflow Tract Velocity Time Interval
LVSV	Left Ventricular Strock Volume
LVVED	Left Ventricular End Diastolic Volume
LVVES	Left Ventricular End Systolic Volume
MRI	Magnetic Resonance Imaging
NHTD	National Hospital for Tropical Diseases at Hanoi
OUCRU	Oxford University Clinical Research Unit
OxTREC	Oxford Tropical Research Ethics Committee
PACS	Picture Archiving and Communication System
POC echo	Point-of-care Echocardiography
POCUS	Point-Of-Care Ultrasound
PRETUS	Plugin-based Real Time Ultrasound
RAILUS	Real-time AI-assisted Lung UltraSound
RAIMUS	Real-time AI-assisted Muscle UltraSound
RAISEG	Real-time AI-assisted left ventricular SEGmentation
RCT	Randomized Controlled Trial
RFCSA	Rectus Femoris Crossed Sectional Area
RMSD	Root Mean Square Deviation
ROC	Receiver Operating Characteristic Curve
SBP	Systolic Blood Pressure
SD	Standard Deviation

SEM	Standard Error Measurements
SOP	Standard Operation Procedure
SV	Stroke Volume
TBM	Tuberculosis Meningitis
TIRADS	Thyroid Imaging Reporting and Data System
US	Ultrasound

1 Introduction

Chapter layout

This chapter introduces the challenges addressed by this thesis. It also provides an overview of the work done in this thesis and the structure of the rest of the document.

I begin with the motivation for the work carried out in this project in Section 1.1. Section 1.2 describes the structure of this thesis, providing an outline of the content of each chapter, and listing the contributions made in each chapter. Section 1.3 provides a description of the original contributions to the work presented in each chapter. Finally, Section 1.4 lists the publications, poster presentations, and manuscripts in preparation arising from this PhD.

1.1 Motivation

Optimal management of critically ill patients requires rapid and reliable evaluation of vital organ function. Previously, this required expensive and invasive equipment often unavailable in low- and middle-income countries (LMICs). The VITAL project (www.oucru.vital.org), where my project sits, aims at developing, and testing the utility of, affordable technology including wearable devices, AI-enabled radiology and smart usage of patient data to support critical care clinical decision making in Vietnamese Intensive Care Unit (ICU) settings. This includes ultrasound imaging (US) technology, which is being increasingly used in ICUs across the world [1–3]. US is safe, affordable, portable and can be used to assess different organs, including cardiac and lung evaluation, and may have superior performance to X-ray in the assessment of several lung pathologies [4, 5]. US performed and interpreted at the bedside can quickly establish a diagnosis and guide therapy in ICU patients. The use of US is particularly attractive in LMICs since it is often the only imaging modality available [6, 7].

A major disadvantage of US is the need for specialist training to reliably obtain and interpret images [8], preventing non-specialist operators from using it in the ICU and hence often denying critically ill patients' timely diagnosis and intervention. Due to the increasing availability and perceived safety of US equipment, non-experts sometimes tend to use it more often, resulting in

variable and potentially sub-optimal examinations. Under these circumstances, the relative affordability and availability of US can be a disadvantage if clinical decisions are based on these data. Guidelines for echocardiography in the ICU have been developed to ensure accurate quantification and interpretation of cardiac function, but the final analysis remains dependent on the operator having the experience and knowledge to adhere to these guidelines [8, 9] For example, guidelines recommend quantitative measures of cardiac chambers and valves during assessment to inform clinical decision making. However, in busy clinical environments such as acute emergency settings and ICUs, quantitative analysis may not be practical because of the additional time required for manual tracing and difficulty obtaining images of sufficient quality to do this. As a result, it is acknowledged that visual estimation remains the mainstay of rapid functional assessment in many areas of clinical practice, although it requires considerable experience [10, 11].

Artificial intelligence (AI) is a rather liberally used term describing computational methods, often data-driven, that can be constructed to perform complex tasks that would typically require *intelligence*, such as image classification or parameter estimation. AI has been used broadly in a range of medical imaging tasks. It has also been used to localise structures in US data, to classify echocardiography standard views, to diagnose lung pathologies, and to estimate cardiac function. In that sense, AI methods can help towards use, interpretation and quantification of US imaging, tasks which would normally be taken up by experts. As a result, the main role of AI in LMICs, where availability of experts is scarce, formal training is lacking, and there is a high patient:staff ratio, is to make US examinations easier, quicker and more reliable, as well as more accessible for less experienced operators. Real-time AI in ultrasound imaging holds paramount importance within the realm of medical diagnostics and clinical practice for several reasons. Firstly, while saved views serve as valuable references, they only capture a part of the comprehensive information that clinicians rely upon during ultrasound examinations. The operator's proficiency in guiding the ultrasound probe and interpreting real-time images significantly influences the quality and accuracy of the resulting diagnostic images, setting ultrasound apart from other imaging modalities. Secondly, unlike some other imaging techniques, the quality of ultrasound images intrinsically depends on real-time probe guidance

and interpretation skills, making real-time AI assistance invaluable in ensuring consistency and precision across different operators. Lastly, there is often a lack of standardised quality control measures in ultrasound imaging, highlighting the need for real-time AI solutions to enhance image quality, increase diagnostic accuracy, and ultimately improve patient care. Addressing these facets of real-time AI in ultrasound imaging not only enhances the diagnostic capabilities but also augments the overall quality and reliability of ultrasound examinations, thereby representing a critical frontier in medical imaging research and practice.

In this thesis, I will investigate the clinical usability of real-time AI-enabled ultrasound technology in a resource limited ICU setting. The three main areas covered are lung ultrasound classification, muscle ultrasound segmentation and cardiac function estimation from echocardiography.

1.2 Structure and original contributions of this thesis

Chapters 1, 2, and 3 provide background on the challenges confronted in this thesis, review the literature on existing approaches and describe the clinical study protocol for data collection and evaluation of AI-assisted ultrasound tools. Chapters 4, 5, and 6 describe the original work, articulated as three clinical studies, in this thesis. Each chapter begins with an Outline section, setting out the structure of the chapter.

In Chapter 2 I introduce the clinical motivation for point-of-care ultrasound (POCUS) and AI-assisted ultrasound applications. In addition, I also review the state of the art in automatic ultrasound image analysis techniques and clinical translational studies on AI-based ultrasound imaging analysis. I will then describe the potential role of AI-assisted ultrasound in resource limited ICUs and state the research gap that the following chapters will address.

Chapter 3 is a short chapter which introduces the data collection and the general method to evaluate the clinical benefit of AI-assisted tools used throughout this thesis. A description of the clinical study is also outlined, including the acquisition protocol, annotation methods, real-time AI-assisted ultrasound system and amount of data collected.

Each of Chapters 4, 5, and 6 proposes a different AI-assisted ultrasound application and evaluates its clinical utility and usability in the ICU in both the critical phase (echo and lung ultrasound) and

the recovery phase (muscle ultrasound) of patients. Each chapter begins with an **introduction** section outlining the problem, followed by a **proposed approach** section setting out the methods used in the chapter. The specific clinical experiments are described in the following sections and the **results** section presents their primary outcomes. These results are then discussed in the **discussion** section, and a high-level message of the chapter is distilled in the **conclusion.**

Chapter 4 proposes a real-time AI-assisted lung ultrasound classification system and evaluates the clinical benefit when deploying it in the ICU. It begins by introducing the challenges of classifying common lung ultrasound features in critically ill patients, describing the CNN-based model development to automate this task and the Real-time AI-assisted Lung UltraSound (RAILUS) system. The clinical benefit of RAILUS is then investigated in a clinical controlled environment and in real-time in ICU patients. I also conduct a real-time external validation study in dengue shock patients by enabling non-experts to carryout LUS at the bedside using RAILUS tool. I find significant improvement in the clinician's performance and a reduction in interpretation time when using the RAILUS tool, as well as additional information regarding its usability in a resource-limited ICU setting.

Chapter 5 assesses a real-time AI-assisted muscle ultrasound segmentation system and evaluates the clinical benefit when using it in critically ill patients, who often suffer from muscle wasting during their ICU stay. I begin by introducing the challenges of manual segmentation of the rectus femoris (RF) muscle ultrasound and describing the UNET model to automate the segmentation task as well as the Real-time AI-assisted Muscle UltraSound (RAIMUS) system. A prospective study is also described which evaluates the benefit of using RAIMUS in tetanus patients. We randomized the patients to undergo muscle ultrasound with and without RAIMUS tool and found better inter- and intra-variability in the muscle size measurement when clinicians used RAIMUS, and less time spent on ultrasound examination.

Chapter 6 assesses the feasibility of a Real-time AI-assisted left ventricular (LV) SEGmentation (RAISEG) tool, which was evaluated when non-experts used it in healthy volunteers to acquire 4 chamber view videos. It begins with introducing the challenges of acquiring diagnostic quality cardiac ultrasound views and describing the RAISEG system. This preliminary prospective study

demonstrated the potential of using AI-assisted automatic LV segmentation as a quality control framework to help non expert operators to collect US images of diagnostic image quality.

Finally, Chapter 7 summarizes the contributions of the work described in this thesis and considers the limitations of the work presented here. I also present possible future directions and outline how work performed in this PhD project may lead to further technical and clinical innovations and implementations around ultrasound imaging.

1.3 Statements of contributions

This project cannot be successful without the interdisciplinary, intercontinental, and collaborative teams that consist of technical experts, clinicians, radiologists, nurses, ethical supporters, and data management professionals. I list their names with the corresponding teams below.

Technical team: Hamideh Kerdegari, Miguel Xochicale, Alberto Gomez, Andrew P. King (model development, study design)

Clinical team: (study design, patient recruitment, ultrasound examination, patient monitoring, patient follow-up, study payment, study management)

> **Radiologist:** Le Ngoc Minh Thu, Truong Thi Phuong Thao, Nguyen Thi Mai Thao, Nguyen Thi Hong Van
>
> **Clinician/Intensivist/Cardiologist**: Lam Minh Yen, Nguyen Van Hao, Phan Vinh Tho, Angela McBride, Ha Thi Hai Duong, Nguyen Quoc Viet, Nguyen Hoang Anh, Nguyen Tan Hoang, Luigi Pisani, Louise Thwaites, Sophie Yacoub, Linda Denehy, Marcus P. Schultz, Reza Razavi
>
> **Nurse:** Le Thanh Phuong, Nguyen Thanh Ngoc, Pham Thi Lieu

Clinical trial management team: Nguyen Than Ha Quyen, Ninh Thi Thanh Van, Nguyen Thi Phuong Dung (ethical approval, study management, clinical trials registration)

Data management team: Ho Van Hien (data management, data sharing).

Author of this thesis: Study design, protocol writing, ethical approval submission, project management, patient recruitment, clinician recruitment, data collection, data management, data anonymization, data annotation, model finetuning, data analysis and paper writing.

1.4 Publications

Peer-reviewed conference papers

1. **Nhat, P. T. H.,** Van Hao, N., Kerdegari, H., et al., "Role of AI-enabled Ultrasound Imaging in a Resource Limited Intensive Care Unit". In *Affordable Healthcare and AI for Resource Diverse Global Health" (FAIR)* at MICCAI 2021.
2. **Nhat, P. T. H.,** Kerdegari, H., Pisani, L., et al., "Lung Ultrasound Pathology Classification for ICU Patient Management in Low-Middle Income Country". In *Affordable Healthcare and AI for Resource Diverse Global Health (FAIR)* at MICCAI 2021.
3. Kerdegari, H., **Nhat, P. T. H.,** Nguyen, V. H., Truong, T. P. T., Le, N. M. T., Le, T. P., ... & Gomez, A. (2023). Automatic retrieval of corresponding US views in longitudinal examinations. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 152-161).
4. Kerdegari, H., **Nhat, P. T. H.**, McBride, A., Razavi, R., Van Hao, N., Thwaites, L., ... & Gomez, A. (2021). Automatic detection of B-lines in lung ultrasound videos from severe dengue patients. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)* (pp. 989-993). IEEE.

Open source released software

1. Real-time AI-assisted Lung UltraSound (RAILUS). Github repository: https://github.com/vital-ultrasound/public-lung
2. Real-time AI-assisted Muscle UltraSound (RAIMUS). Github repository: https://github.com/vital-ultrasound/public-muscle
3. Real-time AI-assisted left ventricular SEGmentation (RAISEG). Github repository: https://github.com/vital-ultrasound/public-echo

Peer-reviewed journal publications

1. **Nhat, P. T. H.,** Van Hao, N., Tho, P. V., Kerdegari, H., Pisani, L., Thu, L. N. M., ... & Gomez, A. (2023). Clinical benefit of AI-assisted lung ultrasound in a resource-limited intensive care unit. *Critical Care*, *27*(1), 257.
2. **Nhat, P. T. H.,** Van, H. N., Lam, M. Y., Nguyen, H. A., Dong, P. K., Kerdegari, H., ... & Gomez, A. (2023). Clinical evaluation of AI-assisted muscle ultrasound for monitoring muscle wasting in ICU patients. (Under review).
3. Kerdegari, H., **Nhat, P. T. H.,** McBride, A., Pisani, L., Nguyen, H. V., Duong, T. B., ... & Gomez, A. (2021). B-line detection and localization in lung ultrasound videos using spatiotemporal attention. *Applied Sciences*, *11*(24), 11697.

2 Literature review

Chapter layout

This chapter contextualizes and reviews the techniques and scientific background related to POCUS in critical illness, the role of AI-assisted US imaging in the resource limited intensive care unit, and the state of current research in real-time clinical application of AI in US imaging. This lays the groundwork for this thesis' research contributions.

I begin with the key differences between ICUs in LMICs and high-income countries (HICs), outline the motivation and clinical need for US imaging in the ICU, and explain the differences between POCUS versus consultative US in Section 2.1. Next, in Section 2.2. I review the state-of-the-art in computational methods (mainly deep learning-based) that have been proposed to date to improve US image analysis and review their clinical implementation. Section 2.3 focuses on the opportunities and challenges of implementing AI in US imaging in a low resource setting. Finally, Section 2.4 provides a summary of research in the implementation of AI-based US image analysis in clinical settings.

Section 2.1 and Section 2.3 in this chapter are adapted from the work submitted and accepted for presentation at the *Affordable Healthcare and AI for Resource Diverse Global Health* (FAIR) workshop at MICCAI 2021 [12].

2.1 Clinical background

2.1.1 High income countries (HICs) and LMICs: key differences in ICU settings

Despite increased numbers of ICU beds in many LMICs, compared to high income settings, ICU capacity in most LMICs remains limited, with between 0.1 and 2.5 ICU beds per 100,000 population compared to between 5 and 30 ICU beds in HICs [13]. ICUs in LMICs typically suffer from a lack of sufficient equipment (monitors, imaging devices, etc) and the equipment available may be old or even obsolete [13]. These ICUs typically manage different patients compared to

high income settings, often with a very different disease profile, for example dengue, tetanus, tuberculosis and HIV [14].

In LMICs, alternative imaging modalities are often unavailable in ICUs and therefore US is a particularly valuable tool in managing many critically ill patients in these countries. For example, Acute Respiratory Distress Syndrome (ARDS), a severe lung condition occurring in many critically ill patients, was initially defined using criteria which included computed tomography (CT) scan findings. However, the Kigali modifications to the original (Berlin) criteria allow for diagnosis in resource-limited settings and include the identification of pulmonary pathology with US instead of chest x-ray (CXR) or CT [15].

ICU staff in LMICs are typically less well trained, especially in imaging. Of the 99 ICUs that responded to a survey by Rajamani et. al [16], 75 had no basic critical care echocardiography training. In the remaining 24 ICUs, the teaching process was widely variable. While competent operators, such as cardiologists and radiologists, are capable of acquiring the optimum scanning view resulting in high-quality images, less experienced clinicians are more likely to obtain images with suboptimal quality.

The are not many options for POCUS training in LMICs. The two main possibilities are certification training at a medical school, which normally takes 6-12 months (usually general US or echocardiography courses) and short courses. Due to lack of staff, many clinicians can be prevented from leaving the hospital for certification training. Therefore, many clinicians learn POCUS by attending short courses or Continuing Medical Education courses that are hosted by the medical school or a group of local experts. These courses offer many advantages to the attendees, including favorable faculty-to-learner ratios, a wide variety of high-end state-of-the-art POCUS equipment, scanning on healthy volunteer models with excellent image quality or normal conditions. The issue for clinicians that attend these courses begins when they return to their own institutions and have to scan with their older, often outdated equipment, on patients that may have pathologies that make US imaging challenging and without an expert to coach them on probe movements to improve images or assist with their interpretation. Thus, the utilization of POCUS in LMICs is mostly only carried out by experienced US clinicians.

In summary, the application of US in LMICs may face various barriers that hinder its implementation in ICUs, particularly due to limited training and accurate image interpretation, as discussed above. Limited training is an ongoing challenge that is actively explored through training; however, the effective utilization of US imaging still relies on continuous practice under the supervision of an expert. Due to the limited training and medical certification opportunities in LMICs, ultilising US becomes challenging. AI can play a crucial role in improving employment of US in daily clinical practice.

2.1.2 POCUS in critical illness

Critical illness is defined as the deterioration of an illness resulting in deranged homeostasis which is associated with high morbidity and mortality caused by both the underlying disease and further secondary organ dysfunctions [17, 18]. This leads to life-threatening organ dysfunction requiring advanced organ support and monitoring techniques such as mechanical ventilation, bedside monitoring systems and medical imaging. POCUS has emerged as an extremely valuable clinical diagnostic tool and a powerful monitoring component for critically ill patients [3–5]. Three examples of clinicians carrying out POCUS examinations in the ICU are shown in Figure 2.1.

POCUS is defined as the acquisition, interpretation, and immediate clinical integration of US imaging performed by a treating clinician at the patient's bedside rather than by a radiologist or cardiologist [22]. POCUS is an inclusive term, it is not limited to any specialty, protocol, or organ system. The advent of smaller and more affordable US machines, combined with evidence that healthcare professionals who are not radiologists or cardiologists can become competent in the performance of POCUS, have meant that POCUS is now used in many practice settings and in all phases of care – from screening and diagnosis to procedural guidance and monitoring - and can support clinical decision making in medical practice [23]. A recent study showed that POCUS facilitated confirmation of the suspected clinical diagnosis in up to 50% of cases and supported a change in the initial diagnosis in 23% of cases [24]. Zieleskkiewicz et al demonstrated that the protocolized use of a handheld POCUS device at the bedside may improve the time to initial treatment and improve the diagnosis and outcome of patients with acute respiratory or circulatory failure [25].

Figure 2.1. Examples of POCUS in critically ill patients. (A) shows a clinician carrying out an echocardiography exam on a child, (B) shows a clinician carrying out a lung ultrasound exam, and (C) shows a clinician carrying out an echocardiography exam in an adult.

POCUS performed and interpreted at the bedside can quickly establish a diagnosis and guide therapy in ICU patients [23, 26]. As cardiac and lung pathologies are common and important reasons for critical illness throughout the world, echocardiography and LUS are rapidly gaining value in ICUs. Echocardiography enables a comprehensive hemodynamic assessment as well as

diagnosis of cardiac pathologies like tamponade, valvular diseases, and left ventricle hypertrophy [27, 28]. LUS can evaluate lung aeration, which can help in management of patients with many lung pathologies. In addition, LUS has been demonstrated to be superior to chest X-rays in detecting pneumothorax, pleural effusion, pneumonia, and interstitial syndromes [4, 29]. Some of the potential applications of POCUS in critically ill patients are shown in Figure 2.2.

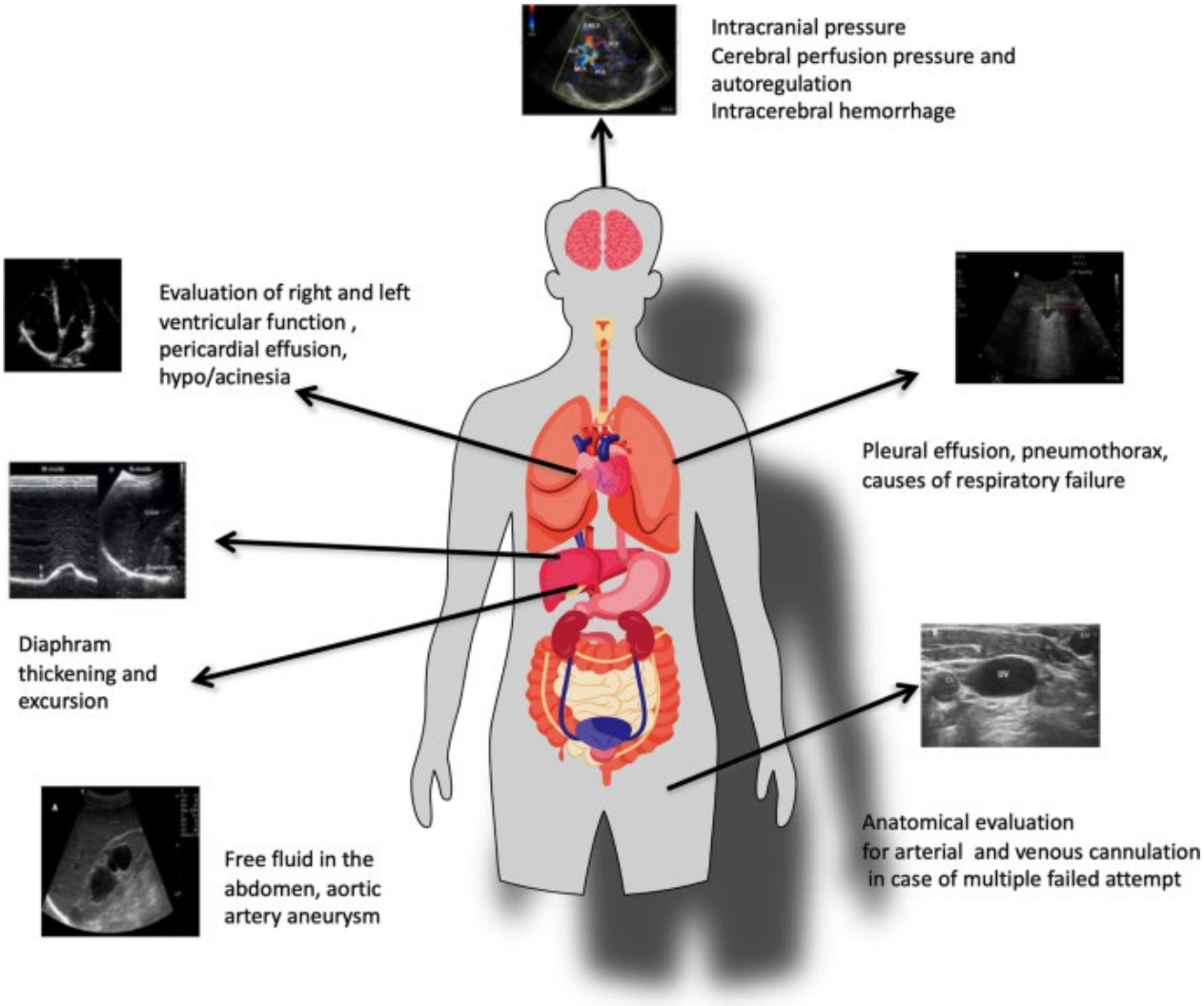


Figure 2.2. Potential POCUS applications in critically ill patients. Ultrasound assessment in critically ill patients for brain, lung, heart, diaphragm, abdomen, and vessels with the main clinical applications is represented. Reproduced from [30].

Many well-established POCUS protocols [27, 31, 32] such as the BLUE protocol and FALLS protocol are designed to assist doctors with the diagnosis and management of pulmonary and

cardiac conditions. In particular, the Kigali protocol [15, 33] has helped to diagnose ARDS in resource-limited settings by using US to identify pulmonary edema instead of chest x-rays or CT. In LMICs, access to CT or magnetic resonance imaging (MRI) may not always be possible, and so patients benefit from low-cost US imaging. In critically ill patients in the emergency department or ICU, POCUS can enable real-time diagnostic capability at the bedside more quickly and efficiently.

Traditional US imaging that occurs in a suite in the radiology department is usually available during daytime hours but may have limitations in terms of understanding the clinical manifestation/information of patients in the ICU. With the advent of POCUS and with new US machines becoming more portable, the imaging paradigm is shifting away from the traditional model of images being acquired by a trained sonographer/technician and then sent to a radiologist for interpretation towards a model where the patient's treating clinician performs and interprets the US imaging in real time, while developing/monitoring the treatment plan for the patient[34].

As a point-of-care imaging technique, POCUS requires direct interaction between the clinician and the patients to establish a clinical diagnosis or guide a procedure. Thus, it differs from consultative ultrasonography, in which the test is ordered by the clinician, typically performed by a technician or sonographer, and then interpreted by a consultant who is not directly involved with the care of the patient. Since POCUS challenges the traditional approach to ultrasonography and involves the clinician directly, it may well result in a reduction in the use of consultative ultrasonography from the radiologist. However, in low resource settings, the POCUS assessments are still carried out by radiologists or cardiologists.

A 2015 retrospective study [35] showed that the introduction of point-of-care echocardiography performed by intensivists led to a decreased number of comprehensive diagnostic echocardiography studies overall but led to a recommendation to perform full diagnostic echocardiographic studies in 10.7% of patients who had undergone point-of-care studies.

In addition, during the COVID-19 pandemic, it is perhaps unsurprising that the utilization and implementation of lung US took a significant step forward. LUS can be used for triage and

monitoring patients when other radiological modalities such as CT were negatively impacted by overcrowding and understaffing [[36].

Unsurprisingly, there is increasing interest in utilising POCUS tools to evaluate respiratory and skeletal muscle to detect muscle weakness accurately to assist with weaning off of mechanical ventilation and assessing muscle wasting in ICU patients.

The use of US imaging requires extensive expertise on the part of the operator and legal certification. In low resource ICUs, US still relies on radiologists and cardiologists, who are not always available. This lack of availability usually leads to a delay in the examination. Delays in undergoing diagnostic examinations not only add to the frustrations felt by both clinicians and patients but are often associated with adverse outcomes related to the disease that caused the critical illness of patients in the ICU in the first place.

Advances in imaging technology have resulted in a proliferation of hand-held devices that provide good quality images along with increasing affordability. This has transformed US into a truly bedside assessment tool that can be routinely used by clinicians across the world and in a variety of healthcare settings. Complementing this, software-based and AI-based solutions have been developed to provide automatic image analysis and real-time feedback to the user. These aspects will be discussed in the section below.

2.2 Technological background

This section provides a comprehensive literature review of AI-enabled US technology covering three aspects: 1) an introduction to AI applications related to US, 2) the elements involved in training an AI model for US imaging and 3) the current state of translation of those methods to clinical practice in an ICU setting.

2.2.1 An introduction of AI applications related to US

Figure 2.3 illustrates the workflow of a US examination and the potential application of AI algorithms in each task during the examination. Traditionally, some tasks are carried out by clinicians/radiologists/sonographers including scan acquisition, interpretation, measurement, diagnostic and report. Other tasks, such as beamforming and image enhancement, are carried

out by the hardware in US machines without interaction from the user. The workflow starts by a user placing the probe on the patient's body surface, which sends the US waves and receives the reflected waves. The US machine then processes the reflected waves to generate 2D images, which are displayed on the screen. Next, the user follows these images displayed on the screen in real-time for guidance and diagnostics (both of which require real-time interpretation and interaction with the probe to acquired diagnostic quality scans). The user then saves some standard views as static images or clips which are further used for interpretation, measurements, and diagnoses during the examination or often offline.

Advances in both computing power and availability of data have paved the way for the application of AI in US image analysis. AI, as a field defined broadly by the engineering of computerized systems to perform tasks that normally require human intelligence, has substantial potential in the medical imaging field. AI technologies, such as machine learning and deep learning, have been applied in automating the abovementioned tasks to enhance the accuracy, reproducibility, efficiency, and clinical utility of the ultrasound examination workflow.

In the following, I will describe each element in the process of carrying out a US examination, and how AI has been used to improve each. As noted above, a summary of these elements is provided in Figure 2.3 including image acquisition, image processing, image analysis, biometric quantification, workflow optimization and diagnosis/reporting.

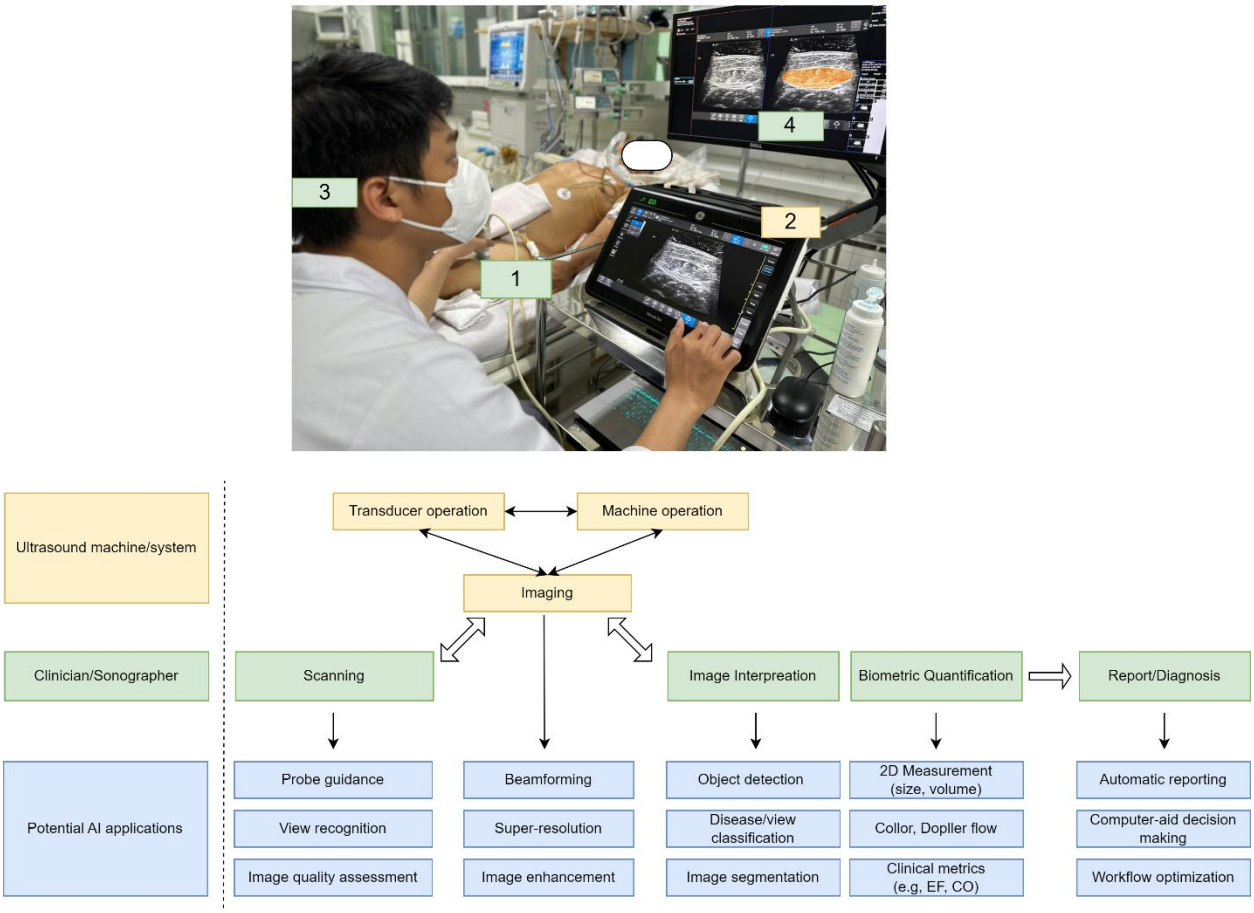


Figure 2.3. US imaging pipeline and potential AI applications at each stage of the pipeline, consisting of scanning, image analysis, biometric quantification and diagnosis/report. Figure was adapted from [37]**.**

Image acquisition: Image acquisition (scanning) is a complex process that requires real-time guidance of the ultrasound probe towards the target anatomical view. This guidance is carried out by expert sonographers/radiologists/clinicians who interpret, in real time, the streamed images presented on a screen. The process ends when the operator decides that images of sufficient quality for the target view have been achieved. AI has been used to improve three elements of image acquisition – probe guidance, view recognition and image quality assessment, as described below:

Real-time probe guidance: Operators must gather several standard views, which are acquired during scanning while positioning the probe at particular angles/positions. AI-

driven real-time feedback involves analyzing the US images as they are being acquired. The AI system can identify suboptimal images due to factors like probe positioning, pressure, and angle. It provides immediate guidance to the sonographer on adjustments needed for optimal image quality. Several works have been done in this field such as AI assisted probe guidance in 2D echocardiography [38], real-time needle guidance to assess blood vessels [39], and fetal plane navigation [40, 41].

View recognition: In US, it can be difficult to determine which part of the region or organ being scanned is currently within the field of view from only a 2D cross-sectional image. Automatic view recognition can provide such information and early work used support vector machines and conventional machine learning, but recently techniques that integrate deep learning have been developed and have greatly improved view recognition performance. For example, recent work has achieved impressive performance in 4 chamber view recognition in echo [42, 43] abdominal and head circumference view recognition in fetal scans [41, 44].

Image quality assessment: While competent operators, such as imaging technicians, sonographers or cardiologists, can find the optimum acoustic window to obtain high-quality images, less-experienced users are more likely to acquire data with suboptimal image quality. AI algorithms for quality control can automatically assess acquired images based on predefined criteria. Images with artifacts, inadequate visibility, or improper settings can be flagged for review before being used for diagnosis or analysis. Several studies have made significant attempts to develop automatic systems that can provide real-time image quality feedback to operators and thus aid the operators in obtaining diagnostic quality views. For example, work has been performed to automated quality assessment for urinary bladder US images [45], cardiac US, breast US and fetal US [[46–49].

Image processing: US image processing involves the utilization of computational methods, such as filtering, segmentation, and feature extraction, to manipulate and analyze the raw ultrasound

data, resulting in improved image quality, anatomical delineation, and the extraction of clinically relevant information for medical diagnosis and research purposes.

Image enhancement: AI algorithms for image enhancement in US typically employ deep learning techniques. Convolutional Neural Networks (CNNs) can be trained to identify noise patterns and artifacts in US images, then generate enhanced versions of the images with reduced noise, improved contrast, and sharper edges [50, 51].

Image analysis: One of the main aims in US examination after image acquisition is to analyse and assess the acquired images. Examples of analysis tasks include image/disease classification, tumor/abnormality detection and delineation of abnormal structures [52]. The similarity between the appearance of normal vs abnormal structures can make this task challenging, particularly for those clinicians with less experience. AI communities have developed a wide range of model architectures to tackle this challenge, as detailed below.

Image (disease) classification: AI models for standard view or disease classification are often based on CNN models such as ResNet, Inception and VGG, or more recently Vision Transformer based models. By training on datasets with both normal and abnormal US images, these models learn to identify subtle patterns and features associated with different types of lesions. Transfer learning, where a pre-trained model on a large dataset is fine-tuned for a specific task, is also used to adapt AI models to US data. Several works have developed DL-based models to classify clinical US severity scores in breast US (BIRADs) and thyroid US (TIRADS) [52–55]. In addition, recently due to the surge in the application of lung US in Covid-19, a number of studies have trained CNN-based models to classify pneumonia lung, Covid-lung and normal lung [56–58]. Hepatocellular carcinoma liver US is also an attractive field of research where several models have been developed to classify whether the tumor is malignant or benign based on liver US images [59, 60]. Those studies show the potential to translate their work into clinical application.

Object detection: The detection of objects of interest is important during US interpretation and is mainly used to find lesions, tumors, nodules and other anatomical structures in US videos/images. The most widely used application is the detection of

tumors/lesions that can support clinicians in optimizing interventions and therapy planning. Breast lesion detection in US was investigated by Yap et al. [61] contrasted three different deep learning methods including LeNet, UNet and pretrained FCN-AlexNet, YOLO (You Only Look Once), Fast R-CNN and the single-shot multibox detector are also popular detection algorithms in breast, thyroid and prostate US [61–63]. To incorporate the spatiotemporal information of US video, RNN-based models and their variants with LSTM and Attention mechanism have been applied in fetal ultrasound and cardiac US [44, 64].

Image segmentation: Manual delineation/contouring in US images is a common task to estimate the size or volume of anatomical structures (e.g. cardiac chambers) or tumors (e.g. liver or breast tumors). Segmentation is a laborious, time-consuming task that often requires repetitive manual tracing, which is subjective, dependent on the experience of the user and prone to variability. To address those challenges, automated image segmentation using AI involves training models to recognize boundaries of different anatomical structures in US images. The U-Net architecture [65] or its variants such as nnUNet [66] are commonly used for semantic segmentation tasks in US imaging [65, 67–69]. By labeling a training dataset with pixel-level annotations, the AI model learns to segment organs, tissues, and other structures accurately. Many studies have shown the potential for automatic segmentation in breast US, fetal US and muscle US [68–72].

Quantitative analysis: AI can automate quantitative analysis by identifying landmarks and making measurements in US images. For example, in cardiac US, AI can measure the dimensions of heart chambers, and blood flow velocities or the size/volume of tumors [61, 72]. These measurements can aid in tracking disease progression and evaluating treatment outcomes.

Clinical metric estimation: Despite their importance in clinical settings, there is often significant variation in the human measurement of clinical parameters such as ejection fraction and cardiac output from echocardiography videos [10, 64, 73], and rectus femoris muscle size, prostate, thyroid, and breast cancer severity scores [55, 63, 74], and this

variation could impact clinical care. Therefore, there is significant interest in improving measurement precision and reproducibility. Studies have shown that AI assistance in interpretation of medical images is more useful to some clinicians than to others and generally provides more benefit to less experienced clinicians [75–77].

Diagnosis/reporting: Diagnosis in US imaging involves the expert interpretation of ultrasound images to identify and characterize anatomical structures, abnormalities, or pathologies within the patient's body. The findings are then typically documented in a detailed diagnostic report, which provides a comprehensive assessment for clinical decision-making and patient management. The application of AI in this context involves leveraging machine learning algorithms to automate image analysis, enhance diagnostic accuracy, and expedite the reporting process, ultimately improving efficiency and healthcare outcomes.

Automatic reporting Conventionally, ultrasound reports are written to record the findings from the US examination. These findings may include important local properties, e.g., boundary conditions and tumor morphologies, biometrics, and other observations. Labor cost and report quality are the two fundamental issues motivating research on automatic medical report generation. First, from the author's personal experience, the average time of writing a single US image report is 5–10 min by experienced doctors. Second, the report quality often significantly varies with doctors' expertise level. To generate highly accurate medical reports automatically, recent research has proposed machine/deep learning-based methods to map from the medical images to diagnostic reports [78–80].

Workflow optimization: AI can automate several steps in the US workflow, such as labeling images, archiving, and report generation. This streamlines the process, reduces human error, and allows medical professionals to focus on the clinical interpretation of images. For example, Zhang al developed an AI framework that was embedded with the picture archiving and communication system (PACS) system to automatic interpreting of echo examination by automatic standard view selection, image segmentation, automatic quantification, disease classification and production of reports that can directly be

interpreted by the attending doctor or radiologist [81]. Recently, multimodal deep learning foundation models [82] were also applied in echocardiography to enable image-to-text, identify clinical changes and identify unique patients across multiple videos, and this work showcased the capabilities of automatic summarization and assessment of cardiac US findings. AI's ability to analyze large datasets can lead to new insights in medical research and may act to enhance computer-aid decision making in daily clinical practice.

It is important to note that while AI holds great promise in US imaging analysis, there are some remaining challenges [12], such as the need for high-quality annotated data, ensuring the interpretability and reliability of AI algorithms, and integrating AI seamlessly into clinical workflows. Ongoing research and collaboration between AI experts and medical professionals are essential for realizing the full potential of AI in US imaging analysis.

2.2.2 Typical AI model development pipeline for US image analysis

In this section, I will explain the pipeline for developing an AI model for automating medical imaging tasks including image classification and segmentation. Classification in US imaging involves assigning a category to an image (or sequence), for example to indicate presence or absence of specific structures or identifying certain patterns or abnormalities. Segmentation in US images involves identifying and outlining the boundaries of specific structures or regions within the image. This can include organ segmentation or tumor/lesion segmentation. Both tasks typically require a training dataset where labels of the (image level or pixel level) categories are available (I.e. supervised learning).

The development of an AI model in healthcare ideally involves interdisciplinary collaboration involving radiologists, clinicians, AI researchers, and regulatory authorities. Such collaboration is essential to ensure that the AI model is safe, effective, and aligned with medical standards and ethical considerations. There are sets of guidelines that have been produced for developing AI models for healthcare, for example the FUTURE-AI guidelines [83] or AI-based prediction model (AIPM) guideline [84].

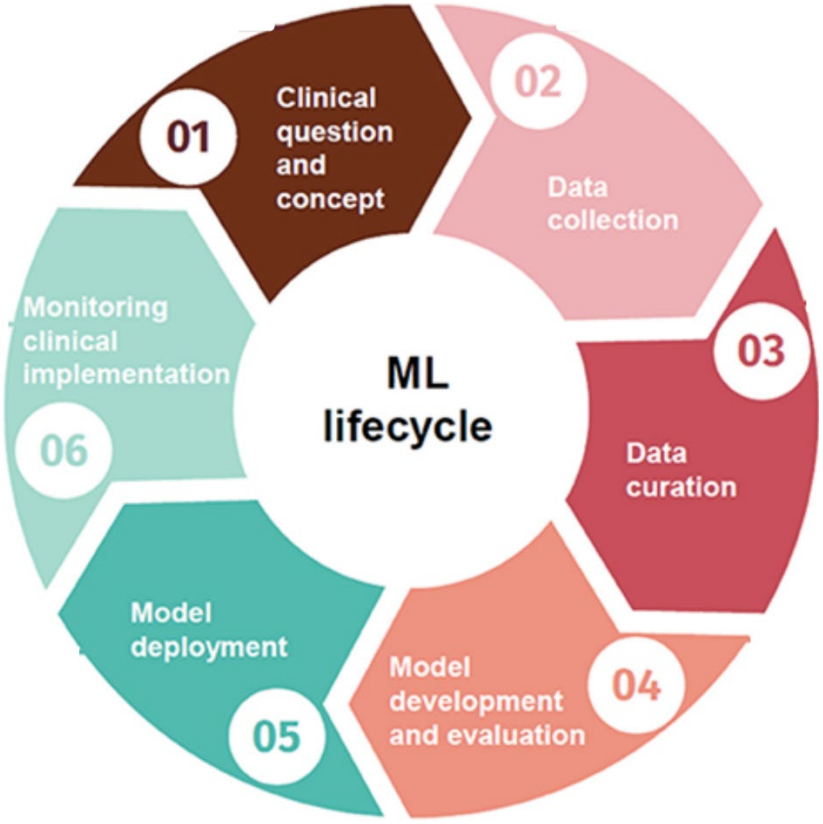


Figure 2.4. Pipeline for development of an AI model in US imaging. Reproduced from [85]

The main elements involved in the model development pipeline are illustrated in Figure *2.*4 and will now be described in turn.

2.2.2.1 Clinical question and concept

Before building an AI application for US imaging, it is critical for researchers including AI engineers, clinicians, radiologists, and other health professionals to identify the clinical needs and the context. These will determine whether an AI solution is relevant and convenient to solve the need and be deployed in the specific context. For example, the clinical needs should be realistic and relevant, improve patients' outcomes or optimize the health care workflow. Defining the clinical problem and its relevance before initiating model development is therefore essential.

2.2.2.2 Data collection

Data collection is an important prerequisite to develop and validate an AI model that performs well. US image data can be collected from various sources, including hospitals, clinics, clinical

research databases, and open-access repositories. It is important to highlight that ethical approval is normally required to collect new US images of patients and/or share the data more widely. In the case of a prospective study, informed consent is necessary. After ethical approval, the relevant data needs to be acquired/accessed, properly de-identified, and securely stored. Nowadays, synthetic data can be generated using a generative model such as a generative adversarial network (GAN), and this is another source of data that is increasingly being used in AI model development. However, using current technology, it can be challenging to generate a sufficiently diverse and realistic set of synthetic data for some applications, especially when abnormalities are considered. Another important aspect that should be considered in data collection is generalizability. For example, if an AI model is trained with images from a European institution and the model is used in an Asian population, then performance may be affected by population differences or disease prevalence. Similarly, if all the imaging training data were acquired using one kind of US machine, it may not work as well on machines from other manufacturers, and this is known as vendor or single-source bias. It is thus advised to use images from multiple diverse sources, because the choice of the dataset will influence the models trained on the data, and the conclusions we can draw from the results. At the community level, we should foster understanding of the datasets' limitations. Good documentation of datasets should describe their characteristics and data collection protocols.

2.2.2.3 Data curation

The next step after collecting data is to perform data preprocessing like resizing images, normalizing pixel values to a standard range, and converting images to a common format (e.g., DICOM, NIFTI,). Next, data augmentation techniques such as rotation, flipping, zooming, and contrast adjustment can be applied to artificially increase the dataset's variability. Thanks to deep learning code packages like PyTorch and TensorFlow, data augmentation can also be performed "on-the-fly" during training to optimize the computer memory usage. Another important step is data annotation. Depending on the specific task of interest, such as image classification or image segmentation, and the type of model being trained, i.e. supervised, semi-supervised or unsupervised, the appropriate data annotations needs to be made. This step links the images to ground-truth information, which can be one or more labels, segmentations, or electronic

phenotype (e.g., biopsy or laboratory results). For tasks like image classification or lesion detection, expert radiologists need to annotate images with labels indicating the presence of abnormalities or pathologies. For segmentation tasks, they will need to manually outline the regions of interest such as anatomical structures or tumors within the US images to create pixel-level annotations. It is noteworthy that radiologists/clinicians often perform segmentation or classification annotation tasks as part of their clinical work, either manually, or using proprietary software provided by the machine manufacturer (GE, Siemens, Philips etc). However, it is often impossible to export those labels to use for AI model development, hence AI practitioners will normally ask clinicians to use other research tools for producing annotations for model development. For example, tools such as VIA VGG annotator [86], ITK-SNAP [87], 3D Slicer [88] are commonly used to annotate medical imaging data for research purposes. For tasks like clinical metric estimation, values such as Left Ventricular Ejection Fraction (LVEF), severity scores like BIRADs and TIRADs can be extracted from the hospital record or US examination report. In general with annotation, quality control is crucial, and multiple annotators may be needed to ensure accuracy and consistency. Publicly available datasets usually come with annotations which can be downloaded directly from the source without further data annotation being needed. It is also crucial to structure the data and annotations in homogenized and machine-readable formats [89].

2.2.2.4 Model development and evaluation

Data Splitting: To ensure generalizability of the AI model, the dataset should be representative of the target data/population. Conventionally, datasets are divided into three subsets: training, validation, and test sets. The training set is used to update model parameters, the validation set is used for model selection and hyperparameter optimization and the test set is used for final validation of the trained model. Empirically, a common split might be 80-10-10 or 70-15-15.

Model Selection and Architecture: Clinicians and AI practitioners work together to choose a model architecture that performs to a desired specification. For example, in classification, supervised learning algorithms are applied, such as Support Vector Machines, Random Forest, or CNNs. These models can take the US image as input and output a vector with probabilities for each possible class. In segmentation, models such as Fully Convolutional Networks, U-Net, are

often employed. These models take the US image as input and output a segmentation mask indicating the structure of interest.

Model Training: Training a deep learning model involves optimizing the model's parameters (weights and biases) to minimize a loss or cost function that measures the difference between the model's predictions and the actual ground truth values. The process starts with initializing the model's weights, either randomly or with pre-trained weights from a related task (transfer learning) [90]. Next, a suitable loss function is selected to quantify the error between the model's predictions and the actual target values. The choice of loss function depends on the problem type (e.g., mean squared error for regression, cross-entropy for classification, Dice loss for segmentation, etc). Next, the model is trained using optimization algorithms like stochastic gradient descent or Adam [91, 92] to update the model's parameters. At this stage its learning rates and other hyperparameters should be optimised during training as described in the paragraph below. The training process may require several iterations of these steps to achieve the desired model performance.

Hyperparameter Tuning: Hyperparameter tuning is a critical step in the development of deep learning models, significantly influencing their performance and efficiency. Hyperparameters, unlike model parameters that are learned during training, are set prior to the training process and include choices such as the learning rate, batch size, number of epochs, and the architecture of the neural network itself, among others. The process of hyperparameter tuning involves experimenting with different sets of hyperparameters to find the combination that yields the best performance on a given task, typically measured by the model's accuracy on a validation dataset. Techniques for hyperparameter tuning range from manual search, where the practitioner adjusts hyperparameters based on experience and intuition, to more sophisticated methods like grid search, random search, and Bayesian optimization, which systematically explore the hyperparameter space. Effective hyperparameter tuning can lead to substantial improvements in model performance, making it a vital aspect of the deep learning workflow.

Model Evaluation: To evaluate a deep learning model, the test set is employed, and various metrics are used to measure model performance and assess its generalization on the test set. Additional steps include visualization, comparisons with baselines or state-of-the-art methods,

cross-validation for robustness, ablation studies to understand model components, and examining potential biases and fairness issues. Common evaluation metrics include accuracy, sensitivity, specificity, precision, F1-score, or area under the ROC curve (AUC-ROC). Accuracy measures the percentage of the input data that were classified correctly. It is a simple measure used in multiple scenarios if there is no class imbalance (i.e., one class has a higher number of samples compared with the others). One of the drawbacks of using accuracy as the metric is that there is a knowledge loss when measuring False Positive and False Negative observations. Therefore, Specificity and Sensitivity are widely used for measuring the performance of the model, this time taking into consideration a possible class imbalance. In order to assess the performance of an algorithm and to understand where there might be a miss-classification issue, a visual representation called Confusion Matrix is used. This specific table is typically used to describe the performance of a supervised classification model. Each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class (or vice versa). This way, an AI engineer can have a wider overview of the performance of the model, or which classes the model is underperforming in. From sensitivity and specificity, we can extract a performance plot representation called the ROC curve. It is created by plotting the sensitivity against one minus the specificity at various threshold settings. From the ROC curve, the AUC-ROC can be extracted as a performance metric. ROC analysis is related in a direct and natural way to a cost/benefit analysis of diagnostic decision making.

2.2.2.5 Model deployment

Model deployment involves first evaluating the final model on the test set to ensure its generalization to new, unseen data. If the model meets predefined performance criteria or clinical baseline criteria, it should then be prepared for deployment in a clinical or research environment. This process may include optimizations for inference speed and memory usage.

2.2.2.6 Monitoring and clinical implementation

Continuous Monitoring and Improvement: After deployment, we should continue to monitor the model's performance in real-world scenarios. Feedback should be gathered from medical professionals and patients to identify areas for improvement. Regular updates to the model

should be considered using new data to account for changes in patient populations, technological advancements, and evolving medical knowledge.

Ethical Considerations and Validation Regulatory guidelines and ethical standards for medical AI systems should be followed to ensure that the AI model's decisions can be validated and justified clinically.

2.2.3 Implementation of AI-based US imaging analysis

In this section I describe the translational aspects of AI-based US imaging analysis. I start by emphasising the gap between technical development and clinical implementation of AI in US imaging. The literature on prospective evaluation studies and the process for FDA[1] approval of AI software for automatic US analysis in clinical settings are also described.

2.2.3.1 The research-implementation gap

In recent years, there has been a rapid increase in the amount of research that utilizes machine learning in all aspects of healthcare including disease screening, medical image interpretation and clinical workflow optimization. Much of this research has demonstrated impressive performance which is starting to match or even exceed the performance of human experts. This has been made possible as a result of the increasing amounts of data generated and recorded from routine clinical care, the available computational processing power, and advances in AI techniques such as deep learning. However, most studies to date have been retrospective and there has been a relatively limited amount of research investigating the applicability of healthcare AI in clinical practice. This research-practice gap can be explained by several reasons. Specifically, AI algorithms may be subject to technical issues, such as model overfitting, dataset bias, and lack of generalizability, limiting the potential translation of AI research into clinical practice [93]. In addition, practical implementation of AI applications in healthcare can be incredibly challenging. For example, the deployment of AI in US imaging requires real-time acquisition and interpretation and should consider factors such as the clinicians-computer interaction, clinical workflow optimization and trust in AI algorithms. Without randomized control trials, reliable evidence must be gathered before we can safely use AI tools in clinical

[1] Vietnamese Ministry of Health approves the use of FDA or CE-marked devices in patients without further testing.

practice. Furthermore, other issues need to be considered, such as data sharing and privacy, lack of algorithm transparency, the changing nature of healthcare work, continuing monitoring and maintenance of the AI model, financial concerns, and the demanding regulatory environment [93]. For example, Robert et al. reviewed 62 published studies on machine learning for COVID but found none with potential for clinical use due to methodological flaws and/or underlying biases [94]. The "inconvenient truth" of AI in healthcare was pointedly described as "at present the algorithms that feature prominently in research literature are in fact not, for the most part, executable at the front lines of clinical practice" [95].

2.2.3.2 Prospective studies using AI in a clinical context

When searching the literature, there are limited papers describing randomized controlled trials (RCTs) to test AI algorithms and even fewer on algorithms related to US. Below are some of the prospective studies that have been performed on AI in US applications, broken down by bodily organs.

Lung US: Kuroda et al. [96], investigated AI assisted LUS for COVID-19 patients in the detection of pneumonia. Patients were validated for pneumonia using CT scan. The AI assisted 12 zone POCUS showed to be highly accurate, sensitive and specific (94.5%, 92.3% and 100%, respectively). It is worth noting that accuracy, sensitivity, and specificity decreased with an 8 zone US and a single zone US. The limitations were that it was a single centre study with a small population group. Finally, the algorithm was only able to detect B-lines and no other lung features, thus, this did not represent sufficient information to accurately diagnose the pathology. Dave et al. [97] conducted a study in the ICU to prospectively validate a real-time deep learning algorithm in classification of normal lung versus B-line patterns showing high accuracy, sensitivity and specificity.

Fetal US: Pluym et al. conducted a prospective study using an AI based tool named SonoCNS Fetal Brain [98]. Participants had two US scans by sonographers and physicians to measure biparietal diameter, head circumference, transcerebellar diameter, cisterna magna and posterior horn of the lateral ventricle with and without SonoCNS. The findings showed SonoCNS was able to reliably identify biparietal diameter and head circumference, however, it was less reliable for the

other three measurements. A limitation of the study was that there was no randomization between the clinicians and patients. In addition, Matthew et al. [99] tested an AI tool which automated image acquisition, biometric measurement, and report production for mid-trimester screenings. Although the results showed no difference in the biometric measurement between the AI assisted group and non-AI assisted group, it was able to save on average 7.62 minutes. The limitations were the small sample size of 23 patients and that sonographers were aware that the patients had a low risk of fetal anomalies.

Liver US: Thodsawit et al. [100] evaluated an AI algorithm developed to detect focal liver lesions in US in a randomized controlled study. Patients with lesions were randomized into an expert and non-expert group and then further divided into AI assisted and non-AI assisted. The study found when assisted with the AI algorithm, the rate of detection of lesions improved significantly in the non-expert group from 21.4% to 36%, but the improvement was not significant for the expert group (63.3% to 66.7%). The authors concluded the results were not overwhelming and the software was not ready to be integrated in clinical practice. A limitation was that the study was a single-centre study which compared to multi-centre studies tend to have results which are not as generalizable or reproducible.

Cardiac US (echocardiography): Papadopoulou et al. [101] demonstrated that an AI algorithm integrated into a handheld US device was able to calculate LVEF with an accuracy of 88% when compared to the standard manual biplane Simpson's method. However, to provide sufficient evidence for this to be incorporated into clinical practice, a randomized control trial should be conducted. Gohar et al. [102] compared an AI assisted echocardiographic tool to point-of-care US experts in a prospective study and found that the tool was able to calculate similar results for automatic LVEF (weighted Cohen's-Kappa of 0.498; $p < 0.001$), automatic inferior vena cava measurement (weighted Cohen's-Kappa of 0.536; $p = 0.009$) and automatic velocity time interval (weighted Cohen's-Kappa of 0.655; $p = 0.024$). The limitations were that a single expert calculated all the readings, the sample size was small (44 patients) and only a few patients had cardiac pathologies. These issues need to be addressed to confidently say that the tool was able to identify pathological parameters. Zhai et al. [103] conducted a study comparing left ventricular outflow tract velocity time interval (LVOT-VTI) measurements between an AI algorithm and

intensive care unit doctors. Measurements were taken by doctors using the manual method and compared to the measurement calculated by the AI algorithm. These were then compared to expert measurements by expert sonographers. Results found that there was not a statistically significant difference between the manual and AI-assisted readings of LVOT-VTI (8.8% ± 1.3% vs. 10% ± 2%, p = 0.6517). The study had a few limitations, however. It was a single centre study and therefore, the generalizability should be explored further. Furthermore, the time limit for ICU doctors to take a measurement was 2 hours after admission; however, it was 24 hours for expert sonographers. During this time, the patient's condition may have changed which raises issues regarding the validity of the results. A prospective study by Varudo et al. [104] on 95 patients testing machine learning algorithms found the real-time assessment of LVEF in critically ill patients strongly correlated with values measured by experts. Both novices and expert clinicians calculated LVEF using the AI algorithm and their values were compared to reference values by experts. The results showed that in the expert group, there was a strong correlation of r = 0.86, whilst in the novice group the correlation was r = 0.81. Regarding the sensitivity and specificity, they were 70% and 98% in the expert group and 73% and 98% in the novice group respectively. However, the weakness of the study was that only 7 clinicians participated in the study, and it focused on patients that were hemodynamically stable. He et al. [105] a model named EchoNet-POCUS which was tested prospectively on 47 patients in the emergency department. The deep learning model was trained to interpret cardiac function and video quality. The results of the EchoNet-POCUS were compared with an expert's interpretation of the echocardiogram. They found that they achieved an AUROC of 0.96 for predicting low ejection fraction and an AUROC of 0.89 for predicting video quality. Some limitations of this study include a small sample size and that it was a single centre study.

Although these AI algorithms may prove useful in real-world settings, the issue of thorough and realistic validation remains a concern. In addition, these models may not be able to generalize easily to new geographic locations, patient populations, disease distributions, and changes in imaging technology without requiring substantial engineering effort and/or new data.

It is important to note that the growth in the amount of research on AI for US imaging does not necessarily lead to clinical progress. The huge potential of healthcare AI applications can only be

realized when they have been integrated into clinical routine workflows. To ensure effective adoption, an AI application should provide scientific evidence for its effectiveness relative to the gold standard, e.g. a randomized controlled trial.

U.S. Food and Drug Administration (FDA) approval AI software for US

Recently, the FDA released the list of AI and ML-enabled medical devices, of which there are 15 AI-enabled US softwares listed as class II/510(k) [106]. Most of the approved softwares are in the heart (6), thyroid (2), breast (2) US and other organs (5). LVEF is the main application of the listed softwares in echocardiography. For example, Caption Health, Ultromics, Kosmos AI, and DiA have developed deep learning-based LV segmentation tools which can derive LV volume and subsequently LVEF. Kosmos AI's automated EF can be obtained by the user in less than 15-20 seconds (including acquiring images) and using cart-based echo machines requires the equivalent time for calculating EF in the range of 2-3 mins. Caption Health had previously reported a Root Mean Square Deviation (RMSD) of 8.29% vs. the RMSD of 7.51% reported by Kosmos. In addition, Ultromics also released a new product that automatic calculation of Global Longitudinal Strain and Heart Failure preserved Ejection Fraction (HFpEF) detection using single 4 chamber view videos. Caption Health pioneered real-time guidance for capturing standard echo views with automatic real-time view detection and quality assessment, and a clinical study was evaluated on novice users [107]. The KoiosDS and ClearView software were described as ML-based software that were able to classify user-selected region of interests (ROIs) in breast US images using the BIRADS category. Samsung and AmCad Biomed released their deep learning-based software for automatic analysis of US images, which provides detailed information on the quantification and visualization of US characteristics of thyroid nodules. Up until now, there have been very few reports on post deployment monitoring of these AI softwares, and this is a key aspect for monitoring their performances that we should keep a close eye on in the near future [108].

2.3 Potential of AI-assisted US imaging in a resource limited ICU

AI has the potential to dramatically increase the effectiveness of US and in particular POCUS, primarily by assisting clinicians in interpretation and clinical metric estimation, which are some of the common obstacles to POCUS implementation.

Although the penetration of deep learning in POCUS imaging is currently estimated to be very small, the readiness of point-of-care machines and the potential of the AI technology indicate that further translation into clinical practice is likely to occur. Many current medical imaging AI applications are designed for radiologists, and there is a need to shift research effort more towards the use of medical-imaging AI for non-radiologist clinicians, or more specifically, intensivists and nurses in the case of POCUS. This trend presents an opportunity for improving access to POCUS, improving accuracy, improving confidence in interpretation and reducing time in low-resource settings, ICUs and emergency departments, where there is often a lack of around-the-clock radiology coverage. Very few randomized controlled trials have shown the safety and effectiveness of such AI algorithms in US, and the lack of real-world evaluation of AI systems for non-radiologist users can pose a substantial risk to patients and clinicians.

The application of AI in low resource settings is much different from other areas where AI is being currently applied, such as high standard ICUs in the United States or United Kingdom. AI enables new discoveries and improved processes in the entire health care continuum; ethical, governance, and regulatory considerations are critical in the design, implementation, and integration of every component of the AI applications and systems. Because of concerns about both utility and safety, new AI applications will generally have to adhere to the same standards applied to daily medical practice. This will require a level of rigor in testing similar to that used in other areas of medicine, but it also can present challenges, such as the “dataset shift” that can result when there is a mismatch between the dataset with which an AI system was developed (usually High-Income Countries) and the data on which it is being deployed, for example LMICs.

2.3.1 How can AI support the use of US imaging in resource-limited ICUs?

As discussed in Section 2.1 and Section 2.2, US has advantages and could be a great candidate for AI-assisted image analysis and acquisition, which could help inexperienced users. In that sense, AI methods can help towards the use, interpretation and quantification of US imaging, tasks that would normally be taken up by experts. Therefore, the main role of AI in LMICs, where the availability of experts is scarce, formal training is lacking, and staff normally care for a very large number of patients, is to make US examinations easier, quicker and more reliable, as well as more accessible for less experienced operators. Clinical AI applications may assist in the

acquisition of standardized images independent of the operator, guiding both sonographers and non-experts to acquire images with diagnostic quality.

Furthermore, the success of AI applications in US requires knowledge of local infrastructures, clear usability requirements and access to adequate training data via new data collection protocol.

In order to be clinically translatable, research into AI-assisted POCUS needs to have two components. First, the research must be structured to answer a clinically meaningful question in a way that can influence the behavior of the health professional and ultimately, lead to an improvement in outcomes for the clinician's task or the outcome of patient. Second, the developed AI tool must be definable, scalable, and applicable to the problem at hand. It must not be influenced by factors outside the domain of the problem and must yield outcomes that can be applied to similar clinical problems across a wide range of populations.

To the best of the author's knowledge, there are still no studies focused on the real-time implementation of AI in US in LMIC ICUs.

2.3.2 Challenges for the adoption of AI-enabled US in resource-limited ICUs

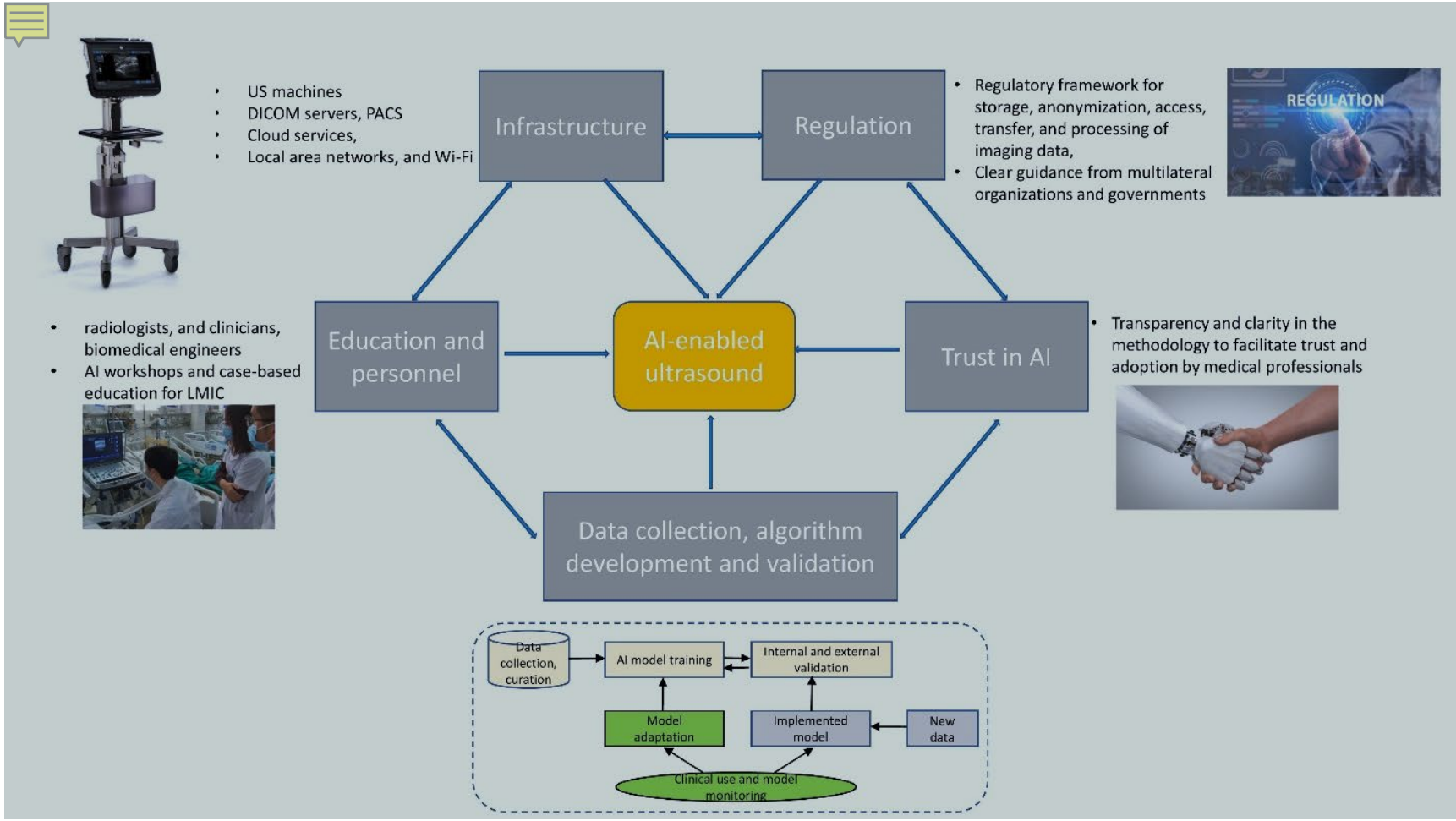


Figure 2.5. Challenges for the adoption of AI in US in resource-limited settings

Whilst AI has advanced rapidly over the past decade, translating AI research into a useful clinical tool has faced unexpected challenges along the way, as the expectations contrasted with the reality. The major impediments for scale-up and adoption are illustrated in Figure 2.5 and described below.

a. **Infrastructure -** The use of AI in healthcare in wealthy countries is increasing, whereas most LMICs do not even have basic digital infrastructure in their healthcare systems. The absence or insufficiency of such infrastructure (US machines, DICOM servers, Picture Archiving and Communication System (PACS), cloud services, local area networks, and Wi-Fi) in most LMICs represents a significant challenge in AI implementation.
b. **Education and personnel –** Effective AI integration into clinical workflows in LMIC's health institutions requires its adoption by physicians, radiologists, and clinicians and close collaboration with biomedical engineers with expertise in AI. Clinical professionals should be trained to interpret and factor in AI-enabled tools together with the rest of the clinical information for effective patient management. An initiative that is being implemented with positive results is to establish AI workshops and case-based education for LMIC hospital partners to try out AI algorithms. These workshops should be integrated with other education and infrastructure deployments, to incorporate AI education into resource-limiting settings, and to provide feedback to practitioners.
c. **Data collection, algorithm development and validation -** A significant challenge to accelerating the use of AI algorithms in LMICs relates to the quantity and quality of the available data. Algorithms are currently trained mostly with data from HICs, which may not be representative of LMIC populations. This means that even if AI algorithms are approved commercially, the efficacy and translatability of such algorithms may still be inhibited by the lack of data from LMICs. This should be addressed both through novel computational methods that can learn from small amounts of data, and through improvements in local resources and infrastructure that allow for more ambitious data collection initiatives. In addition, the metrics used to evaluate the model performance sometimes do not reflect clinical applicability and are not easily understandable by many healthcare professionals, for example Dice similarity score in image segmentation.

However, none of the commonly used metrics ultimately reflect what is most important to patients, namely whether the use of the model results in a beneficial change in patient care.

d. **Retrospective versus prospective studies -** While existing studies have encompassed very large numbers of patients with extensive benchmarking against expert performance, the vast majority of studies have been retrospective, meaning that they use historically labelled data to train and test algorithms. Only through prospective studies will we begin to understand the true utility of AI systems, as performance is likely to be worse when encountering real-world data that differs from that encountered in algorithm training.
e. **Regulation –** Because AI in US is an emerging field, many LMICs also lack a comprehensive and consistent regulatory framework for storage, anonymization, access, transfer, and processing of imaging data, particularly for AI companies outside of the health facilities. It is crucial to create clear guidance from multilateral organizations and governments on how and when regulation on AI applications should be applied.
f. **Trust in AI -** When it comes to AI-based approaches for use in medical decision making, it is important to provide transparency and clarity in the methodology to facilitate trust and adoption by medical professionals.

to address these challenges and impediments, and to explore and appropriately accelerate the use of AI-assisted medical imaging in global health, it is critical to develop translational networks and innovations in LMICs, which will allow development and scalability of the new approaches. It is crucial to understand the challenges and identify gaps in the settings where AI approaches will be implemented, therefore, the strategy of promoting local innovations is more likely to be sustainable.

2.4 Summary

Previous work in utilization of AI-enabled ultrasound in critical patients where resources are limited is scarce. All methods identified in the existing literature are insufficient for optimal use in a LMIC ICU setting, where information must be extracted from imaging at the time of the scan, be reliable and contain minimal variation despite a variety of non-expert operators. For instance,

in automated LUS, no published method can classify all clinically required features. In muscle US, no clinical evaluation study has been carried out. More generally, none of the abovementioned studies supply information on how to implement the models for clinical use in real-time in the ICU, neither do they investigate what accuracy is required in the clinic. Specifically, previous methods are trained on images or frames, however real-time US is essentially video data, where the dynamics of the imaging may play an important role in how operators use the system and how information is inferred. Additionally, most of the previous works use models trained on data from HICs, and the performance in data from LMIC ICUs is unknown [93, 109–111].

In addition, clinical validation is needed to evaluate the performance and impact of AI algorithms in diverse clinical settings and environments with users of varied experience and backgrounds.

To date, most healthcare AI studies have been proof-of-concept studies that have focused on AI algorithm development and validation using retrospective clinical datasets. In contrast, only a handful of studies have implemented and evaluated AI in a clinical environment. To ensure safe adoption, however, an AI application should provide solid scientific evidence for its effectiveness relative to the standard of care.

Within this context, the works presented in this thesis describe how I have closely worked with a multi-disciplinary team including engineers, clinicians, radiologists, and healthcare providers to address the aforementioned unmet needs and to demonstrate the potential of AI in US in real-life clinical settings at an ICU in Vietnam.

3 Clinical protocol, data collection and annotation

Chapter overview

This chapter first outlines the motivation for data collection in a LMIC in Section 3.1 and provides an overview of the clinical study protocol in Section 3.2. This study was conducted to collect data and test the AI-assisted tools throughout this thesis. Section 3.3 describes the datasets that were collected in the clinical study and where available, public datasets that were used. Section 3.4 covers the annotation protocol for each application: lung US classification, muscle US segmentation and cardiac US ejection fraction estimation.

This clinical study was approved by the Oxford Tropical Research Ethics Committee (OxTREC) and local hospitals Ethics Committee including Hospital for Tropical Diseases at Ho Chi Minh city (HTD) and National Hospital for Tropical Diseases in Hanoi (NHTD).

3.1 Introduction

Infectious diseases such as dengue, tetanus, and tuberculosis [112–114], are one of the main causes of hospital admission in LMICs. These diseases continue to account for a significant burden of disease in LMICs. As a consequence of climate change and the evolution of pathogens, infectious diseases (for example, dengue [115, 116] pose significant problems not only in LMICs but also in HICs countries. Ultrasound imaging has been widely used in diagnosis and monitoring infectious diseases patients due to its portability, relatively inexpensive and safety. For example, ultrasound data can be used to diagnose respiratory diseases, abdominal infections, and identify cardiac related problems, but often this cannot be done effectively in LMICs due to a lack of available expertise [117, 118]. The availability and portability of ultrasound machines has meant that large datasets are now available that can be used to train AI algorithms. AI has the potential to play an important role in many common health challenges related to infectious diseases in both high-income settings and LMICs [119, 120]. However, these data have mostly been acquired in high-income settings and there is currently very limited data collected from patients in LMICs. For example, there is a large amount of lung ultrasound of COVID-19 patients in the US and

Europe but there is currently no lung ultrasound data of dengue patients on the publicly available dataset [121]. The lack of LMICs data for training AI model poses a risk to the fairness, generalizability and efficiency of AI tool. Obtaining an adequate amount of training video/image datasets for AI algorithms can be challenging, particularly in resource-limited settings where there is often an absence or insufficiency of infrastructure to acquire and store such data [13, 109]. As a result, a significant challenge to accelerating the use of AI algorithms in LMICs relates to the quantity and quality of the available data. Algorithms are currently trained mostly with data from high-income countries (HICs), which may not be representative of LMIC populations. This means that even if algorithms are approved commercially, the efficacy and translatability of such algorithms may still be inhibited by the lack of data from LMICs. Several techniques have been proposed to address this challenge, such as transfer learning that allows the training of AI models in limited-data scenarios [90, 122, 123]. Data from local hospitals plays a critical role in AI/ML development, and the quality as well as quantity of data have a large influence on AI model development, from data collection, training, testing, and deployment in real life. Clinical data is often diverse, noisy, and ever-changing; thus, deploying a model in the real world requires a data-centric approach in order to have a reliable system.

To address the lack of ultrasound data in LMICs, particularly for patients diagnosed with dengue, sepsis, tetanus, and central nervous system infection, as part of this thesis, I collected, curated, and labelled new data (cardiac, lung and muscle ultrasound), and (where relevant) made use of data available from other publicly available sources. The ethical approvals, inform consent forms, and case report form to recruit ICU patients in HTD are attached in Appendix A and B.

3.2 Clinical study protocol for data collection

In this section, I briefly describe the clinical study protocol for data collection including study design, inclusion/exclusion criteria, informed consent process, study procedures and data management plan. The study covers data acquisition for all three clinical applications (lung, muscle, and heart). The study was conducted in collaboration with the ICU, Emergency Department (ED), Viet Anh (Central Nervous System Infection department) and Radiological

departments at two tertiary hospitals in Vietnam. The members for each team are detailed in Section 1.4.

3.2.1 Study design

I conducted a prospective observational study of patients aged ≥16 years from June 2020 until December 2023. The patients in the study were recruited into one of two groups. Patients participating in Group 1 were selected for cardiac and lung US and recruited at the HTD in Ho Chi Minh City only (Adult ICU, Emergency Department, Viet Anh Ward). Patients participating in Group 2 were selected for muscle ultrasound and were recruited at HTD (AICU and Viet Anh Ward) and the National Hospital for Tropical Diseases at Hanoi (Adult ICU).

3.2.2 Inclusion/Exclusion criteria

Inclusion criteria

All patients admitted to participating wards were checked for eligibility to participate in the study using the following criteria.

Group 1

- Age ≥16 years
- Written informed consent
- Clinical diagnosis of sepsis, dengue, or tetanus

Group 2

- Age ≥16 years
- Written informed consent
- Diagnosis of Grade 1 or 2 Tuberculosis meningitis (TBM) or Ablett Grade 3 or 4 tetanus or CNS infection
- Within 72 hours of ICU admission
- Duration of ICU stay expected at least 5 days.

Exclusion criteria

- Informed consent not given
- Contraindication to ultrasound scan (allergic to ultrasound gel)

3.2.3 Informed consent process

Informed consent was taken by the attending doctors in Vietnamese, all of whom received specific training in the study and Good Clinical Practice and were authorized to take consent by the principal investigator. These doctors also assessed whether the patient had the mental capacity to provide informed consent. If the doctor judged that the patient did not have this capacity, they attempted to obtain informed consent from the patient's representative (usually a relative). If the patient lacked capacity and no representative was available, the patient was not included in the study. It was made completely and unambiguously clear that the patient (or their representative) was free to refuse to participate in all or any aspect of the study, at any time and for any reason, without incurring any penalty or affecting their treatment. Those who refused consent were treated as per the best available standard of care and did not have any study related procedures performed.

The patient or their representative was required to personally sign and date two copies of the latest approved version of the informed consent form. The study staff also signed and dated the two copies. The patient/ representative received one copy.

If the patient/representative was illiterate, a witness who was not a member of the study staff was present during the informed consent discussion. The informed consent form was read to the patient/representative in the presence of the witness. If the patient/representative agreed to participate, the form was signed and dated by the witness. If the patient was a minor (defined as < 18 years of age) assent was obtained in addition to parental or guardian consent.

3.2.4 Study procedures

Eligible patients/patient representatives were approached by attending doctors for informed consent after the patient had been stabilized. No ultrasound assessments were performed until the patient had been stabilized and the study did not delay critical medical care. Further details on the informed consent process are given in section 3.2.3.

3.2.4.1 Group 1 (cardiac and lung US)

Patients with dengue, sepsis or tetanus were eligible for this group as they were likely to have cardiac and lung complications as the primary reason for ICU admission or its in-hospital complications. Once informed consent had been given, standardized clinical information was recorded in the Case Report From (CRF) as follows: clinical diagnosis, demographic data, comorbidities, admission medication, height, weight and outcomes.

Patients underwent ultrasound examinations according to a predefined standard operating procedure (SOP), with a total of 3 ultrasound examinations over 5 days. We aimed to recruit approximately equal numbers of patients with tetanus, dengue and sepsis and with varied severities to provide a diverse set of data for training and evaluation.fThe equal numbers of patient between these conditions reflect the real-life epidemiology of this ICU and is representative of the relative incidence of infectious diseases in low resource settings in tropical countries like Vietnam. Ultrasound images were taken by appropriately trained staff (radiologists, intensivists, sonographers) according to a predefined SOP using Sonosite M Turbo, GE Venue Go, GE VIVID IQ, or similar FDA approved ultrasound machines as shown in Figure 3.1. For cardiac US data, the wide-band phased array 3Sc probe (GE Venue Go, GE VIVID IQ) was used. For lung US data, the wide-band convex array probe C1-5-RS (GE Venue Go, GE VIVID IQ, Sonosite M Turbo) was used. Video of the ultrasound examinations were recorded (both in DICOM and .mp4 format). Images were used for clinical assessment although results were available on request by ward staff. The time taken to perform the scans was also recorded.

Figure 3.1. Machines and probes used for cardiac US and lung US examinations.

The recorded videos from this group were used for AI model development in Chapters 4 and 6. To acquire data for AI model validation, the ultrasound system was connected to a computer using a HDMI connection through a low-latency frame grabber, and the computer showed, in real time, the acquired images and the outputs of the AI algorithms to assist the operator. The details of each developed real-time AI-assisted US system will be explained in each following chapter 4, 5, and 6. It is important to note that the AI assistance was for evaluation purposes only and did not replace the clinical examination nor was it used under any circumstance to alter the patient management or clinical pathway.

In both data collection and validation phases, images/videos were also verified by a further experienced independent cardiologist and a lung US expert blinded to the method of image acquisition.

3.2.4.2 Group 2 (muscle US)

Patients with tetanus (Abblett grade 3 or 4) or with TBM (grade 1 or 2) or with CNS infection were eligible for this group, as these patients usually suffer from muscle wasting due to prolonged ICU stay. Once informed consent had been given (using the same process as Group 1), clinical

information was taken as in Group 1. Disease specific scores (TBM and Tetanus Severity Scores) were collected. Similar to Group 1, ultrasound images were taken by appropriately trained radiologists according to a predefined SOP using the wide-band 12L-RS linear array probe. The full SOP for muscle US is described in detail in Section 5.1. In addition to the ultrasound machines used in Group 1, in Group 2 the handheld Lumify (Phillips) probe was also used for muscle US scans in NHTD as shown in Figure 3.2. The use of multiple vendors will potentially improve generalizability of the developed models.

Figure 3.2. Machines and probes used for muscle US examination.

Measurements of rectus femoris cross sectional area (RFCSA) from standard views and videos were taken on 3 occasions: day 1 of the study, between days 2-7 and at ICU discharge. Additionally, functional outcome was measured at discharge and by telephone questionnaire at 6 months using EQ5D[124] and SF36-V1 [125]. These are standard internationally recognized assessment scores previously used in studies at HTD and allow comparison with previous studies in other patient groups at HTD as well as internationally. Videos of the ultrasound examinations were recorded (both in DICOM and .mp4 format).

3.2.5 Ethical considerations

This protocol and the associated informed consent form were submitted to the ethics committee of the participating hospitals and the (OxTREC. The investigator submitted and, where necessary, obtained approval from the above parties for all substantial amendments to the original approved documents.

The investigator ensured that this study was conducted in accordance with the principles of the Declaration of Helsinki (Seoul 2008) [126] and the terms of approval of the appropriate ethics committees.

The investigator ensured that this study was conducted in full conformity with relevant regulations and with the ICH Guidelines for Good Clinical Practice July 1996.

3.2.6 Data management

The study used a pre-defined data management plan following OUCRU's policies. Source documents were generated during the study by the site ward and study staff. Source documents included all original recordings of observations or notations of clinical activities, and all reports and records necessary for the evaluation and reconstruction of the study. Source documents included, but were not limited to, the participant's medical records, research case record forms (paper or electronic), echocardiography and ultrasound images, progress notes, pharmacy records, and any other similar reports or records of procedures performed during the patient's participation in the study.

Access to applicable source documents was required for study purposes. The site investigators were responsible for maintaining any source documentation related to the study. Source documentation supported the data collected on the CRF when the CRF was not the original site of recording, or else the reason for the difference was documented. Source documentation was available for review or audit by the sponsor or designee and any applicable regulatory authorities.

CRFs were used as a data collection tool. The site Investigators were responsible for maintaining accurate, complete and up-to-date records. These forms were completed on an ongoing basis during the study by authorized individuals.

Corrections to paper CRFs were initialed and dated by the person making the correction and did not obliterate the original entry. All CRFs were reviewed by the designated study staff and signed as required with written or electronic signature, as appropriate.

Selected study members were trained on how to enter all clinical data as source information from the CRFs and from laboratory source documents into an internet-based computerized data entry system called CliRes hosted by OUCRU. Source documents and electronic data were verified according to the Data Management Plan.

Data sharing was performed in accordance with Welcome Trust and OUCRU policy in a timely manner.

Record retention

The investigator (myself) is responsible for retaining all essential records for at least 15 years after the completion of the study. Original paper documents will be maintained for a minimum of 5 years and electronic documents retained thereafter. All stored records are to be kept secure and confidential.

Data use and storage

Images and clips were collected for this study as stated in the protocol and stored in the OUCRU server in an anonymized form. The participants were identified only by a study specific participant number and/or code in the database. The name and any other identifying detail were NOT included in any study data electronic file. All captured data were appropriately anonymized in agreement with the EU GDPR guidelines before they were shared with collaborators or used for research.

3.3 Datasets

In total 7 healthcare professionals including 4 radiologists, 1 cardiologist, 1 intensivist and 1 operator were involved in the data collection for this study. Two external ultrasound experts have validated a subset of the dataset to ensure the quality of the data.

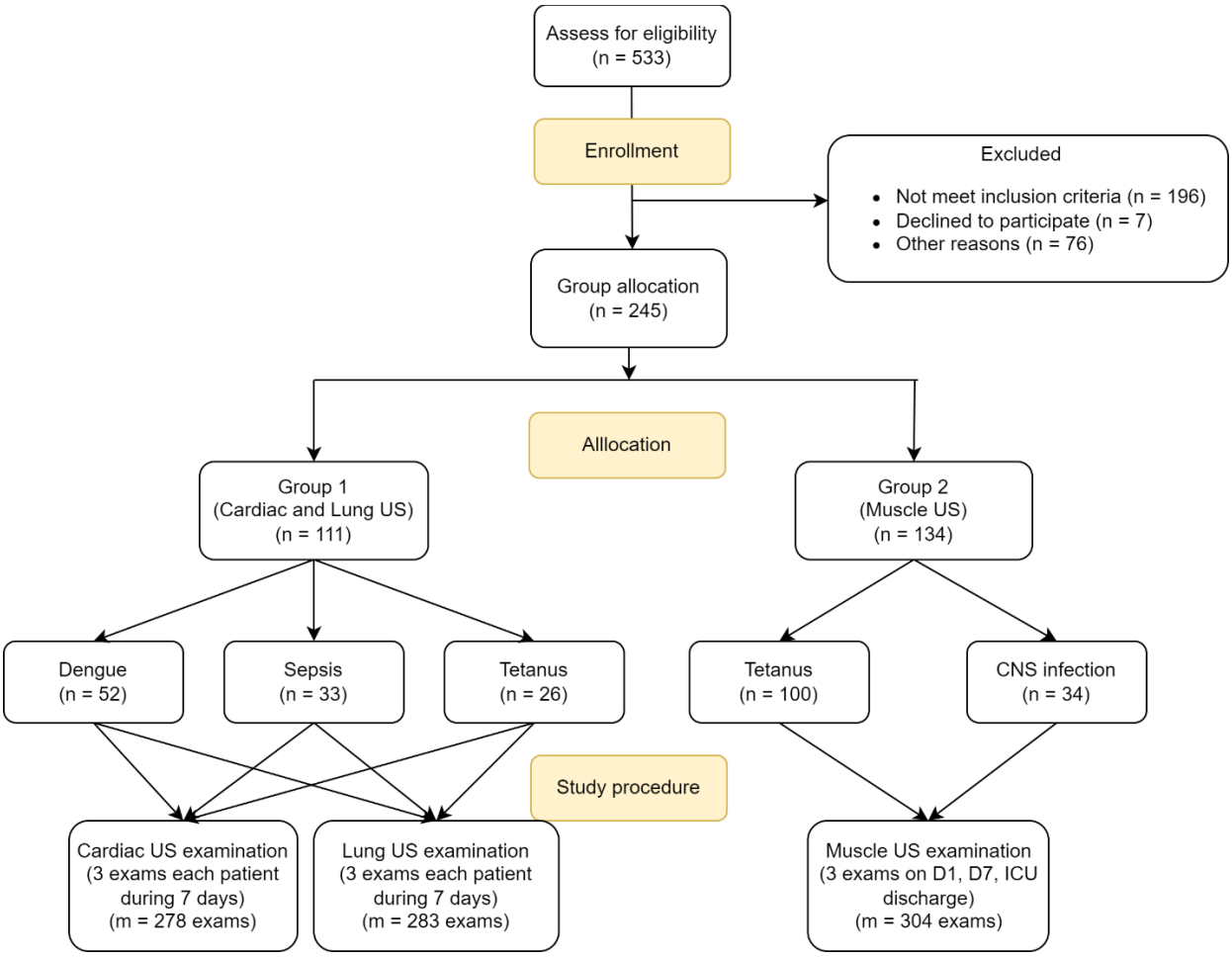


Figure 3.3. The flowchart of the clinical study consists of 4 mains step: screening, enrollment, allocation, and study procedure.

This prospective study was conducted and started recruiting patients in HTD from June 2020. Due to the COVID outbreak in Ho Chi Minh City, the study was paused twice (April 2020 to June 2020, and June 2021 to December 2021), and there was also a pause between Feb 2022 to Jan 2023 when I was in the United Kingdom. After the COVID-19 outbreak, the number of patients admitted to HTD reduced significantly compared to before the outbreak. However, the study achieved 50% target recruitment with 111/200 patients for cardiac and lung ultrasound, and 134/160 patients for muscle ultrasound. The recruitment flowchart and the progress are shown in Figures 3.3 and 3.4, respectively. The demographics of the patients are given in **Table 3.1**.

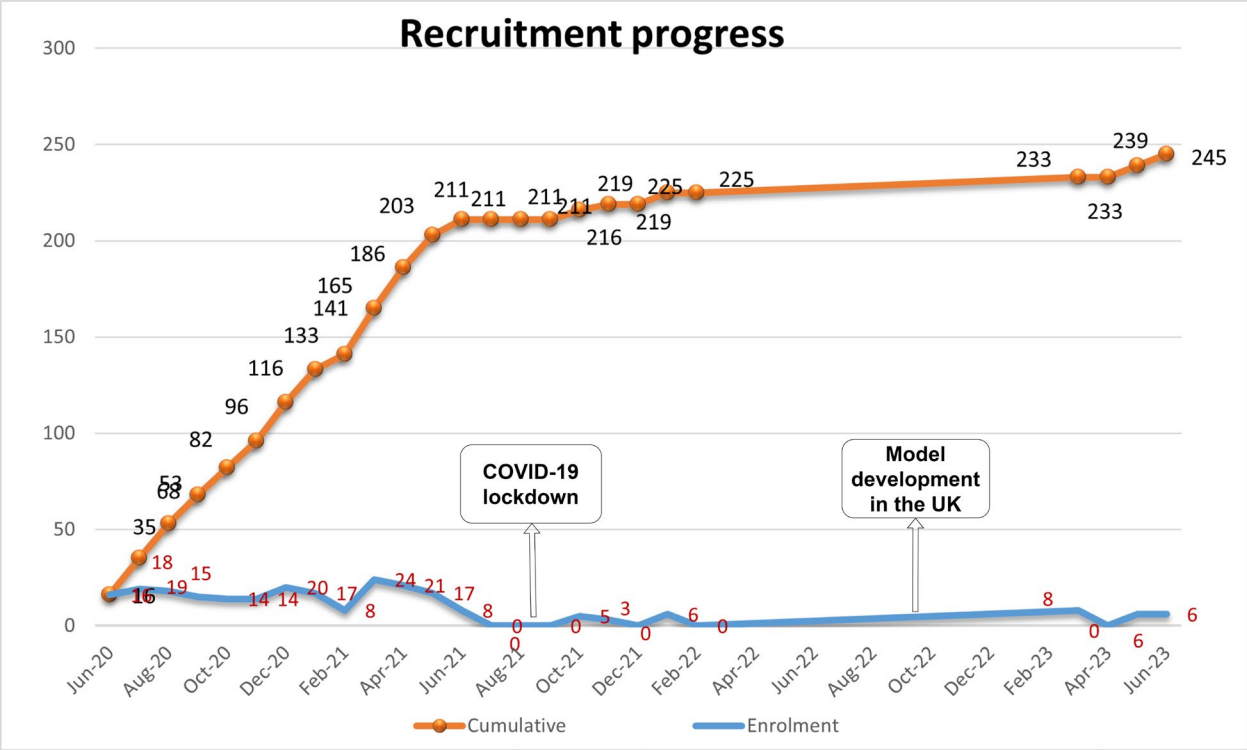


Figure 3.4. Recruitment progress of prospective study at HTD

Table 3.1. Demographics of patients in the study

Patient characteristics	N = 245
Age (year)	47 ± 17
Gender (Female)	75 (30.9%)
Weight (kg)	59 ± 14
Height (cm)	161 ± 13
Diagnosis	
Dengue	52 (21.2%)
Sepsis	33 (13.5%)
Tetanus	126 (51.4%)
CNS infection	34 (13.9%)
Length of hospital stay (days)	17 ± 15
Length of ICU stay (days) (n = 214)	20 ± 17

Outcome	
Home	213 (86.9%)
Hospital Transfer	16 (6.5%)
Death (at home or in hospital)	16 (6.5%)

The kind of collected data including the study and the publicly available datasets utilised are shown in **Table 3.2**. Two publicly available datasets in echocardiography - Cardiac Acquisitions for Multi-structure Ultrasound Segmentation (CAMUS) [127] and Echonet-dynamic [128] - are used for the study.

3.4 Data annotations

The US videos collected from the clinical study were anonymized to remove any patients' identifiable information by cropping out the region outside of the US window. Depending on the clinical task, each US modality was annotated and labelled differently. For LUS, annotation focused on classification of 5 common features. For muscle US, annotation focused on segmentation of the rectus femoris muscle, and involved manual contouring of the muscle. Finally, for echo, annotation focused on left ventricular segmentation, and the annotations were manual contours of the LV endocardium at the end-diastolic (fully dilated) and end-systolic (fully collapsed) frames. Below, the annotation process for each US modality is described in more detail.

Table 3.2. Detail of collected datasets and publicly available datasets

Modality	Source	No. patients	No. exam per patient	No. acquisition in exam	No. total examinations	Full exam recording	Standard view	Note
Cardiac US	Our clinical study	111 patients	3 (within 7 days)	PLAX, PSAX, A4C, A5C, A2C, sub costal IVC, measurements	278 examinations (260 A4C video)	US screen recorded in .mp4 (25mins/exam)	DICOM	6 cardiac circles with ECG
	Echonet-dynamic [128]	10030 patients	N/A	A4C	10360 A4C videos	N/A	.avi	3-5 cardiac circles, Heavily down sampled, without ECG
	CAMUS [127]	500 patients	N/A	A4C	500 A4C videos	N/A	.png	1 cardiac circles, without ECG
Lung US	Our clinical study	111 patients	3 (within 7 days)	12 lung zones + 2 base	283 examinations (3396 LUS videos)	US screen recorded in .mp4 (21 mins/exam)	DICOM	4 seconds video
	OUCRU study	60 patients	5 (within 7 days)	12 lung zones + 2 base	250 examinations (3000 LUS videos)	N/A	.mp4	4 seconds video
Muscle US	Our clinical study	134 patients	3 (ICU admission, DAY 7, ICU discharge)	3 repeats standard RFCSA view for each leg	305 examinations (1800 RFCSA videos)	US screen recorded in .mp4 (17 mins/exam)	DICOM	4 seconds video

3.4.1 Lung US annotation

LUS videos were stored as 4 second clips each in .mp4 format and were annotated offline. Sample lung US images are shown in Figure 3.5. The nature of LUS is that the features can appear and disappear during respiratory circulation. Therefore, the saved videos can contain the features of interest (e.g. pleural effusion, B-lines, etc.) only in a segment of the video.

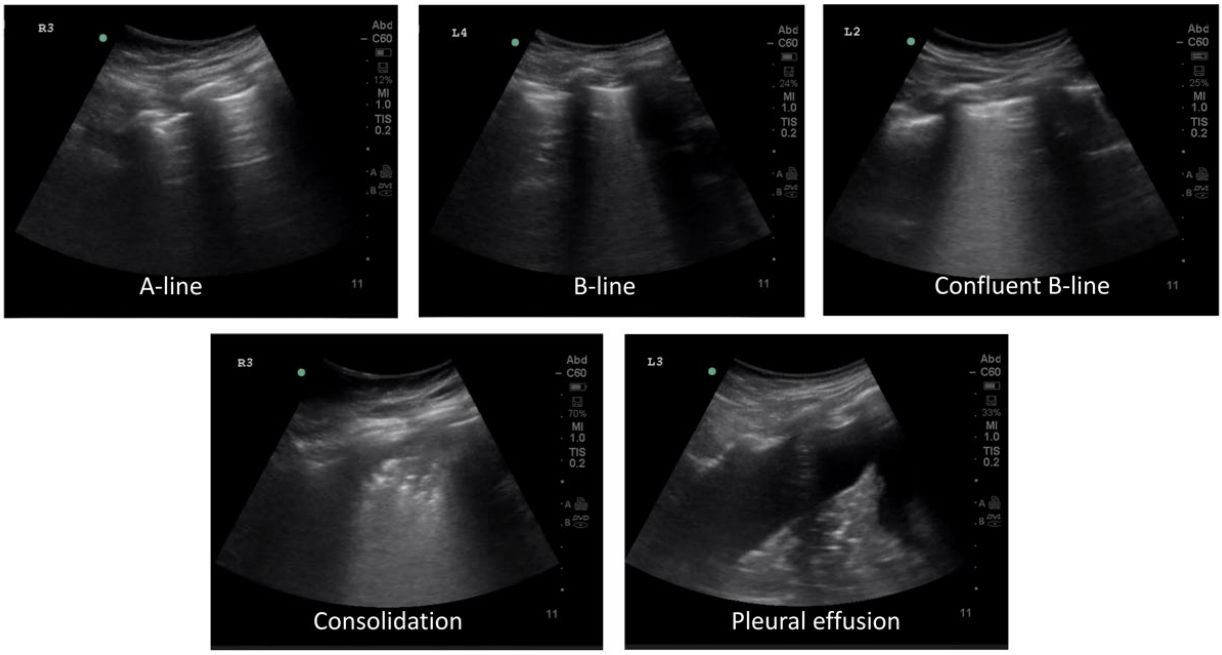


Figure 3.5. Five common features of LUS. A-line (representing a healthy lung, horizontal reverberation artifacts of the pleural line caused by multiple reflection), isolated B-lines (vertical hyperechoic artifacts deriving from the pleural line spreading to the end of screen, moving synchronously with the lung, more than 2 B-lines is considered abnormal), Confluent B-lines (many B-lines merge together and occupy a large area in the intercostal space, which is a sign of pulmonary interstitial syndrome), Consolidation (an echo-poor image juxtaposed to the pleural line and delimited by irregular boundaries, usually seen in pneumonia), Pleural effusion (hypoechoic space between the parietal and visceral pleura, is the build-up of excess fluid between the layers of the pleura outside the lungs due to heart failure, plasma leakage, pneumonia or pulmonary embolism).

Therefore, annotations had to take account of this temporal variation. The anonymized LUS video clips were annotated by qualified clinicians using the VGG annotation tool which is suitable for

video annotation [129]. During the annotation procedure, as illustrated in Figure 3.6, the clinicians loaded the 4 second video to the tool and played the video. Then, the clinicians identified and marked when the feature appeared as the beginning and when it disappeared as the end. After that, the label (i.e. either A-line, B-line or Confluent-B-line, Consolidation or Pleural effusion) was assigned to this segment of the video clip. In the case of B-lines videos, the B-line frames in the videos were annotated frame-by-frame using a straight line to draw the B-lines in that frame for the temporal localization task.

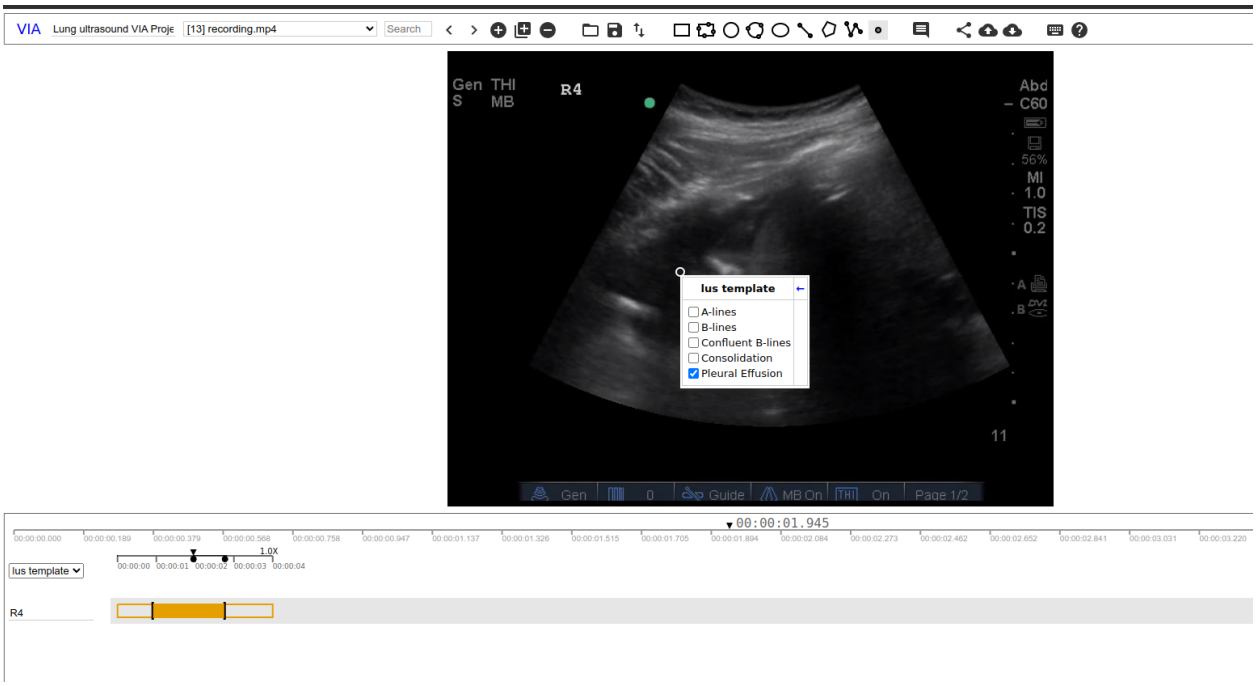


Figure 3.6. VGG VIA annotation tool used for annotating lung US videos. Sample illustrates a 4 second clip of pleural effusion being annotated. In that 4 second video, the clinician identified that the pleural effusion appeared from second 1 to second 3 (yellow bar).

The annotation outputs were saved in Java Script Object Notation (JSON): https://www.json.org/ (accessed on 1 December 2021) format to be used as ground truth. 3000 clips were annotated at video-level by 2 clinicians and at frame-level by 1 operator and a subset of data were validated by an external LUS expert.

3.4.2 Muscle US annotation

Establishing a well-defined ground-truth segmentation was of utmost importance for muscle ultrasound segmentation. The Rectus Femoris (RF) muscle was annotated and measured at the examination time. Radiologists carried out the examination and did the manual segmentation of the standard muscle scans. During the exam, the radiologists held the probe to acquire the standard view, then they froze the image. Next, the radiologist used the tracing function to manually delineate the RF muscle and save the raw images and their corresponding masks.

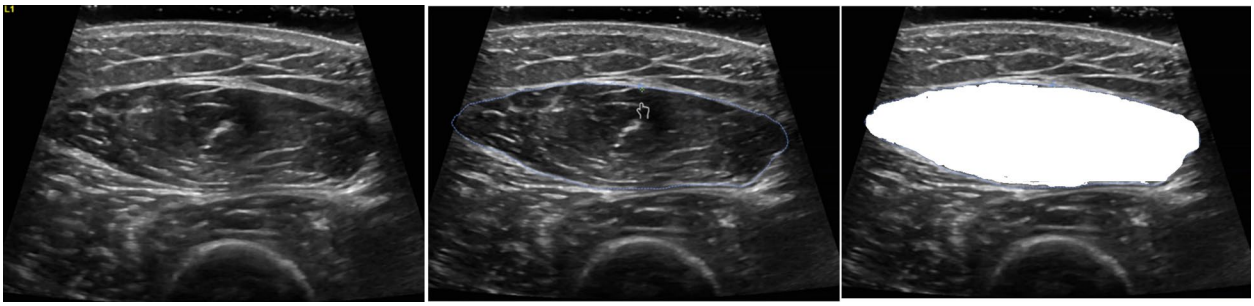


Figure 3.7. Sample RFCSA scan. (A) standard view image, (B) manual contour by a radiologist and (C) extracted mask to use as ground truth for model development

A Python script was written and used to anonymize and extract the standard view and corresponding segmentation mask of the stored data - a sample can be seen in Figure 3.7. The images and masks were saved in JPEG format to be used as ground truth for model development. In total 1200 muscle images were annotated by 3 radiologists and 1 operator.

3.4.3 Cardiac US annotation

Two radiologists and 1 intensivist carried out the echo exam following the standard operating procedure illustrated in Figure 3.8 and manually traced the endocardial contours at end-systole and -diastole on the apical 4 chamber view in all study patients with sufficient image quality using Simpson's method of discs [130]. End-diastole was defined at the peak of the electrocardiographic R-wave and/or 1 frame before mitral valve closure. End-systole was defined as 1 frame before mitral valve opening or when end-systolic volume was deemed to be at its smallest by the clinicians.

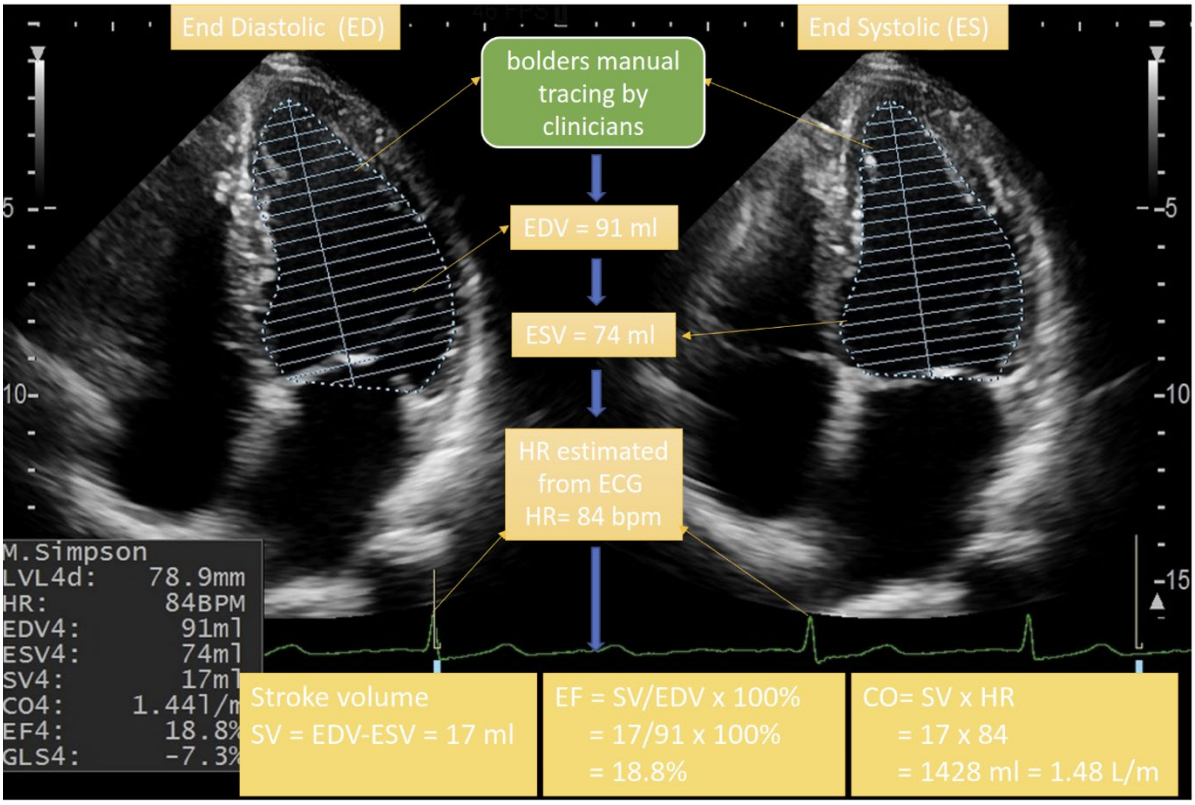


Figure 3.8. Example of computation of the SV, LVEF, CO from 4 chamber view US video using Simpson's method implemented in the ultrasound machine's software. Example of computation of the SV, LVEF, CO from 4 chamber view US video using Simpson's method implemented in the ultrasound machine's software. The process includes 5 steps: Step 1 – Identifying the End Diastolic (ED) frame (based on ECG if available or based on experience of user), Step 2: Tracing the endocardium (inner LV border) of the LV at ED for measuring the ED volume (EDV), Step 3: Identifying the End Systolic (ES) frame (based on ECG if available or based on experience of user) of the same cardiac cycle, Step 4: Tracing the endocardium of LV at ES for measuring ES volume (ESV). Step 5: Calculate the SV, EF and if ECG is available the CO will be estimated by multiplying the SV and heart rate (HR).

In addition to clinical study dataset, I also used publicly available datasets including Echonet-Dynamic and CAMUS datasets for model development. The full description and annotation protocol of Echonet-dynamic [128] can be found and downloaded in https://echonet.github.io/dynamic/ and CAMUS [127] in https://www.creatis.insa-

lyon.fr/Challenge/camus/databases.html. The former dataset has 2D A4C and two-chamber view sequences of 500 patients. For each sequence, the manual annotation for the ED and ES frames of the left ventricle structures - endocardium, epicardium, and left atrium are provided as the ground-truth for 450 patients. The Echonet-dynamic dataset consists of 10030 different A4C echocardiography videos with corresponding number of ED and ES frames. For each video, two tracings from experts are provided of both the ED and ES. The US images for ED and ES stages are extracted from the video and frame information, and the ground truth is created from the expert tracings of the left ventricle.

3.5 Discussion and conclusion

AI application in low-resource settings should build into existing systems and institutions rather than starting from scratch or hoping to replace existing systems. The success of AI applications requires knowledge of local patients' population, infrastructure, staff, clear usability requirements and access to adequate training data via field testing. In order to be robustly evaluated, and utilised in LMICs, the AI model needs a sufficient amount of data to build or retrain the model. The potential approach is to establish valid data collection protocols in LMICs that allow collecting high quality, representative of the patient population where AI tools will be deployed in the future. This study was able to collect dataset for cardiac, lung, and muscle ultrasound in patients with infectious diseases in the ICUs in Vietnam using the available devices. The dataset was used for model development, validation, and testing in Chapter 4, 5, and 6 and will be made publicly available following Wellcome Trust and OUCRU policies.

4 AI-assisted lung ultrasound

Chapter overview

This chapter presents a prospective study to evaluate the clinical benefit of AI-assisted lung ultrasound in low resource settings. The main contribution in this chapter is the clinical evaluation of an AI solution that assists LUS practitioners and the assessment of its usefulness specifically in a low resource ICU.

The chapter begins by introducing the clinical needs and challenges of implementing LUS in the ICU, with a particular focus on the LMIC setting, followed by motivating the need for an AI-assisted LUS interpretation solution in Section 4.1. I propose a three-phase approach to assess the clinical benefit of the AI tool in resource limited ICU in Section 4.2. In the first phase (Section 4.2.1) I confirm the use-case for our tool and set a minimum performance target for the AI system. The second phase (Section 4.2.2) develops and assesses whether the tool could help clinicians interpret LUS video in a clinical controlled setting. Finally, in the third phase (Section 4.2.3) I evaluate the benefit when clinicians use the AI tool to carry out LUS exams in critically ill patients at the hospital. I present the results of each phase in Section 4.3 and discuss them in Section 4.4.

A journal paper of the work presented in this chapter has been published in the Critical Care journal [131].

4.1 Introduction

In recent years, POCUS has proved to be a useful bedside imaging technique for the assessment of critically ill patients for both diagnosis and therapeutic management [1, 25, 29]. LUS does not expose patients to radiation and has been shown to be more sensitive and specific in the diagnosis of many pulmonary pathologies when compared to chest x-ray [2]. LUS is fast, low-cost and able to detect and diagnose lung pathologies such as cardiogenic (heart-related) pulmonary edema, inflammatory interstitial lung diseases, pneumonia and pleural effusion. Many of these

conditions are prevalent in LMICs so the potential for application of LUS in LMIC settings is high [7, 132].

Respiratory failure due to infectious disease is one the most common reasons for ICU admission in LMICs, for example, due to dengue, sepsis, or malaria and more recently Covid-19. In severe cases progression to ARDS can occur, which has high mortality and leaves survivors with significant pulmonary morbidity [133–135]. As a result, high quality management of these patients is crucial.

Many well-established LUS protocols such as the BLUE protocol and FALLS protocol [32] are designed to assist doctors with the diagnosis and management of pulmonary and cardiac conditions. In particular, the Kigali protocol [15] has helped to diagnose ARDS in resource-limited settings by using ultrasound to identify pulmonary edema instead of using CXRs or CT. In LMICs, access to high resolution (and high cost) volumetric imaging modalities such as CT or MRI may not always be possible, and so patients would benefit from the use of low-cost ultrasound imaging. In critically ill patients in the emergency department or ICU, LUS can enable real-time diagnostic capability at the bedside more quickly and efficiently.

A recently performed meta-analysis [56] confirmed high sensitivity (96%) and specificity (93%) of LUS for detecting pneumonia. However, a common challenge in LMICs is the inability to regularly perform LUS on patients with suspected pneumonia, due to lack of equipment and expertise. In addition, like other ultrasound techniques, LUS is operator-dependent and requires training for image acquisition and interpretation. Specifically, LUS focuses on imaging artefacts except for pleural effusion and consolidation, which are produced at the pleural surface. These artifacts change over time with the respiratory cycle, so image interpretation is challenging especially in critically ill patients with respiratory distress and tachypnea (rapid and shallow breathing). Therefore, the lack of qualified ultrasound professionals, most likely due to lack of training programs in developing countries, is an obstacle to the implementation of LUS in LMIC settings.

4.1.1 Related works

Artificial intelligence, particularly deep learning, has made substantial advances in ultrasound image analysis during the last decade. For LUS, most existing work has been limited to

classification of a single artefact (B-lines), or, more recently, multi-class classification for a specific lung disease, focusing on Covid-19 [57, 136–139]. B-lines identification can be a challenging skill for novice LUS operators. Kerdegari et al [139, 140] showed that the combination of a convolutional neural network (CNN) with a long short-term memory (LSTM) network and a temporal attention mechanism can classify videos based on the presence or absence of B-lines (to an accuracy of 81%), and temporally localise the frames which have B-lines in them (to an accuracy of 87%). However, this method was limited to B-lines and did not consider other artefacts, such as consolidation or pleural effusion, which are crucial for efficient patient management in the ICU. Most research related to classification of multiple artefacts in LUS was published recently and focused specifically on COVID-19 patients. Roy at al [57] introduced several models to classify and localise COVID-19 markers (A-lines, vertical artifacts, consolidation and white lung) in POCUS LUS, and included frame-based score, video-based score prediction and semantic segmentation. For video-based classification, the best model achieved F1, Precision and recall of 61%, 70% and 60%, respectively. Another study conducted by Liu et al [136] proposed a new multi-symptom multi-label model with Active Learning methods [13] that was able to learn from a smaller amount of annotated data to classify multiple COVID-19 lung features, achieving an accuracy of 100%, 95.72%, and 80.98% for A-line, B-line and pleural lesions, respectively. Both studies have made the code and datasets publicly available. Arntfield et al [58] developed a CNN to classify LUS images with B-lines of different etiologies (COVID-19, non-COVID respiratory distress syndrome and hydrostatic pulmonary oedema) and showed a better performance than physicians.

Most of the previous studies mentioned above, work on static LUS images that have been identified, selected and saved by the clinician. Two studies, conducted by Liu et al [14] and Born et al [141], proposed video-based classification for COVID-19 related image features and differentiated COVID-19 from bacterial pneumonia and healthy subjects. Recently, Camacho et al [142] showed preliminary results of an AI-assisted tool for offline LUS score calculation named ULTRACOV in 28 COVID-19 patients and found the agreement between the findings of their algorithm and the expert were 88.0% for B-Lines, 93.4% for consolidations and 99.7% for pleural effusion detection. In addition, the average interpretation time with the ULTRACOV prototype

was 5.3 min, while with a conventional scanner it was 12.6 min. Tan et al [143] developed a LUS B-line detection AI system using YOLACT [144] and Mask-RCNN [145] for B-line counting, to assist in estimating fluid overload in dialysis patients. This tool was used retrospectively to evaluate LUS videos collected by physicians and achieved good agreement with physicians' annotated labels (r=0.825).

There are also several commercial B-line quantification software applications, for example Auto B-line of GE Healthcare and B-line Quantification of Phillips Lumify. It is worth noting that both software packages work by counting the maximum number of B-lines in a single frame. Specifically, Auto B-line (GE Healthcare) assists clinicians in counting the number of B-lines in a pre-recorded LUS scan, and the software has been evaluated by comparing the accuracy of AI versus physician assessment for B-lines in some studies [146, 147]. In a study conducted by Gottlieb et al [146], a research assistant used Auto B-Line to retrospectively analyse the same clip as a sonographer to distinguish between multiple B lines and one dense B line and showed that the Auto B-line software performance was comparable to experts. Phillips Lumify developed a feature for a standard 12 lung zones scan that includes a B-line counting algorithm and many studies have shown the feasibility of this tool [148]. However, this tool only works with a linear probe which is suitable for pleural lines and B-line quantification but not for other important lung features such as consolidation or pleural effusion.

Moreover, a group led by Robert Arntfield[2] is working actively on developing real-time validation of a LUS deep learning model in an ICU in Canada. The authors built and evaluated the real-time AI tool to distinguish between A-line pattern and B-line pattern in 100 critically ill patients The real-time inference achieved an accuracy of 95%, a sensitivity of 93%, and a specificity of 96% for identifying B-line pattern. Importantly, their DL model was able to run inference at a rate of 30 predictions per second [97].

Many guidelines suggest that instead of counting the number of B-lines in one lung zone, quantifying the percentage of lung zone occupied by B-lines pattern is a better approach to determine the overall severity of the lung pathology. With this in mind, Brusasco et al [149]

[2] https://www.deepbreathe.ai/

developed several LUS scores for B-line quantification in the measurement of extravascular lung water (EVLW) including (1) maximum number of B-lines detected (nLUSS), (2) visual percentage of lung area occupied by B-lines (%LUSS), (3) B-line coalescence (cLUSS), (4) modified B-line coalescence score (qLUSS), and (5) computer-aided score (QLUSS). The results showed that QLUSS (computer-aided score) of the pleural line percentage affected by B-lines has the potential to assess EVLW. QLUSS showed a stronger association with EVLW ($R2 = 0.57$) than cLUSS ($R2 = 0.45$) and nLUSS ($R2 = 0.000$), while there was a lower association than qLUSS ($R2 = 0.85$) and LUSS ($R2 = 0.72$).

Kuroda et al. [96], investigated AI-assisted LUS to detect pneumonia in patients with COVID-The presence of pneumonia in the patients was validated using CT. The AI-assisted 12 zone point-of-care US was shown to be highly accurate, sensitive and specific (94.5%, 92.3% and 100%, respectively). However, it is worth noting that the accuracy, sensitivity and specificity decreased with an 8 zone US and a single zone US. Furthermore, a limitation of this work was that it was a single centre study with a small population group. Finally, the algorithm was only able to detect B-lines and no other lung features, thus it was not sufficient information to accurately diagnose lung pathology.

Overall, the implementation of developed algorithms for US analysis in clinical settings remains very limited and has focused mainly on fetal and cardiac ultrasound [42, 44, 107]. In the case of LUS there are, to date, no published investigations on real-time deployment of AI-enabled tools for five common LUS feature classification in lung ultrasound. In addition to improving accuracy, automatic recognition can help improve confidence in, and reduce the time spent by the clinician on image interpretation; this may be especially beneficial where resources are limited, and highly trained healthcare staff are not always available.

In this study, we aim to narrow the implementation gap by assessing the utility of AI-enabled LUS. To this end, we investigate the performance of clinicians with a range of clinical expertise with respect to LUS interpretation and investigate to what extent a real-time AI-assisted LUS model can help improve this performance in a resource limited ICU setting.

4.2 Proposed approach

Our study was carried out in Vietnam in three phases as illustrated in Figure 4.1. In the first phase we confirmed the use-case for our tool and set a minimum performance target for the AI system. This was an online interactive survey of participants from multiple hospitals in Vietnam. In the second and third phases we assessed the utility of the AI tool at the HTD in Ho Chi Minh City. We chose to focus on non-expert clinicians in phases 2 and 3, as they are the eventual target users for our tool. The study was approved by the Scientific and Ethical Committee of the HTD and the OxTREC. All participants gave written informed consent.

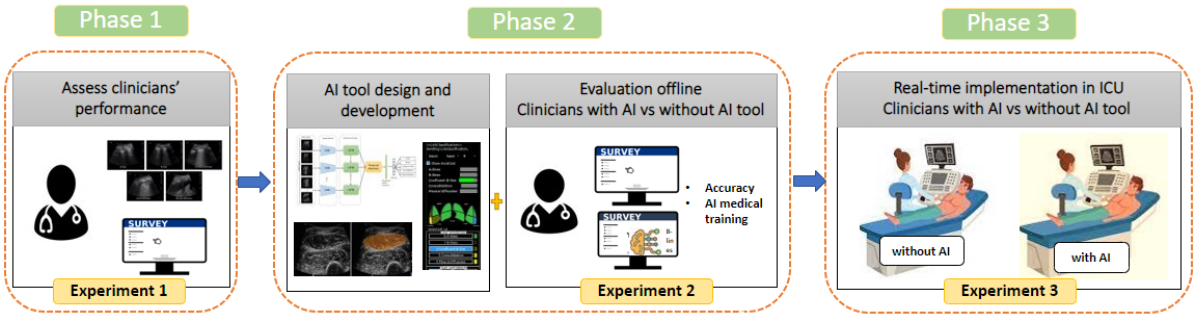


Figure 4.1. Overview of the three phases of this study: first, we assessed clinician performance on LUS interpretation. Second, we designed and investigated the impact of using the proposed model in offline image interpretation. Finally, we investigated the clinical benefit of real-time AI assisted LUS.

4.2.1 Phase 1: Baseline characterization of user performance in LUS interpretation without AI support

An online survey was completed by 276 participants during an online LUS training course attended by doctors from multiple centres across Vietnam, on September 4th, 2021. The level of expertise was self-assessed by clinicians using four pre-defined categories: 1) beginners, defined as "just know about LUS but have not practiced in patients", 2) intermediate, a clinician who has been carrying out LUS (<2 times/week) but have not used the findings for clinical assessment, 3) advanced, a clinician who uses LUS in daily practice and their findings are used for clinical assessment, and 4) expert, a clinician who has specialised in LUS and has more than 5 years'

experience. The participants were asked to identify the following findings in a series of 10 LUS clips from adult patients hospitalized with dengue shock and septic shock: A-lines, B-lines, confluent B-lines, consolidation and pleural effusion (2 of each) given in a set order (samples are shown in Figure 4.2). Responses were compared with the expert-defined labels consistent with our data curation process. All clips were sourced and labelled with agreement by three ultrasound-trained clinicians and one expert.

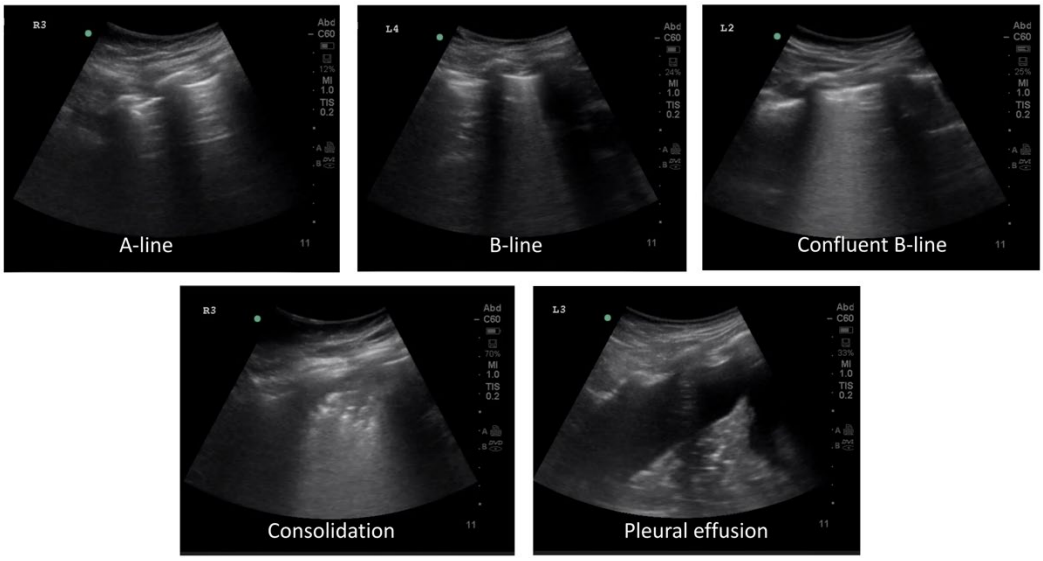


Figure 4.2. Sample LUS images. The LUS artifacts reflect the ratio between air, lung tissue, fluid, or other biological components. A-line (representing a healthy lung, horizontal reverberation artifacts of the pleural line caused by multiple reflection), isolated B-lines (vertical hyperechoic artifacts deriving from the pleural line spreading to the end of screen, moving synchronously with lung, more than 2 B-lines is considered abnormal), confluent B-lines (many B-lines merge together and occupy a large area in the intercostal space, which is a sign of pulmonary interstitial syndrome), consolidation (an echo-poor image juxtaposed to the pleural line and delimited by irregular boundaries, usually seen in pneumonia), pleural effusion (hypoechoic space between the parietal and visceral pleura, is the build-up of excess fluid between the layers of the pleura outside the lungs due to heart failure, plasma leakage, pneumonia or pulmonary embolism).

4.2.2 Phase 2a: Design, Development and Deployment of AI-assisted LUS video classification

4.2.2.1 Dataset

In this study, the dataset was collected from 65 patients with dengue shock or septic shock admitted to the HTD in Ho Chi Minh city, Vietnam between June 2019 and June 2020. The research was approved by the OxTREC and the HTD Institutional Review Boards. The dataset was split into training, validation and test set as follows: 90% (3079 videos, four seconds each) was used for training and validation and the remaining 10% of the patients (322 videos, four seconds each) were left out for the test set. LUS examinations were carried out using a Sonosite M-Turbo machine (Fujifilm Sonosite, Inc., Bothell, WA) with a low medium frequency (3.5-5 MHz) convex probe by qualified sonographers. LUS was performed using a standardised operating procedure based on the Kigali ARDS protocol [33]: assessment for B-lines, consolidation and pleural effusion, performed at 6 points on each side of the chest (2 anterior, 2 lateral and 2 posterolateral).

These videos were annotated by expert sonographers using the VGG annotator tool [86]. Five lung patterns were selected for multi-class classification, as introduced above and illustrated in Figure 4.2: A-line (normal lung), isolated B-lines, confluent B-lines, consolidation, pleural effusion. The distribution of the overall data is shown in **Table 4.3**. Class imbalance was addressed during training by weighting each sample's contribution to the loss with the inverse of the number of samples in its class. Class imbalance may affect model performance and in our study, there was relatively few data on consolidation as compared to other classes (21 videos versus 2000 videos of pleural effusion). However, it also reflects real-world clinical settings in this ICU, in which there are relatively few patients with pneumonia (consolidation presented in the LUS image) and dengue patients with plasma leakage (may cause pleural effusion in LUS image) are predominant

AVI-format videos were cropped and masked to remove text and information outside of the scanning sector. The 640x480 pixel videos were downsampled using OpenCV into 64x64 pixels. For training, each four-second clip was converted into shorter clips of one second with an overlap of 20% between consecutive frames in the video.

4.2.2.2 Model architecture

Our proposed model is an extension of the model from [139] that was used for B-line classification and localization in LUS videos. The model was developed by the technical team and was described in detail in Appendix C It consists of three parts: a convolutional neural network (CNN) to extract frame-wise spatial features, a bidirectional long short-term memory (LSTM) network to extract temporal features from the video and a temporal attention mechanism to increase the weighting of frames that carry more useful information for the classification task. In this paper, we replaced the classification subnet, after the temporal attention mechanism, with a fully connected layer (with ReLU non-linear activation and dropout) and a 5-element final layer that produces a 1-hot 5D vector for 5 class classification.

4.2.2.3 Integration into a real-time framework: Real-time AI-assisted LUS framework (RAILUS)

The model above was integrated into PRETUS — a Plugin-based Real Time Ultrasound software platform for live ultrasound image analysis and operator support [150] that allows simple and quick deployment of real-time ultrasound analysis algorithms. We named the integration of our method RAILUS (Real-time AI-assisted LUS). The user interface is shown in Figure 4.3 (top left), with a detail of the interactive widget in Figure 4.3 (top right) and images showing the clinical setup in Figure 4.3 (bottom). RAILUS and PRETUS are open source and publicly available in https://github.com/vital-ultrasound/public-lung.

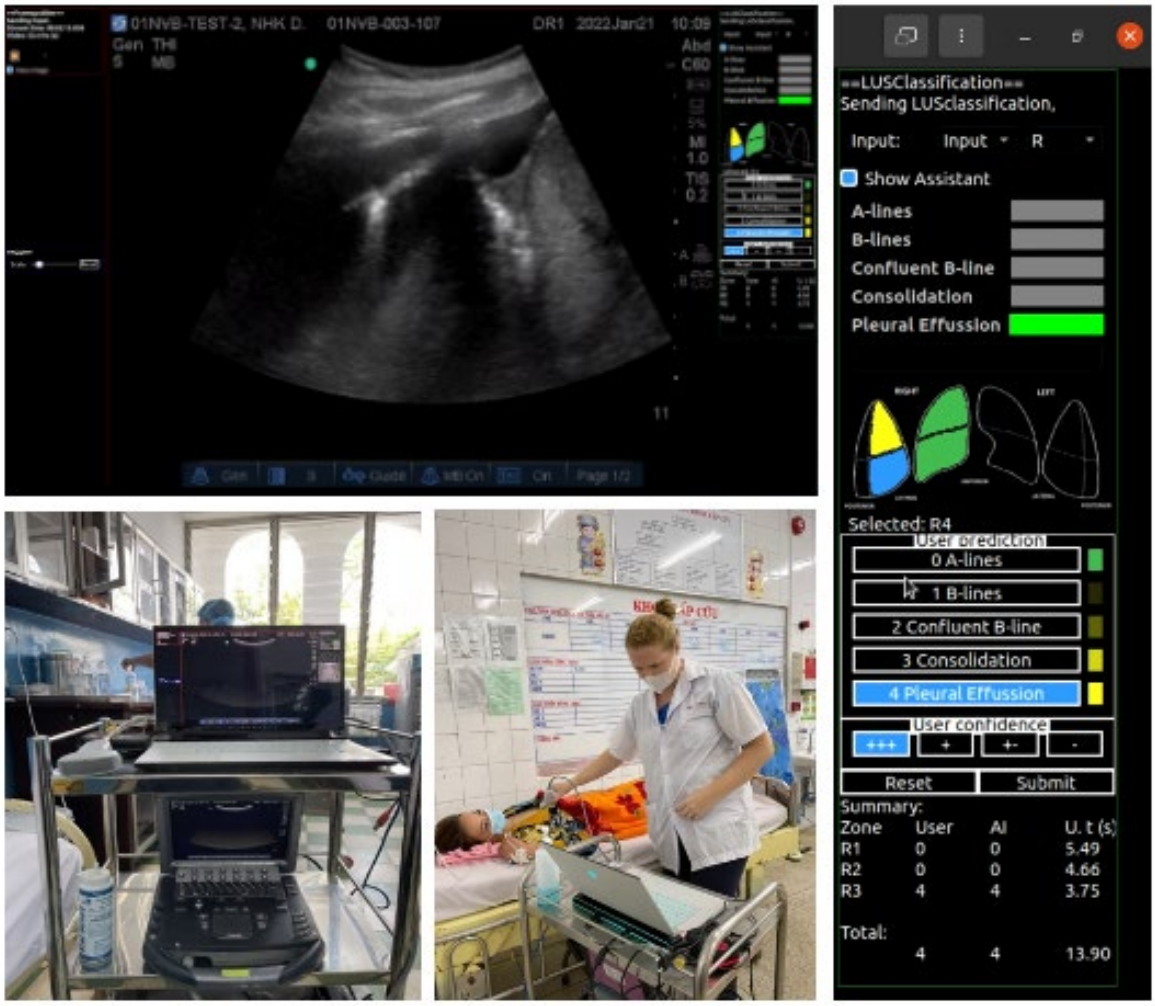


Figure 4.3. Real-time AI-assisted LUS framework (RAILUS). Please refer to the text for details.

The full standard operation procedure of how to use the RAILUS software is described in Appendix C. In short, on the right of the screen is the widget of the RAILUS software, where the model provides continuous real-time prediction (as the green bar) of the class of corresponding clips. The clinician interacts with the laptop by mouse/mouse pad to select the lung zones (12 lung zones) and labels. Therefore, the tool acts as an assistive AI method rather than an automated AI method because the final decision was made by clinician while using the tool. After choosing the label and submitting, the lung zone in the lung diagram changes color (e.g. from green-normal lung to yellow-pleural effusion) according to the class of the label. For the purpose of testing and comparison between clinicians with the AI tool and without the AI tool, the bottom

right shows the user prediction, model's prediction, user's confidence and total time taken for interpretation.

The RAILUS software can be used in both pre-recorded LUS clips and the real-time clinical environment with clinicians carrying out the LUS examination by reading the machine's video output.

4.2.3 Phase 2b: Clinical evaluation of RAILUS tool in clinical controlled environment

In this phase, we evaluated the impact of our bespoke LUS AI system RAILUS on clinicians' performance in a controlled environment, using a set of images already obtained by LUS experts.

We evaluated the performance of RAILUS in a controlled environment in workshops for 57 non-expert clinicians in three different clinical settings (tertiary referral centre, COVID-19 field hospital and academic ventilation training course). Participants were given 1 hour of training to become familiar with the RAILUS software, then were asked to identify the presence of 5 different lung US features (A-lines, B-lines, confluent B-lines, consolidation, pleural effusion) in the expert-acquired LUS videos (as in phase 1). They were first tasked to do this without the RAILUS tool and then with the RAILUS tool. When performing the task without RAILUS, participants were blinded to the AI-assisted interpretation by turning off the AI prediction feature on the RAILUS software. The responses were used to calculate the average accuracy of the interpretation (5 class classification task) with and without the RAILUS tool and compared to expert-defined labels.

4.2.4 Phase 3: Real-time evaluation of RAILUS software in critically ill patients

Real-time evaluation was then carried out prospectively by a subset of the non-expert clinicians from Phase 2, who were now asked to perform a LUS with and without the AI system in the Emergency Department or on ICU patients. Eligible patients included adults aged ≥18 years with dengue shock admitted to HTD between December 2021 to Feb 2022. Patients who were allergic to ultrasound gel or had open wounds in their chest were excluded. Participating non-expert clinicians were randomly assigned to perform the LUS scans on patients following a standard 12 zone LUS protocol with or without the RAILUS software (Figure 4.4). When assigned to the non-AI group, the AI-assisted interpretation was turned off. A LUS expert then performed the same

12 zone scan on the same patient within 2 hours of the non-expert clinicians performing their LUS scans. To assess whether each scan was of adequate diagnostic quality, we performed an independent expert validation (blinded to whether the study was performed using the AI tool or not).

Results were evaluated to quantify the accuracy of the interpretation against the expert, the time required to interpret single lung zones, and the clinicians' perceived confidence in their interpretation (from 1 to 4. with 1 not confident and 4 very confident).

The end points that were evaluated for this study were: (1) accuracy of the interpretation against the expert, (2) time required to interpret single lung zones in seconds, and (3) confidence level with their interpretation.

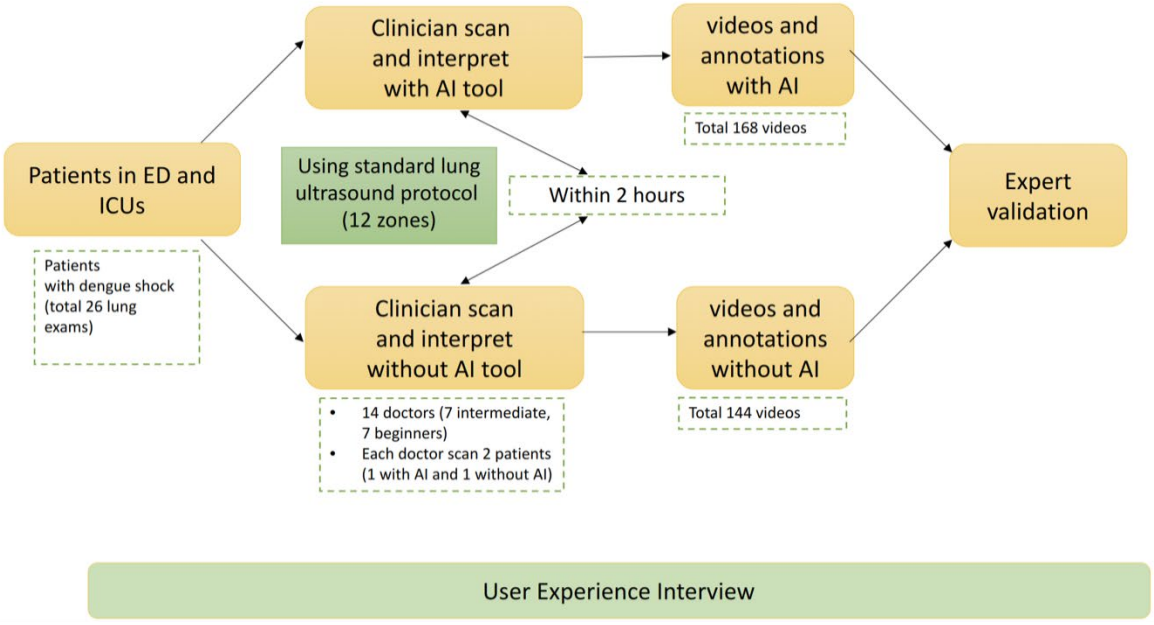


Figure 4.4. Prospective study schema. Patients admitted to ED or ICU were recruited to the study after stabilization. The patients were scanned twice: first, by a clinician with AI-assisted tool (RAILUS), and second, by another clinician without AI-assisted tool 2 hours later. The video and annotations of the collected LUS video were validated by an independent LUS expert.

4.2.4.1 *Usability questionnaire*

A questionnaire was administered to the clinicians who took part in the real-time implementation study (Phase 3). After completing the training and using RAILUS to perform LUS in dengue shock patients, the clinicians were then asked to fill in the questionnaire. The answers of the questionnaire were valued by the five-point Likert scale. The questionnaire was developed and delivered by Google Forms. Because the interaction between AI-based tools and clinicians is poorly understood, the survey aimed at investigating how our LUS tool should be deployed in clinical settings, the level of clinician's trust in the tool and the level of concern about its utilisation. The full questionnaire can be found in Appendix F.

4.2.5 Statistical analysis

To assess the performance of the video clip classification task, we assessed 1) Overall accuracy, calculated as the number of correctly classified clips as a fraction of the total number of clips; 2) average accuracy, calculated as the average over all lung pathologies of per-pathology accuracy; 3) F-score; 4) precision and 5) sensitivity (recall). The proportion of clips that were accurately classified were reported with 95% confidence intervals (95% CI). Confusion matrices were calculated and reported. The discriminative variable of demographic information was reported as a percentage.

4.3 Results

4.3.1 Phase 1: User performance in LUS interpretation without AI support

Table 4.1 shows the demographic information of the clinicians and **Table 4.2** summarises the main challenges they encountered during ultrasound imaging. The main findings include: 70% of clinicians were beginners, there were very few experts (1%) and 72% identified "image interpretation" as the main challenge.

Table 4.1. Demographics of the participants. Distribution of participants according to training, expertise and previous US experience.

	N = 276 (%)
Level of training	

Radiology doctor	105 (38%)
Intensivist	57 (21%)
Doctor	77 (28%)
Resident	20 (7%)
Other	17 (6%)
Experience in general ultrasound	
Beginner	121 (44%)
Intermediate	121 (44%)
Advanced	30 (11%)
Expert	4 (1%)
Kind of ultrasound exam	
Echocardiography	115 (42%)
Abdomen	178 (64%)
Blood vessel	90 (33%)
Lung	117 (42%)
Intervention guidance	117 (42%)
Other	18 (7%)
Experience in lung ultrasound	
Beginner	194 (70%)
Intermediate	71 (26%)
Advanced	7 (3%)
Expert	4 (1%)

Table 4.2. Participant's opinion on challenges of ultrasound and available equipment in their unit/department

Principal challenges in lung ultrasound	
Image interpretation	200 (72%)

Identifying landmark	139 (50%)
Finding the right view	137 (50%)
Writing report	66 (24%)
Instruction for user	78 (28%)
Image quantification	58 (21%)
Is lung ultrasound used in your unit/department	
Yes	177 (64%)
No	99 (36%)
Type of ultrasound machine available	
Handheld devices (phone/tablet-based)	10 (4%)
Point-of-care	169 (61%)
"Full feature-trolley" big machine	156 (57%)

The results of human-only video classification are shown in the confusion matrices in Figure 4.5 (showing total count and percentage). Experts showed excellent ability to accurately classify LUS clips. For all other clinician categories, there was more difficulty in differentiating A–lines, B–lines and confluent B–lines, and relative ease at identifying consolidation and pleural effusion. In Figure 4.5, the horizontal and vertical axes show predicted label and expert–defined label, respectively. The numbers in each cell indicate total count (percentage of total). Cells are colored by percentage.

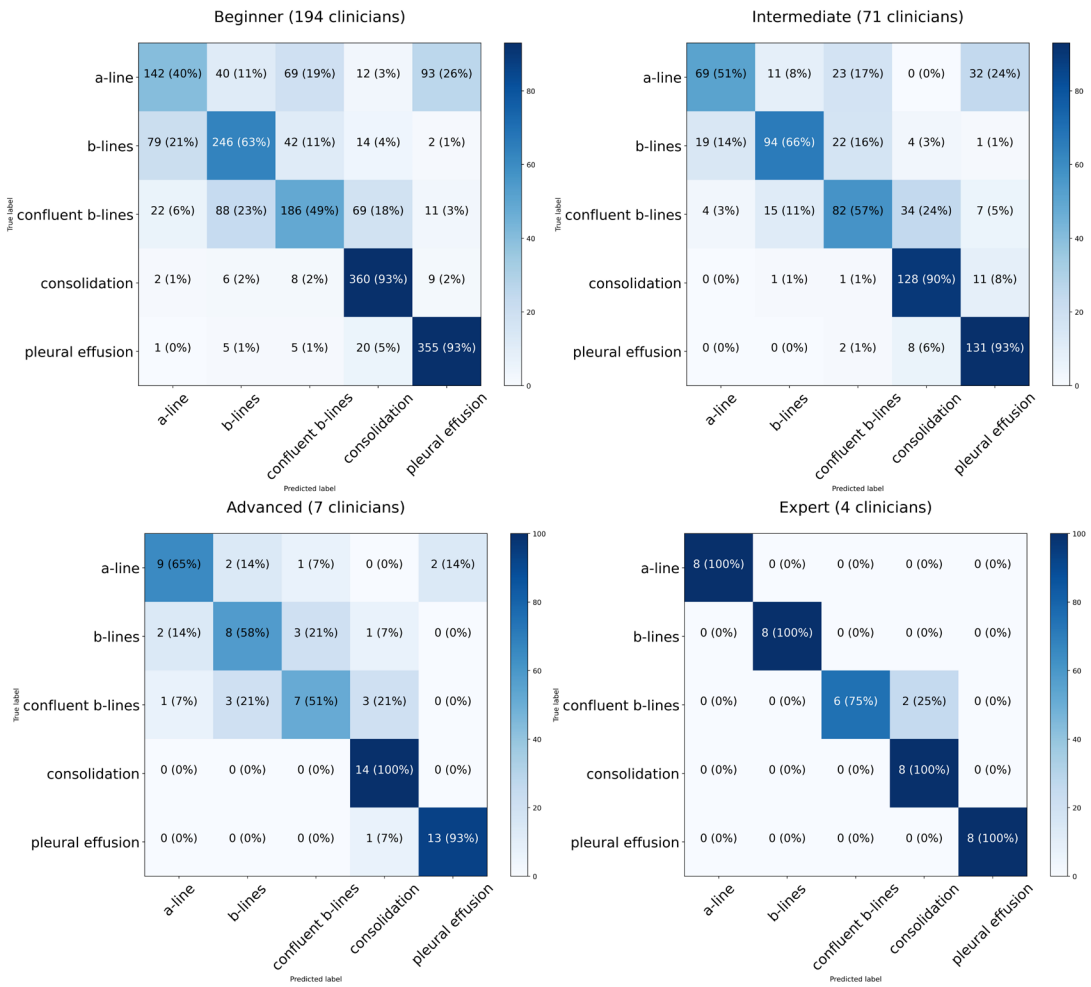


Figure 4.5. The results of the manual video classification by clinicians with 4 levels of expertise. Horizontal and vertical axes show predicted label and expert-defined label, respectively.

Although the average accuracy is correlated to the level of expertise (beginners: 68.7% (95%CI, 66.8%–70.7%), intermediate 72.2% (95%CI, 70.0%–75.6%), advanced 73.4% (95%CI, 62.2%–87.8%), and experts 95.0% (95%CI, 88.2%–100.0%)), ($p<0.001$), the difference in performance between beginners, intermediate and advanced users is relatively small (<5%). However, experts have significantly better performance.

4.3.2 Phase 2a: Performance AI-assisted LUS video classification algorithm

The proposed AI model was trained with more than 3000 lung ultrasound videos from dengue and septic shock patients in HTD, acquired using a Sonosite M-Turbo (Fujifilm Sonosite, Inc.,

Bothell, WA) machine. The details of each class and the corresponding class weights (to balance the training) are shown in **Table 4.3**.

Table 4.3. Number of samples per class and training weight values of each class.

Class	Number of videos	Weight
A-line	1825	1.0
B-lines	102	17.8
Confluent B-lines	138	11.4
Consolidation	21	78.3
Pleural effusion	993	1.6

Our model achieved an average accuracy of 81%±11%, with the test F1-score for each class (A-line, B-lines, Confluent B-lines, Consolidation, Pleural Effusion) being 92%, 67%, 75%, 67%, and 90%, respectively. The relatively low accuracy for consolidation is due to consolidation being a relatively rare condition for which we have significantly less data, as shown in **Table 4.3**. Detailed results on F1 score, precision and recall are provided in **Table 4.4**.

Table 4.4. Classification performance (F1-score, precision and recall) on the test set

	F1	Precision	Recall
A-line	92%	96%	89%
B-line	67%	57%	80%
Confluent B-lines	75%	68%	83%
Consolidation	67%	75%	60%
Pleural effusion	90%	87%	93%
Average (±std)	78%±11%	77%±14%	81%±11%

To put the model performance in context with the baseline operator performance reported in the previous section, we provide the confusion matrix (in relative numbers) in Figure 4.6. This can be compared directly with those in Figure 4.5.

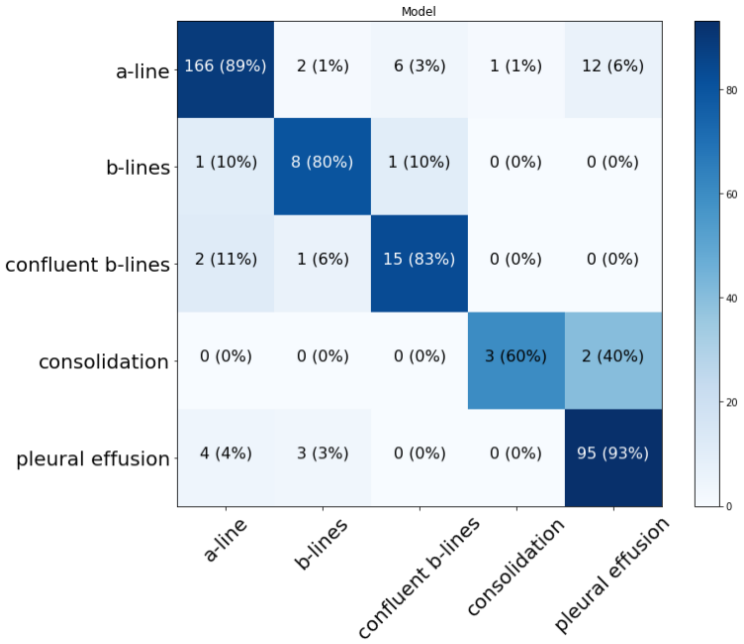


Figure 4.6. Confusion matrix of our proposed model

These results indicate clear trends supporting the suitability of our model to support clinicians. Our model outperforms beginners, intermediate and advanced users on average and in every class except for consolidation, for which we had significantly less data. Interestingly, all users have >90% accuracy in this class.

4.3.3 Phase 2b: Performance and clinical validation of the RAILUS software in a controlled environment

In this experiment, 57 clinicians who are beginner or intermediate users were asked to interpret LUS videos with and without the RAILUS software. The results on the LUS video classification are shown in the confusion matrices in Figure 4.7. The performance of clinicians that used the RAILUS software was better than when they did not use it, with a mean accuracy of 82.9% (95%CI, 86.7%–79.1%), compared to 68.9% (95%CI, 65.6%–73.9%), ($p<0.001$). The accuracy of all classes increased significantly except for the case of the B–lines class, which reduced slightly from 63% to 59% when using RAILUS.

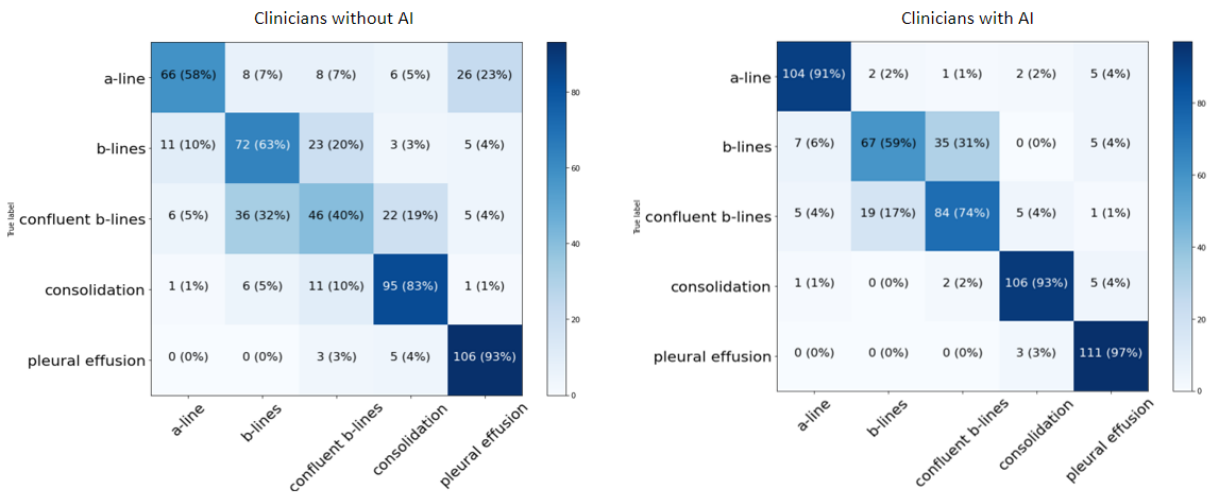


Figure 4.7. Confusion matrix of clinicians without (left) and with (right) RAILUS, in a controlled environment using retrospective data.

4.3.4 Phase 3: Real-time implementation of RAILUS software in critically ill patients

In total, seven dengue shock patients were recruited for real-time testing of the RAILUS software. 14 clinicians (7 beginner and 7 intermediate) were invited to carry out the lung examination in those patients with and without RAILUS. **Table 4.5** shows the characteristics of the patients. Overall, 26 LUS exams were carried out, resulting in 168 LUS videos (4 seconds each) with the AI tool and 144 LUS videos without the AI tool.

Table 4.5. Baseline of characteristics of the study patients

Baseline characteristics (n=7)	
Age (median, IQR)	19 (17-31)
Gender (male)	5/7
BMI	22.4 (20.6-24.6)
Heart rate (bpm)	100 (93-105)
Blood pressure	
SBP (mmHg)	105 (99-120)
DBP (mmHg)	75 (70-85)

Respiratory distress with oxygen requirement	3/7
Respiratory rate (bpm)	20 (22-29)
SpO_2/FiO_2	
Shock	7/7
Plasma leak	7/7

4.3.4.1 Performance of clinician with and without RAILUS software

Accuracy of image identification was higher in those using the RAILUS AI tool than those using the standard LUS technique: 93.4% (95% CI 89.0–97.8%) compared to 68.1% (95% CI 57.9–78.2%), ($p < 0.001$). Performance was better in all classes for clinicians using our AI tool compared to those without AI assistance as shown in Figure 4.8. In particular, A-line detection accuracy rose from 74 to 98%, and Confluent B-line detection accuracy rose from 6 to 92%.

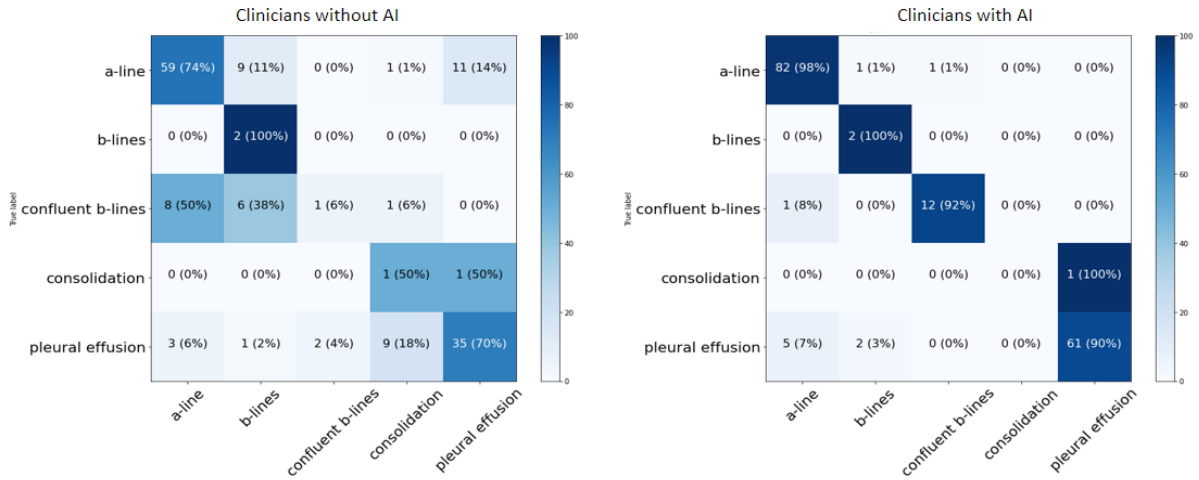


Figure 4.8. Confusion matrices of clinicians with and without RAILUS in real-time

4.3.4.2 Time spent on interpretation

When using the RAILUS software, the time required by clinicians to interpret one video was quicker compared to those who did not use the RAILUS software. The median time was 5.0 seconds (IQR 3.5–8.8) without RAILUS and 12.1 seconds (IQR 8.5–20.6) with RAILUS ($p<0.001$), respectively.

4.3.4.3 Clinician confidence in their LUS interpretation

The usability questionnaires showed an overall positive judgement of the RAILUS software when used in real time. Most clinicians (93%) found the AI–assisted tool useful in the clinical context and wanted to use the tool in the future (86%). The main suggestions included having a function to automatically report the results and a comparison function to evaluate progression/resolution of lung pathologies from serial examinations. 64% of clinicians thought the tool was useful for both real–time and post-exam evaluation, while only 7% thought it was only useful for post-exam evaluation. Interestingly, 71% of clinicians wanted the radiologist/expert to re-evaluate their interpretation with the AI tool. Moreover, 64% of clinicians felt most confident in their assessment with the AI tool enabled, compared to only 7% being most confident without the AI tool.

Regarding the concerns of clinicians when using the LUS AI tool, some clinicians (64%) were concerned about legal responsibility in the event that the AI tool made an error. 36% of the clinicians were concerned about how their results compared to the AI tool while 28% were concerned about privacy of the patient data.

4.3.4.4 Usability of RAILUS software in real-time

A summary of the findings related to perceived real-time usability and benefit of RAILUS are shown in Figure 4.9. The figure shows the distributions of the answers collected from operators after performing LUS with and without RAILUS on critically ill patients. Overall, most operators found the tool usable, useful, and beneficial.

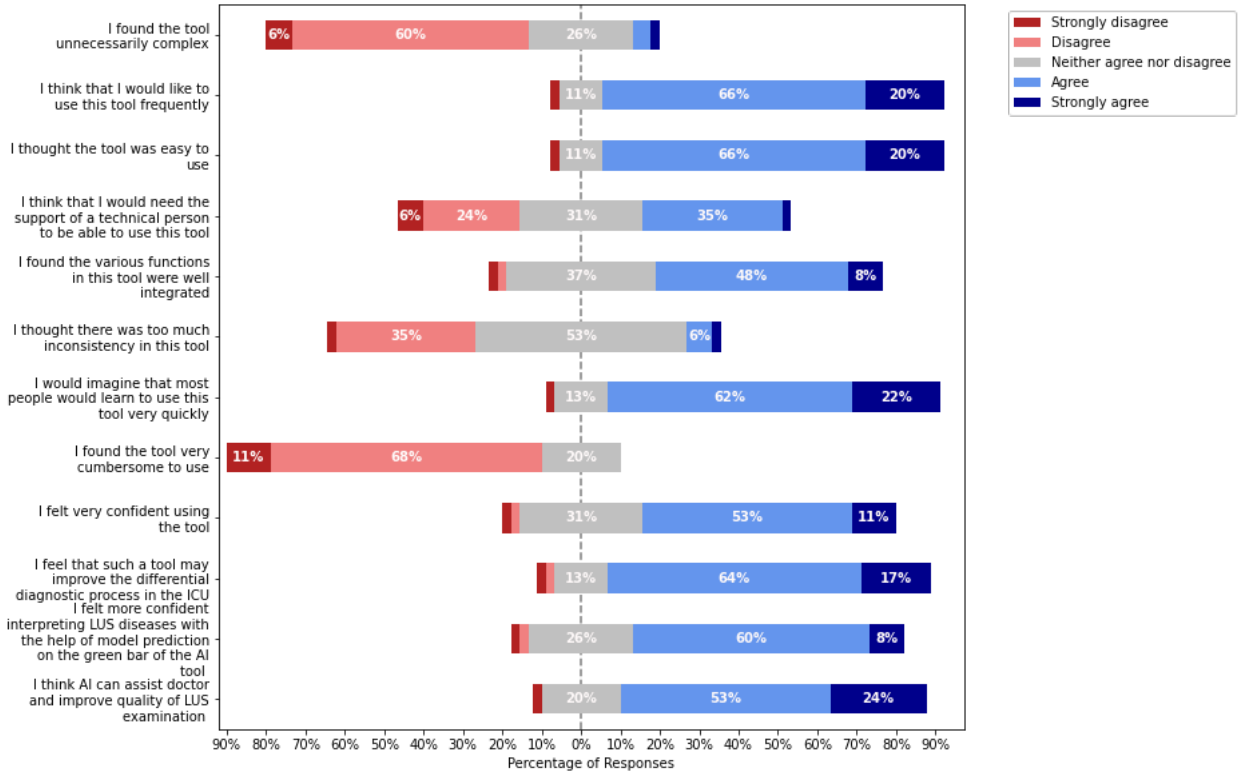


Figure 4.9. Usability of the RAILUS software in real-time, from a survey carried out after operators performed LUS with and without RAILUS.

4.4 Discussion

In this study, we have evaluated an AI-assisted LUS system in a resource-limited ICU setting. To this end, we initially assessed the baseline performance of clinicians in a LUS image classification task. Our results show that there is a significant gap between beginners, intermediate, advanced clinicians, and experts in LUS interpretation, particularly in B line interpretation. This is consistent with our survey, where the majority of the participants stated that image interpretation was their most significant challenge in performing LUS. This performance gap and the challenging nature of this task may prevent non-expert clinicians from carrying out LUS examinations in clinical practice. In low resource settings like Vietnam, there are very few experts or even advanced clinicians in LUS. As our survey in phase 1 showed, across Vietnam, only 1% of users identified as experts, hence the need for our tool. Even in our setting for phase 3 of this study –a large

teaching hospital- there are no experts in LUS. Developing an AI tool that can assist inexperienced users could be highly beneficial for patient outcomes and improving quality of care.

A crucial aim of our study was to investigate whether operators improved their baseline performance when assisted by our RAILUS AI system. Our study showed that performance improved by 15% when using RAILUS in a controlled environment (with expert-obtained clips), and by 25% when using RAILUS prospectively in real-time LUS assessment. Notably, this represented a level exceeding the baseline set by advanced clinicians in Phase 1. In addition, interpretation was approximately twice as fast when using the AI system. Of note, the clinicians using the tool were those already involved in routine care of these patients (imaging staff, infectious disease doctors and ICU staff) but they still showed a significant improvement in performance with the AI tool in both phases 2 and 3. The performance of clinicians in interpreting B-lines reduced slightly in the second phase. We note that this is mainly due to difficulty in distinguishing B lines from confluent B lines. Our User Interface (UI) allowed several possibilities to be simultaneously displayed, and commonly this meant that both B lines and confluent B lines were predicted for the same clips (although with varying degrees of certainty as represented by the green line in the UI (Figure 4.3). When using RAILUS, the ultimate decision as to which of the 5 lung US features was present was left with the clinician, and hence this introduces interesting questions about trust in AI and clinical decision making. For example, Bernstein et al [77] investigated whether incorrect AI results impact radiologist performance and found that when AI is wrong radiologists make more errors than they would have without AI. In phase 3 there were only 2 loops with B-lines, thus a small sample size from which to make definite conclusions in this cohort. We also cannot exclude other contextual influences on decision making. In future studies, sample size calculations should take the incidence rate of each lung pattern into account and also make efforts to better understand clinicians' reasons for following AI predictions (or not).

Finally, these quantitative results were supported by the post-study surveys, which revealed that the AI-assisted tool was felt to be useful in the clinical context and most clinicians confirmed they were keen to use the tool in the future. However, the concerns raised about data privacy and legal responsibility when using an AI-assisted tool are valid. As the application of AI in ultrasound

is in its infancy, there are relatively few regulations on how to legally implement it in routine healthcare practice or who will be responsible for AI-derived medical errors, particularly in low resource settings. By improving the accuracy, speed and confidence of bedside LUS, it might help clinicians in ICUs in LMICs to better manage critically ill patients with various lung pathologies. This could especially benefit the monitoring of patients during fluid resuscitation where fluid balance is critical to achieve a stable haemodynamic status without causing fluid overload e.g., pulmonary oedema. If successfully deployed, RAILUS can help better management of dengue patients in our settings, which causes a major endemic every year, however as there is only supportive care for patients with dengue shock syndrome admitted to ICU with careful fluid balance.

We believe our study represents an important step towards real-time implementation of AI in LMIC ICUs, but nevertheless has some limitations. The dataset used is from patients with severe dengue/sepsis so it remains to be seen how these results would translate to patients with other diseases. Furthermore, while we have designed our system to be agnostic to ultrasound devices, clinical validation of the tool on other types of devices (such as portable devices) is yet to be performed. Our study focused on the individual findings of LUS. Clinical practice requires a more nuanced interpretation of these findings for optimal benefit, for example, whether the confluent B-lines and consolidations are focal or non-focal to discriminate between pneumonia or pulmonary edema. For the first phase, the order of questions in the survey was not randomized but it was distributed electronically to a pool of participants distributed across Vietnam who could not share the survey with each other. In terms of clinical validation, the study did not randomize the patients to either use the AI tool or not. In addition, the order of AI and not AI was not randomized so there could be a learning effect. Future studies should explore the potential benefits of AI tools for advanced or expert users, and the regulatory/ethical and cultural issues of the clinical use of AI methods in different healthcare settings. In addition, further technical development including but not limited to AI interpretability, AI fairness, quantification of pleural effusion, or more complex LUS patterns such as pneumothorax would be beneficial fields of research.

4.5 Conclusion

This is the first study of real-time implementation of AI-assisted LUS interpretation in critically ill patients. We have demonstrated the feasibility of our system for non-expert clinicians with limited LUS experience, to acquire and interpret lung ultrasound in critically ill patients. It is also important to evaluate AI-assistance in other clinical applications and the next chapter will do this for muscle US for monitoring muscle wasting during ICU stay.

5 AI-assisted muscle ultrasound

Chapter overview

This chapter presents a prospective study conducted at the ICU in the Hospital of Tropical Diseases at Ho Chi Minh city, Vietnam. Section 5.1 provides an introduction and motivation for this work. In section 5.2 the approach to develop and evaluate an AI tool to support non-expert operators in measurement of the Rectus Femoris Cross Sectional Area (RFCSA) in muscle ultrasound is described. The aim of this AI-assisted muscle ultrasound tool is to remove the need for manual tracing to increase reproducibility and save time. Section 5.3 presents the results and Sections 5.4 and 5.5 discuss the findings of the chapter and draw conclusions.

5.1 Introduction

5.1.1 Muscle wasting in ICU patients

There is a large body of research demonstrating that patients admitted to the intensive care unit (ICU) suffer significant morbidity including functional impairments and early and rapid loss of muscle mass [151]. Loss of muscle mass contributes to muscle dysfunction and may impact overall function. However, the reasons for these changes are multifactorial and may include impaired muscle protein synthesis associated with sepsis [152]; patient comorbidity, organ dysfunction, duration of mechanical ventilation and length of ICU stay.

Measurement of muscle changes in the ICU is challenging due to patient sedation and subsequent difficulties with following commands when using traditional volitional techniques such as the Medical Research Council sum score [153]. The use of a non-volitional measure such as point-of-care ultrasound (POCUS) offers the potential to more reliably examine muscle changes [154].

Patients admitted with tetanus, of whom 50% are intubated and ventilated, administered muscle relaxant drugs and benzodiazepines, spend approximately three weeks in the ICU [114]. Patient index admission diagnosis and sequelae from the ICU admission (such as sepsis) are associated

with loss of muscle mass, weakness and impaired functional outcomes [155]. Assessment of muscle function is difficult in these patients. Therefore, the use of POCUS has gained traction and in patients with muscle failure such as tetanus allows serial monitoring.

5.1.2 Why measure Rectus Femoris musculoskeletal US?

The quadriceps femoris muscle is composed of four components: rectus femoris, vastus medialis, vastus intermedius and vastus lateralis. Rectus femoris is an ideal choice for monitoring changes in muscle mass due to its size and superficial location on the thigh. Firstly, the quadriceps femoris muscles are one the largest muscle groups in the body. Secondly, rectus femoris is the most superficial of the quadriceps femoris muscles, making it easy to locate using an ultrasound probe. Furthermore, the thigh is not an intimate part of the body and hence is more comfortable for the patient.

5.1.3 Ultrasound to assess muscle mass, RFCSA, value of RFCSA in ICU

Recently, muscle ultrasound has been shown to be a valid imaging modality for muscle assessment that is comparable to standard MRI or CT imaging whilst being safer, cheaper, and easily repeatable [156–159]. Furthermore, it is widely accessible in most ICUs and, with basic training, can be used by a variety of non-specialist clinicians. In contrast, MRI and CT require transferring the patient to the radiology department, which is high risk and may not always be possible depending on the clinical stability of critically ill patients. Therefore, muscle ultrasound is a convenient approach to investigate muscle changes over time after admission to the ICU. In a recent systematic review and meta-analysis on assessment of muscle wasting during critical illness, muscle ultrasound was used for 85% (28/33) of studies to assess muscle mass in the ICU [160].

To quantify muscle strength, the most frequently used parameters for muscle morphology are muscle thickness, fascicle length, echogenicity and muscle cross-sectional area. The CSA of the quadriceps rectus femoris muscle is one of the measures most commonly assessed as it is easily identifiable through its size and location. Puthucheary et al. conducted a study in ICU patients with serial muscle ultrasound examinations and found that muscle thickness measurements significantly underestimate ICU muscle wasting compared to RFCSA [161]. In addition, the study

also showed that RFCSA is a more reliable parameter for muscle strength in a setting where volitional and nonvolitional muscle strength are challenging to measure, such as the ICU. An example of muscle mass reduction in an ICU patient is shown in Figure 5.1.

Furthermore, recent evidence suggests that the assessment of changes in muscle size over time may improve prognostication and enhance the choice of rehabilitation interventions [162–165]. The strong association between reduced muscle mass and morbidity and mortality highlights the need for simple methods that can monitor changes in RFCSA [164, 166, 167].

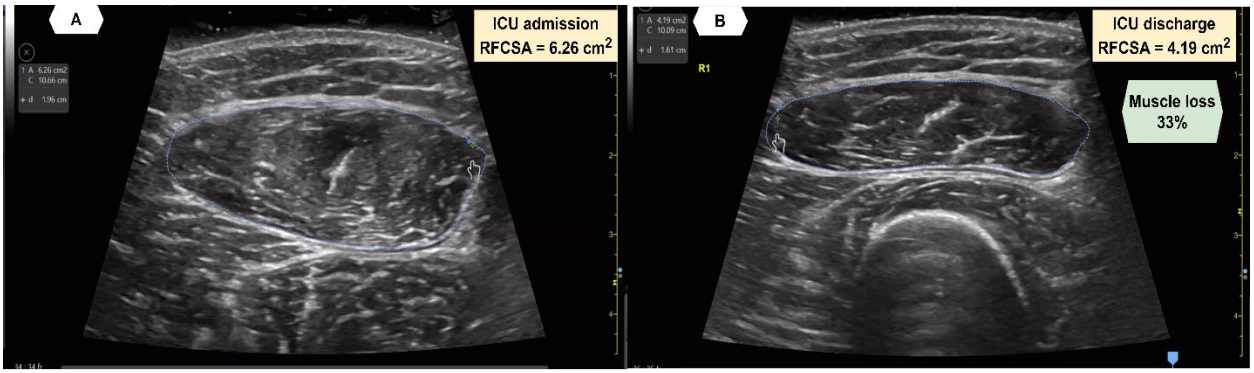


Figure 5.1. Example of muscle ultrasound of a patient in the ICU for 21 days. (A) ICU admission muscle size RFCSA of 6.26 cm^2 and (B) ICU discharge RFCSA of 4.19 cm^2. The patient had a 33% reduction in their muscle cross sectional area.

5.1.4 Muscle mass predicts ICU stay, outcome, long term functional outcome

A meta-analysis on more than 3000 ICU patients, of which 48% of the patients had intensive care unit acquired weakness (ICUAW), showed the rate of RF muscle wasting during the first week of critical illness was 2.10% for RFCSA and 1.75% for RF thickness per day [160]. Puthucheary et al. showed that muscle wasting occurred early and rapidly during the first week of critical illness with a reduction of 17.7% in the RFCSA on day 10 [152]. In the same study, patients with failure of four organs showed muscle loss of more than 27% by the end of 10 days. Similarly, Junior et al. conducted a study in severe COVID-19 patients in the ICU and observed a reduction of 30.1% in RFCSA during 10 days of ICU stay. Mayer et al. showed that RFCSA at baseline was 2.99 ± 0.99 cm^2 and that this decreased with a median percentage change of 18.5% in the 7 days from ICU admission [164]. Many studies have shown that RFCSA and other parameters such as RF

echointensity and muscle power are predictors of diagnosis of ICUAW and functional outcome function assessed at hospital discharge, 6 months and 12 months after discharge [155, 165, 168, 169]. Changes in muscle ultrasound can be used for the diagnosis of ICUAW because electrophysiological recordings or muscle biopsies are not always easily carried out in the majority of ICUs.

5.1.5 Inter- and intraobserver variability in measurement of RFCSA

The process of measurement of RFCSA from ultrasound is a time-consuming task and often suffers from significant intra and interobserver variability [157, 158, 161, 170, 171], hindering its use to inform patient management. Repeated RFCSA muscle recordings can have significant differences due to variations in probe positioning, angulation, and tilt, as well as the manual delineation of muscle by the operator. Conventionally, to enhance the accuracy of the measurement of the muscle size, the RFCSA is measured three times consecutively and an average of the measurements is calculated and saved, which increases the time taken for scanning and data acquisition [157, 158, 168, 172, 173].

In the quantitative analysis presented in this chapter we measure three types of variability in RFCSA measurement: reliability, inter-observer variability and intra-observer variability. These are described below, and these categories are taken from the guidelines in [174].

Reliability: is defined as the extent to which measurements can be replicated. In other words, it reflects not only the degree of correlation but also the agreement between measurements. Agreement parameters assess how close the results of the repeated measurements are, by estimating the measurement error in repeated measurements. Reliability parameters assess whether observers/operators, can be distinguished from each other, despite measurement errors [175]. Historically, Pearson's correlation coefficient, paired t-tests, and Bland-Altman plots have been used to evaluate reliability. However, the paired t-test and the Bland-Altman plot are methods for analyzing agreement, and Pearson's correlation coefficient is only a measure of correlation, hence, they are nonideal measures of reliability. A more desirable measure of reliability should reflect both the degree of correlation and agreement between measurements. Therefore, we argue that the intraclass correlation coefficient (ICC) is a more appropriate index,

specifically the reliability of different raters to measure subjects similarly where values < 0.5 are considered poor, 0.5–0.75 moderate, 0.75–0.9 good and >0.9 excellent [176, 177].

Inter-variability: Inter-variability refers to the variation in measurements observed between different individuals or different examiners performing the same measurements. It can occur due to several factors, including variation in ultrasound probe placement, angulation, pressure applied, and the interpretation of images. To minimize inter-variability, it is essential to establish standardized protocols and guidelines for ultrasound measurements. These protocols may include specific landmarks for probe placement, consistent angulation, and standardized techniques for image acquisition and analysis. Training and calibration sessions for examiners can also help to reduce inter-variability by ensuring a consistent approach across different operators.

Intra-variability: Intra-variability refers to the variation in measurements observed within the same individual or examiner over multiple measurements or sessions. It can arise from factors such as probe positioning, muscle contractility, anatomical variations, and measurement errors. To minimize intra-variability, it is crucial to maintain consistent probe placement and technique throughout repeated measurements. Care should be taken to avoid excessive pressure on the muscle, which can affect the measurement. Additionally, automatic segmentation software or tools that allow for image comparison and tracking of previously measured areas can help to ensure consistency and reduce measurement errors.

5.1.6 Manual delineation

Manual delineation of muscle is a subjective, time-consuming, and laborious task which normally requires a lot of experience. Overall, increasing reliability and minimizing both inter- and intra-variability in RF muscle ultrasound measurements requires standardized protocols, appropriate training, consistent technique, and regular quality control procedures. These measures help improve the reliability and reproducibility of measurements, allowing for more accurate assessment and monitoring of muscle characteristics over time. It is worth noting that despite efforts to reduce inter- and intra-variability, some degree of variability may still exist. It is recommended to report the level of variability observed in studies to provide a better understanding of the reliability and consistency of the measurements. For example, in the study

conducted by Zhang et al., the correlation coefficient of measurement accuracy of two researchers was 0.90 for the RF muscle [166]. Conventionally, to enhance the accuracy of the measurement of the muscle size, the CSA is measured three times consecutively and an average calculated to produce the final value. Inter-investigator comparisons have revealed ICCs, standard error measurements (SEM) and mean bias ranging from 0.85 to 0.999, 0.07 to 1.16 cm^2 (0.9–7.6%) and − 0.16 to 0.66 cm^2 (− 0.6 to 3.2%) respectively for RFCSA measurement. Intra-investigator comparison revealed ICCs, SEMs and mean bias between 0.883–0.998, 0.07–0.93 cm^2 (1.1–7.6%) and − 0.80 to 0.15 cm^2 (− 3.4 to 1.8%) respectively [178].

5.1.7 Semi-Automated Segmentation of the RFCSA

To tackle the challenge of inter- and intraobserver variability in muscle segmentation, AI techniques have been proposed. Chen et al. [179] developed a CNN model to automatically segment the RF in ultrasound images. However, although it achieved good accuracy the proposed method needed 0.2 seconds to segment a single frame, making it unsuitable for real-time processing. Ritsche et al proposed a semi-automatic tool named ASCAuto [178] that performs RFCSA evaluation in panoramic ultrasound. The authors later extended their work to incorporate a deep learning-based technique (DeepACSA) [72]. However, panoramic ultrasound is not always available in all ultrasound machines and requires experts to operate. Furthermore, the tools described in [29,30] are also not suitable for real-time analysis. Recently, Katakis et al developed and evaluated a Transformer-based model for automatic segmentation of the RFCSA [180]. The results showed that the model achieves high ICC and high Pearson's correlation when compared with manual measurement in a validation dataset, but the validation remained offline on a retrospective dataset.

In summary, AI techniques have been proposed for RFCSA measurement [72, 179, 180] they have typically been designed and evaluated for offline use, i.e, they are not suitable for bedside use and real-time analysis. Moreover, to date these tools have been subject to limited validation in clinical settings.

In this study, we aimed to investigate the feasibility of using a real-time AI-assisted tool for RFCSA measurement from muscle ultrasound that would be suitable for clinical use, particularly in an

LMIC setting. We hypothesized that this tool would have improved reproducibility and reduced interobserver variability compared to current methods. We tested our tools in a cohort of patients with tetanus as this is a group of patients in whom muscle wasting is an important problem with long ICU stays and in whom such a tool would be used if proved reliable.

5.2 Proposed approach

5.2.1 Study design and participants

This was a prospective observational study to test the reliability of AI-assisted measurement of RFCSA from muscle ultrasound at the patient's bedside compared to standard ultrasound. The study was conducted in the adult ICU at the HTD Ho Chi Minh City, Vietnam. Measurements were performed in adult patients with severe tetanus (Ablett Grade 3 or 4) admitted to the Adult ICU at HTD expected to stay at least 5 days. Patients were receiving standard treatment including mechanical ventilation, muscle relaxation and neuromuscular blockers following the hospital guideline [114]. All patients or their representatives provided informed consent to take part in the study. Ethics approval was obtained from the HTD Ethics Committee, and the OxTREC. The study was registered at ClinicalTrials.gov number NCT06034093. We followed the guidelines outlined in the CONSORT-AI Extension checklist [181] in the reporting of the study.

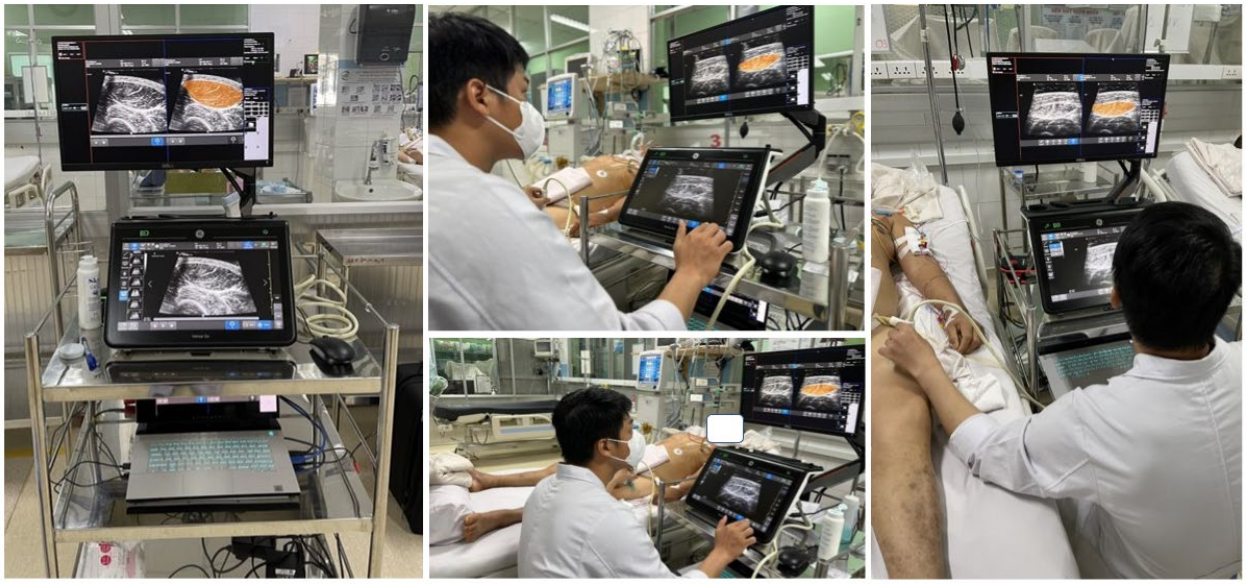


Figure 5.2. Real-time AI-assisted muscle ultrasound (RAIMUS) system

5.2.2 Real-time AI-assisted muscle ultrasound (RAIMUS) software

The developed AI assistant is described in detail in Appendix D and named RAIMUS (Real-time AI-assisted Muscle UltraSound) and was developed by the technical team. We deployed a U-net architecture for the RFCSA sematic segmentation task as it has shown good performance for image segmentation in several medical imaging modalities even if the datasets were small. The main architecture consists of contracting and an expanding paths. The contracting path (or so-called encoder) consists of convolution and max pooling layers whereas the expanding path (or decoder) uses transposed convolutional layers. Skip connection are used between layers of the same resolution in these two paths. The model outputs a pixelwise binary label (background and CSA). The input images for the model were resized to 256 x 256 pixels. Augmentation was applied on-the-fly during training including rotation, horizontal flipping, zoom in/out and increase/decrease of contrast gain. In this study it was deployed in real-time using the PRETUS tool [150]. The ultrasound machine HDMI output was connected to the laptop via a USB frame grabber. This allowed the user to use an external screen with an AI overlay instead of the screen of the ultrasound machine.

The interface to RAIMUS, illustrated in Figure 5.2, is as follows. On the right of the screen, there is a widget containing information from the automatic muscle segmentation, including the

muscle delineation continuously overlaid onto the ultrasound image and the corresponding cross-sectional area in cm^2. The segmentation overlay and related information can be enabled or disabled by the user.

5.2.3 Study procedures

After informed consent was obtained, the participants were randomized to either standard examination (i.e. the operator performed manual segmentation without AI assistance) or the RAIMUS tool (i.e. AI-assisted RF segmentation). Image acquisition (guidance to the chosen muscle view) was manual in both cases. Randomization was done using a computerized protocol at a ratio of 1:1. Each participant was scanned three times by different operators (ICU admission, 7 day and ICU discharge). The flowchart of this study is illustrated in Figure 5.3.

All measurements were carried out according to a standard operating procedure where patients were in the supine position with the leg in neutral rotation. Patients not receiving muscle relaxants were reminded to relax the muscle. Measurements were taken using a 12L-RS linear probe and a Venue Go ultrasound machine (General Electric Healthcare, London, UK) from a location which was three-fifths of the distance from the anterior superior iliac spine to the superior patella pole. This position was used as a landmark for subsequent measurements to provide consistency and allow reliable comparisons to be made over time. The transducer was placed perpendicular to the skin and transversally in relation to the longitudinal axis of the thigh to observe the cross-sectional area of the muscle. An excess of ultrasound gel was placed when performing the muscle scan and the pressure on the skin was kept minimal to ensure good image quality [155, 160, 182]. For each examination 3 separate measurements (scan-rescan) were made for each leg (removing the probe between each one). The examinations' durations were recorded.

The operators selected were 3 clinicians and 2 nurses, all 5 with limited training in muscle ultrasound as our target users for the AI tool are non-expert operators to reflect the common setting in a LMIC. We provided muscle ultrasound training and RAIMUS software training to allow operators to use the tool effectively. Each operator was asked to scan five patients (2 legs, each

leg 3 repeat scans) as part of their muscle training with the AI software. The images were saved and manually delineated. with the AI software.

For patients in the standard measurement arm RFCSA was determined by manual delineation of the cross-sectional image of the muscle. For patients in the AI-assisted imaging arm measurements were made in real time using the automated AI tool.

5.2.4 Evaluation of reliability

We assessed three types of variability including (1) scan-rescan variability, (2) intraobserver variability in delineation and (3) interobserver variability in delineation (Figure 5.3). To assess (1) scan-rescan variability, the operators were asked to scan each leg three times for each of the two allocated methods. To assess (2) intraobserver variability in delineation over time, the same operator subsequently delineated each image one further time 2-4 weeks after the images were acquired (from stored raw images). To assess (3) interobserver variability in delineation, each image acquired by the first operator was delineated by 2 additional operators.

The examination durations (measured from when the operators put the probe on the leg of patient to when they finished delineations and measurements) were compared between the 2 methods.

5.2.5 Sample size

The sample size for this study was estimated following Walter et al [183] with the minimum acceptable reliability (Intraclass correlation coefficient ICC) (ρ_0) of 0.9, expected reliability (ICC) (ρ_1) of 0.96, significance level two-tailed (α) of 0.05, Power ($1 - \beta$) of 80% and the number of raters/measurements per subjects (k) of 3. After using the formula, the minimum required sample size was 27 examinations for each group. With the expected dropout of 10% the total sample size used in the study was 30 examinations for each group.

5.2.6 Statistical analysis

All statistical analyses were performed with R version 4.0.4. Continuous variables are expressed as mean ± standard deviation (SD) or as median (interquartile range), according to the symmetricality of the data distribution, and compared using an unpaired Student's t-test or Wilcoxon rank sum test, as appropriate. Categorical data, presented as numbers and

percentages, were compared using the χ2 test. *P* values less than 0.05 were considered statistically significant.

The variability in RFCSA measurements was assessed using the two-way random effects for multiple raters/measurements ICC with 95% CI (ICC (2, k) [176]. The standard error of measurement (SEM) was also calculated to make a judgment about the degree that measurements vary for an individual. The SEM values indicate the precision of the measurement and were calculated based on the ICC and the SD of the mean of differences between the two measurements $SEM = SD\sqrt{1 - ICC}$. There was no measure-remeasure variation for the automated AI software because the model always outputs the same measurement and hence the same RFCSA result. A modified Bland-Altman analysis for multiple observers in a single plot, as proposed by [184] was used to assess the agreement between RFCSA measurements. The examination duration was compared between the 2 methodologies of measurement using an unpaired Student's t-test.

5.3 Results

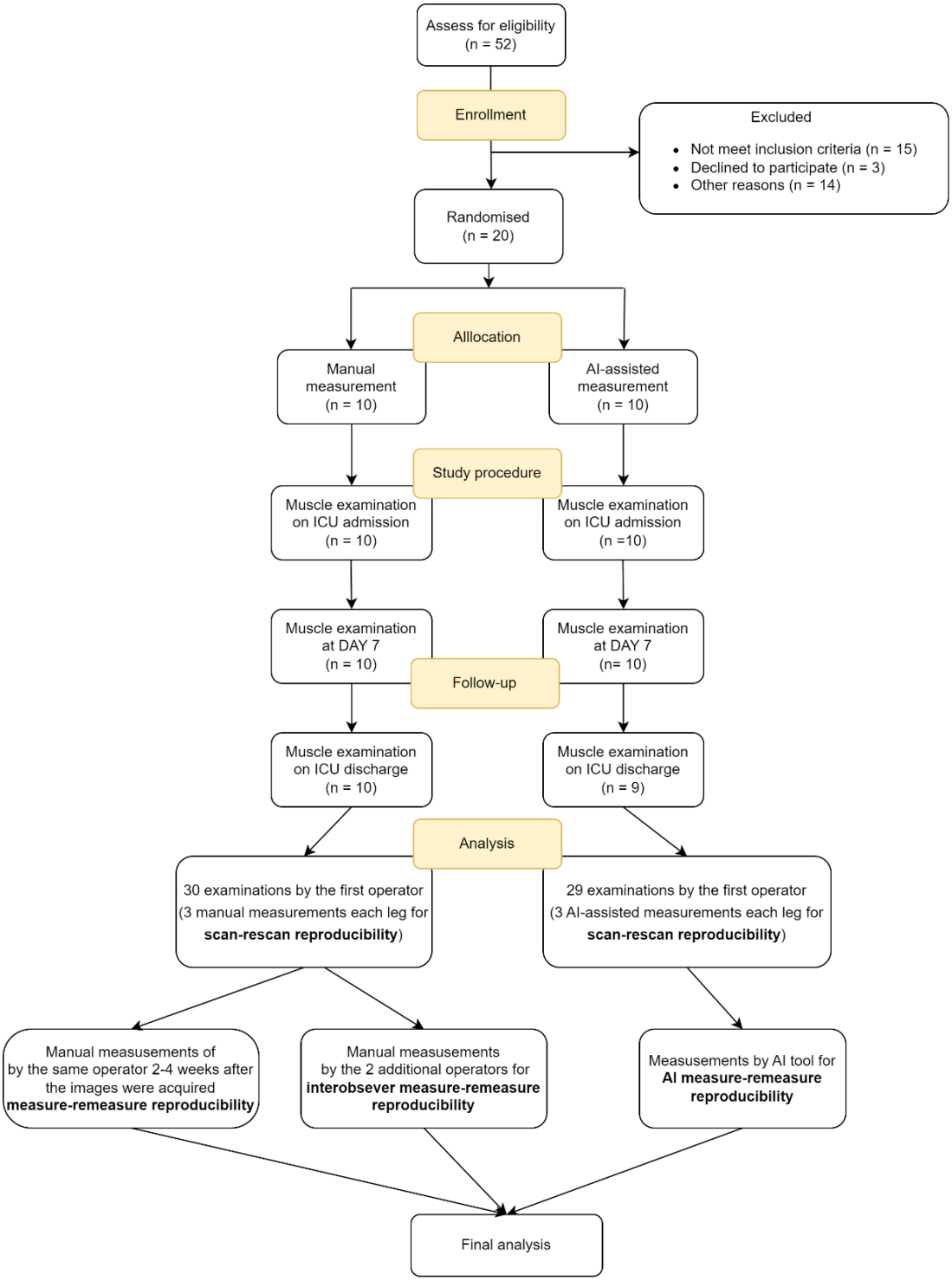


Figure 5.3. RAIMUS study flowchart

We enrolled 20 patients with tetanus at the Adult ICU at HTD between Feb 2023 and July 2023. The mean ± SD age of patients in the AI group and non-AI group were 67 ± 13 years and 56 ± 17 years, respectively. Two (20%) patients in each group were female. 17 (85%) patients had at least one episode of hospital acquired infection (HAI) during ICU stay. The ICU stay and hospital stay were comparable between the two groups (**Table 5.1**). In total 59 muscle ultrasound examinations were carried out, 29 examinations with the AI tool and 30 examinations without the AI tool. After visual inspection of 29 examinations with AI, 28 examinations were successfully delineated by the AI tool and one examination was rejected by the expert. All examinations without the AI tool were accepted by the expert. The average muscle loss during ICU was similar in the two groups, 26 ± 15% for the AI arm and 23 ± 11% for the non-AI arm. The full characteristics of the patients are shown in **Table 5.1**.

Table 5.1. Characteristics of patients in the RAIMUS study (n = 20)

	AI arm (n=10)	**Non-AI arm (n=10)**
Age	67 ± 13	56 ± 17
Sex (female)	2 (20%)	2 (20%)
Comorbidities (1 or more)	6 (60%)	4 (40%)
Sedative use during ICU	10 (100%)	10 (100%)
Use of non-depoplarising neuromuscular blocking agents during ICU stay	10 (100%)	10 (100%)
Length of ICU stay (days)	26 ± 11	24 ± 5
Length of hospital stay (days)	32 ± 13	31 ± 6
Mechanical ventilation duration (days)	19.7 ± 8.2	17.8 ± 5.5
ANSD duration (days)	12.4 ± 6.7	12.0 ± 2.8
Total dose of Pipecuronium	438 ± 190	430 ± 250
Enteral nutrition	10 (100%)	10 (100%)
Rehabilitation duration (days)	10 ± 9	8 ± 4
HAI during ICU stay	8 (80%)	9 (90%)

RF CSA D1 (cm^2)	4.37 ± 1.08	4.76 ± 1.50
RF CSA D7 (cm^2)	4.09 ± 1.01	4.51 ± 1.59
RFCSA Discharge (cm^2)	3.25 ± 1.24	3.65 ± 1.15
% change in RFCSA during ICU stay (%)	26 ± 15	23 ± 11
ICU survival	10 (100%)	10 (100%)

(MV: Mechanical ventilation, ANSD: Autonomous Nervous System Dysfunction, HAIs: Hospital-Acquired Infections, RFCSA D1: Rectus Femoris Cross sectional Area on ICU admission, RFCSA D7: Rectus Femoris Cross sectional Area on Day 7, RFCSA Discharge: Rectus Femoris Cross sectional Area at ICU Discharge)

Scan-rescan variability

The scan re-scan variability of the AI group was lower compared to the non-AI group (ICC 0.999 95%CI 0.998 - 0.999, vs ICC 0.982 95%CI 0.962 - 0.993).

Figure 5.4 shows modified Bland-Altman plots illustrating the percentage difference in three repeated RFCSA measurements from the mean. The figures showed better agreement in the AI arm where the limits of agreement were lower in the AI group (± 1.9%) compared to the non-AI group (± 6.6%).

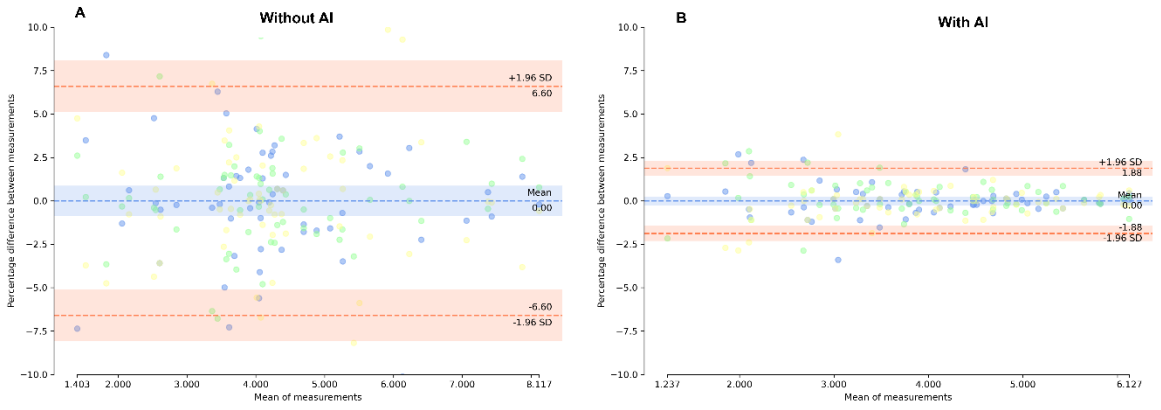


Figure 5.4. Plot of scan-rescan agreement in RFCSA. A) Without AI. B) With AI. Horizontal dotted lines indicate the limits of agreement from the mean (LoA) of the three measurements. Some

symbols are superimposed. The percentage differences of all measurements with the mean (y-axis) are plotted against the mean RFCSA for all participants (x-axis). The horizontal dashed lines indicate the limits of agreement with the mean of the three repeated measurements and were ±6.6% for the non-AI group and ±1.9% for the AI group.

Intraobserver variability in delineation over time

The manual intraobserver in delineation (initial vs 2-4 weeks later) resulted in good reliability with an ICC of 0.984 (95% CI 0.973, 0.990). The modified Bland Altman plot for intraobserver agreement results is shown in **Figure 5.5** (left). The intraobserver agreement without AI was ±5.9% and there was no intraobserver variation for the AI group for the reason stated in the statistical analysis section.

Interobserver variability in delineation

The manual interobserver ICC was 0.974 (95% CI 0.965 - 0.981). The interobserver agreement results are shown in **Figure 5.5** (right). The limits of agreement without AI were ±8.2% and there was no interobserver variation for the AI group for the reason stated in Section 5.2.6.

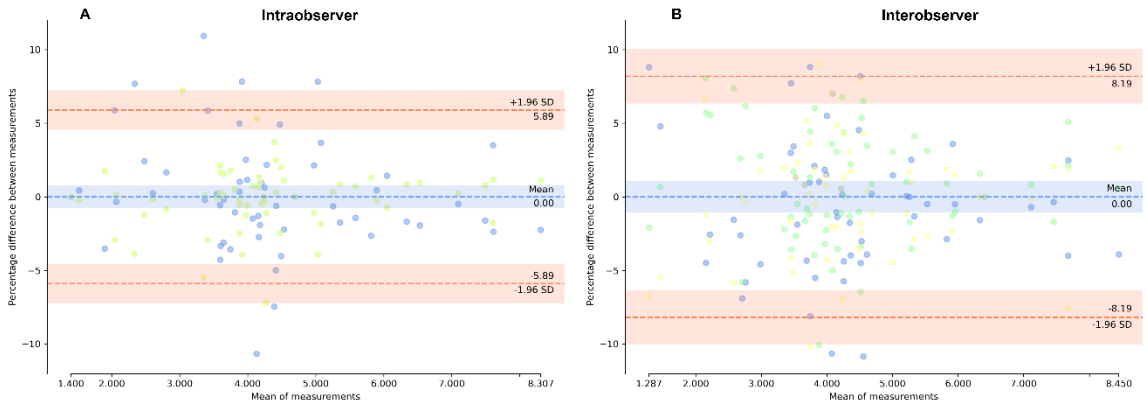


Figure 5.5. Intraobserver agreement plot (A) between the same operator over time and interobserver agreement plot (B) between 3 observers in RFCSA measurement. Observers represent different symbols. The percentage differences of all measurements with the mean (y-axis) are plotted against the mean RFCSA for all participants (x-axis). The horizontal dashed lines

indicate the limits of agreement with the mean of the three observers and ranged from ±5.9% for intra-observer variability (left) and from ±8.2% for interobserver variability (right).

Examination duration

Examination duration (including acquisition and measurement) was shorter in the AI group compared to the non-AI group: a median of 9.4 minutes (IQR 7.2–11.7) compared to 19.6 minutes (IQR 16.9–21.7) ($p < 0.001$).

5.4 Discussion

This chapter has presented, for the first time, a prospective study to evaluate the impact of an AI tool for real-time RFCSA estimation compared to the traditional manual measurement technique for monitoring muscle mass in the ICU. The AI tool succeeded in supporting operators in assessing muscle wasting in patients with a range of RFCSAs and varying image qualities by correctly delineating the RF muscle and measuring the RFCSA with less variability than standard non-AI measurements. Furthermore, the time spent on measuring RFCSA using the AI tool was approximately half that of standard measurements.

The study showed a reduction in scan-rescan variability in the AI arm. This variability may involve both acquisition variability (taking the probe back to the same plane every time) and intraobserver delineation variability. In addition, the scan-rescan variability, intraobserver and interobserver delineation variability are similar which indicates that the main variability of the standard technique is manual delineation. This suggests that it may be possible using the AI tool to perform a 1-scan measurement instead of standard average of 3-scan measurement, further reducing time.

Thus, with the help of the automated AI tool, monitoring of muscle changes in ICU patients could be more practical and feasible than before. The reliability of operators with limited training in our study was already good without the AI tool but the time spent on manual measurements was twice that when using the AI tool. However, further research should be focused on the use of AI to guide accurate and reproducible probe placement.

Limitations

It is important to emphasize that removing the measure-remeasure variability and interobserver variability completely by the use of a deterministic AI model makes repeated measurements more reproducible but does not necessarily make them more accurate. Although the performance of our AI tool is good, future work should more thoroughly assess its robustness and bias, especially in the presence of poor image quality. Further research should also focus on validation of the AI tool in different patient populations, using different machine manufacturers and different settings, for example, monitoring muscle wasting in patients with cancer.

Echogenicity was not investigated in this study. When muscle echogenicity increases, determining the muscle boundaries is very challenging because muscle tissue is replaced by intramuscular fibrous and fat tissue. As a result, the contrast between the muscle boundaries and other structures decreases. Future work should develop AI based methods for assessing muscle echogenicity as this can provide useful information on both quality and quantity of the muscle.

5.5 Conclusions

Real-time AI-assisted muscle ultrasound removes the need for manual tracing, increases reproducibility, and saves time. Our study has shown that much of the variability between measurements is related to manual delineation of the muscle and hence potentially an even faster single-scan protocol could be adopted for AI-assisted RFCSA measurement. Such a system would significantly assist routine clinical monitoring of muscle changes in ICU patients and help in assessing the effectiveness of interventions.

6 AI-assisted cardiac ultrasound

Chapter overview

This chapter presents a preliminary prospective study to evaluate the benefit of AI-assisted left ventricle segmentation conducted at the ICU in the Hospital of Tropical Diseases at Ho Chi Minh city, Vietnam. The introduction in Section 6.1 explains the role and utility of point-of-care echocardiography in the diagnosis and management of cardiac involvement in critically ill patients in the ICU and the motivation for this study. Section 6.2 describes the study design which aims to evaluate an AI tool to support non-expert operators in acquiring high-quality standard apical 4 chamber view echocardiography videos. Section 6.3 presents the results and Section 6.4 discusses the key findings.

6.1 Introduction

Point-of-care echocardiography (POC echo) has shown to be a useful tool in cardiovascular diagnosis, management and follow-up in critically ill patients at the bedside in ICUs and emergency departments [[9, 16, 28, 35]. Its relatively low cost, real-time nature, and its portability make it an imaging modality of choice in ICU settings. Several studies have demonstrated the utility of POC echo in the ICU for answering specific clinical questions (e.g., whether the patient has any structural or functional cardiac abnormalities) and making optimal decisions, especially around fluid management for patients [16, 185]. Several organizations have established guidelines [186–190] for both standardized training and practice for POC echo. However, barriers to implementing wider use of POC echo in low-resource settings include limited training programs, time constraints within training programs, a paucity of faculty with credentials for supervision, the need for established national quality assurance protocols, and a lack of standardized assessments [191, 192]. Due to these barriers, most intensivists/treating clinicians in low-resource ICUs do not routinely perform POC echo or continue to use it with minimal training and oversight [193, 194], thus underscoring the need for alternative tools to democratize the use of POC echo in low-resource settings. The use of US imaging requires

significant expertise on the part of the operator as well as legal certification. Unlike in high-income countries, POC echo exams for patients in low-resource ICUs still rely on radiologists and cardiologists, who are not always available. This lack of availability often leads to a delay in the examination. Delays in undergoing diagnostic examinations not only add to the frustrations felt by both treating clinicians and patients but are often associated with adverse outcomes. Additional challenges in the ICU can be attributed to obtaining optimal views and image quality, due to patients being unable to move, restricted acoustic windows between the ribs and often being on respiratory support.

6.1.1 Related works

6.1.1.1 Real-time automatic EF measurement

Left ventricular ejection fraction (LVEF) is the most commonly used cardiac functional parameter to evaluate critically ill patients and to adjust treatments [190, 195]. The major strength of LVEF is that it is a universally known and accepted parameter, not only by cardiologists but also by general physicians and intensivists. A further advantage is that in clinical practice, LVEF can be estimated noninvasively from echocardiographic images, by delineating the ventricular chamber over the cardiac cycle. LVEF represents the proportion of blood in the ventricle at end diastole that is pumped out of the ventricle during systolic contraction.

Other commonly used cardiac functional parameters include stroke volume (SV), which is the *amount* of blood that is expelled from the ventricle with each contraction, i.e. SV = EDV - ESV. Hence, SV is related to LVEF as follows: LVEF = SV / EDV. Cardiac output (CO) is the amount of blood that the heart pumps per minute [130, 185]. CO = HR x SV, where HR is the heart rate. An example of how to estimate SV, EF, and CO can be found in Figure 6.1. LVEF = (EDV – ESV) / EDV, where EDV and ESV are the end diastolic volume and end systolic volume respectively.

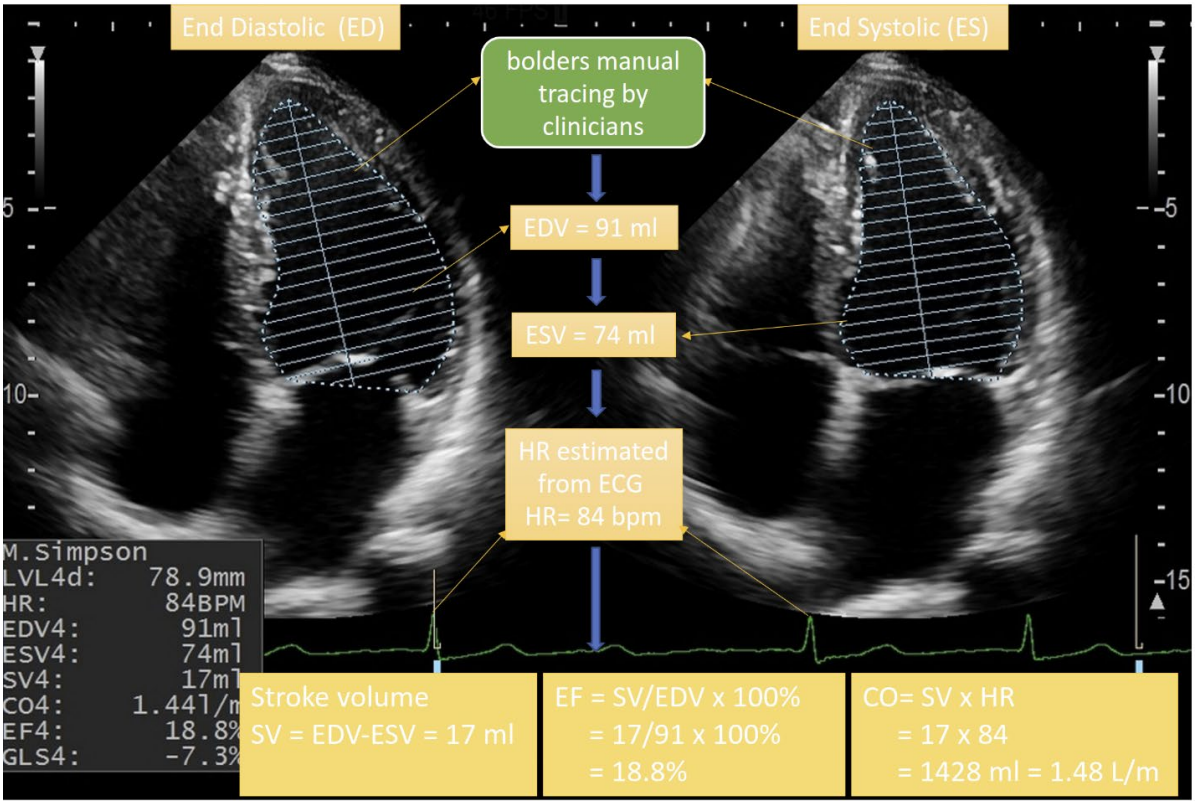


Figure 6.1. Example of computation of the SV, LVEF, CO from 4 chamber view, using Simpson's method from the ultrasound machine. The process includes 5 steps: Step 1 – Identifying the End Diastolic (ED) frame (based on ECG if available or based on experience of user), Step 2: Tracing the endocardium (inner LV border) of the LV at ED for measuring the ED volume (EDV), Step 3: Identifying the End Systolic (ED) frame (based on ECG if available or based on experience of user) of the same cardiac cycle, Step 4: Tracing the endocardium of LV at ES for measuring ES volume (ESV). Step 5: Calculate the SV, EF and if ECG is available the CO will be estimated by multiplying the SV and heart rate (HR).

As described above, quantitative evaluation of LVEF requires measurement of end-systolic and end-diastolic volumes, which can be derived from imaging by using traced endocardial boundaries at these 2 phases of the cardiac cycle. From these 2D contours, the volumes can be calculated using, for example, the modified Simpson's rule [130]. However, boundary identification is prone to errors due to suboptimal image quality, artifacts, and unusual LV shape in different pathologies, all resulting in considerable interobserver variability. This is time-

consuming and operator dependent so many groups have proposed methods to automate the contouring process. Ideally, such automated tools would work quickly and be integrated into the acquisition process, to make them suitable for use at a patient's bedside in the Emergency Department of the ICU.

Early methods for automatically contouring the LV endocardial border were based on classical machine learning [196, 197] However, such approaches were not generally fast enough to be suitable for real time use (> 10 s per frame). Recently, deep learning-based methods [67, 104, 198] have achieved state-of-the-art results while also being very fast (< 1 s per frame). Furthermore, recently a number of companies have incorporated automated LVEF calculation capabilities into their commercial products (e.g. Caption AI, Kosmos US, Venue Go GE Healthcare).

Examples of classical ML approaches for LV contouring include Knackstedt et al., who utilized a ML algorithm for automated measurements of LVEF and compared them with manual findings in 255 patients [199]. The analysis time for the automated measurements was 8 ± 1 second per patient. The interclass correlation coefficients and Bland-Altman analysis showed good agreements with automated LVEF (ICC from 0.78 to 0.86), local or reference center manual tracking, but not visual LVEF calculations. Cannesson et al. explored the role of AI for rapid and reproducible measurements of LVEF in 218 patients [196]. The ML approach correlated well with manually calculated LVEF ($r = 0.98$; 6% limits of agreement) and required less time per patient (48 ± 26 s vs 102 ± 21 s, $p < 0.01$). In addition, it correlated well with visual estimates of LVEF by expert readers ($r = 0.96$, $p < 0.001$). Another study by Rahmouni et al. suggested further testing of ML-based protocols for LVEF assessment prior to their routine use [197]. The study utilized an ML-based method (AutoEF, Siemens Medical Solutions, Erlangen, Germany) of LVEF measurement and compared it with visual estimates, manual planimetry, and cardiac magnetic resonance (CMR). The visual measurements by an expert ($r = 0.86$) and novice reader ($r = 0.80$) correlated more closely with manual planimetry using Simpson's method than did AutoEF ($r = 0.64$).

More recently the research literature on automated echocardiography contouring has shifted towards the use of deep learning models. For example, EchoNet-Dynamic [67] is a deep learning algorithm developed using 10,030 echocardiogram videos to estimate LVEF and classify patients with heart failure and has similar accuracy to that of experienced cardiologists. As stated earlier, human interpretation of echocardiograms relies on manual segmenting of the LV in the end diastolic and end systolic phases, which can have high interobserver variability. The EchoNet-Dynamic algorithm incorporated information across multiple cardiac cycles and accurately segmented the LV, predicted LVEF (mean absolute error 4.1%) and classified heart failure with reduced LVEF (AUC 0.97). This performance was validated using an independent dataset and was more reproducible than that of clinicians. Smistad et al proposed a real-time and fully automated system based on several deep learning components, such as view classification, cardiac cycle timing, segmentation and landmark extraction, to measure LV volume, and LVEF [198]. The model was trained with a data set of 500 patients from an outpatient clinic and evaluated on a separate data set of 100 patients from another clinic, where LV volume and LVEF were measured by an expert using clinical standard protocols and software. The bias and standard deviation of the automatic LVEF measurements were -3.6 ± 8.1%, while the mean absolute difference was measured at 7.2%, which are all within the range of interobserver variability (±15% and 10%). In this context, interobserver variability (also known as interrater reliability) quantifies how much two or more different observers' results vary when they measure or evaluate the same LVEF from A4C video independently. This variability can arise due to differences in the level of expertise or experience of the observers, interpretation of what they are observing or the methods or tools they use for the observation. Studies have showed that LVEF results can differ by up to 15% in either direction from the average value and the mean absolute difference in results between observers can be as much as 10% [67]. The proposed real-time pipeline allowed for a continuous acquisition and measurement workflow without user interaction and had the potential to significantly reduce the time spent on analysis. He et al conducted a prospective pilot study in 47 patients in the emergency department by utilizing the real-time low-cost hardware point-of-care echo (by connecting the card-based GE Venue Go machine with the Raspberry Pi to run their deep learning model predictions) for assessment of LVEF and video quality that achieved an

AUROC of 0.92 (0.89-0.94) for classifying between normal and abnormal cardiac function and an AUROC of 0.81 for video quality classification [105].

6.1.1.2 Real-time image quality assessment, recognition of cardiac views and probe guidance

Techniques for the calculation of cardiac functional metrics, as described in the previous section rely largely on the acquisition of good quality echocardiography images from standard views.

The acquisition of the A4C view can be broken down into three steps. First, the users need to find the correct intercostal acoustic window where the anatomical apex of the heart is closest to the transducer. In the second step, users need to rotate and tilt the transducer around the left ventricular centerline (between the apex and the center of the mitral valve) to produce a standard A4C imaging plane with minimal foreshortening. The last step consists of optimizing the images for both anatomical correctness and image quality to establish a more detailed cardiac examination.

However, a major limitation of echocardiography is that the process of acquiring images in a clinical setting presents inherent difficulties related to the skill and experience of the operator [200, 201]. Without mechanisms in place to guarantee consistent collection of high-quality images, it becomes challenging to achieve precise diagnoses or make comparisons with previous scans of the same patient [201].

These challenges give rise to uncertainty in the reliability of clinical metrics measurements such as EF, which affects the reliability of diagnostic results [202]. Currently, the evaluation of echocardiography image quality is largely subjective, and precisely what factors contribute to this quality is not fully understood [[203]. Therefore, the challenge of reliably acquiring diagnostic quality images in a clinical setting inhibits the adoption of echocardiograms, as a reliable imaging modality for cardiac diagnosis despite its many advantages. This is particularly evident when less experienced operators are involved, where the reliability of measurements can potentially lead to inconsistent assessments in patient management and clinical trials.

As noted above, the standard clinical method of image quality assessment is a subjective, manual process, where a cardiologist/sonographer/radiologist visually inspects the images and decides on what anatomical features present in the image are pathologically and clinically relevant. This process can have significant variability in clinical opinions and decisions, for example when operators assess multiple images or when an image is retrospectively reassessed by the same operator [202].

A number of studies have attempted to measure or assess the variability in echocardiography image quality and the impact of AI assistance on this. For example, Hong et al [204] conducted a non-inferiority pilot study with 4 novice operators who had no prior echo experience using AI-guidance for acquisition of A4C images. The study compared the image quality score produced by an AI tool (real-time guidance for A4C view acquisition) versus a quality score made by trained sonographers with respect to an expert assessment. The study findings suggest that the AI tool can recognize nuances of widely varying images during scanning, capturing the sonographer's expertise. Pasdeloup [205] developed and evaluated a DL-based tool for real-time AI-guided apical view acquisition. Training data were generated by slicing 3D ultrasound volumes, which permits simulation of the movements of a 2D transducer. The models were further trained to calculate the transducer position using a regression-based approach. The method was validated and tested on 2D images from several data sets representative of a prospective clinical setting. The method proposed an adequate transducer movement 75% of the time, when averaging over all degrees of freedom and 95% of the time, when considering transducer rotation solely.

A real-time automatic foreshortening detection using DL was proposed in [198]. The tool provides real-time feedback to the operators if the image represents a foreshortened view, as well as feedback on how to adjust the transducer position to reach the target standard apical view and a moving bar for prediction of standard view (either A4C or A2C). The proposed real-time pipeline allows for a continuous acquisition and measurement workflow without user interaction and has the potential to significantly reduce the time spent on analysis and reduce measurement error due to foreshortening, while providing quantitative volume measurements in the echocardiography lab.

There are currently very few studies comparing POC echo image acquisition and interpretation with AI among inexperienced users. Baum et al [206] conducted an RCT to compare internal medicine residents using an AI-enabled device (Echonous) with a non-AI device (Butterfly) in image acquisition and interpretation. The results showed that the AI group had faster scan times (72 seconds [IQR 38-85] versus 85 seconds [IQR 54-166]; p=0.01), higher image quality scores (4.5 [IQR 2-5.5] versus 2 [IQR 1-3]; p<0.01) and correctly identified reduced systolic function more often (85% vs 50%; p=0.02) compared to the non-AI group.

Clinical validation of the impact of AI tools on echocardiography image acquisition was conducted in a study conducted by Narang et al as well as several case studies reported in Covid-19 patients [107, 207]. Narang et al reported that their algorithm (Caption Health Inc., Brisbane, CA, USA; now part of the GE Healthcare family) can help non-expert operators to acquire recordings of diagnostic quality [107]. This was an FDA cleared AI-guided US system, providing real-time guidance on how to position and manipulate the transducer on a patients' body, automated quality assessment and automatic calculation of EF from any combination of up to three cardiac views commonly acquired at the point of care: parasternal long-axis, A4C, and apical 2 chamber. In this study, 8 nurses without prior experience in echo used a Caption AI enabled ultrasound machine to scan 30 patients. They found AI could help the nurses acquire diagnostic images that were suitable for evaluation of several cardiac ultrasound metrics including LV size, function and presence of pericardial effusion. This study showed the potential for translation to implementation in resource limited settings. The same system was tested with 19 medical students without previous ultrasound knowledge to scan real patients, who reported the algorithm was helpful in acquiring the A4C and A2C views [208]. Each student was trained in the basics of echocardiography in a 2.5 hour online video tutorial and then scanned three patients with the help of the AI tool. The study showed an excellent agreement between the machine's LVEF calculations from images acquired by the novices with the LVEF measured by experts (bias of 3.5% ± 5.6 and r = 0.92, p < 0.001). However, these studies lacked a control group of users who were scanned without the AI tool's assistance.

In summary, in parallel with technological advancements, the application of AI to optimise echo image acquisition has increased significantly but has been largely developed and tested in high-

income countries. The benefits of AI applications to improve the acquisition of diagnostic-quality images by less experienced operators with minimal training may be of particular importance in low-resource settings.

To address this gap, in this study, we aimed to test the feasibility of an AI-assisted LV segmentation tool for assisting non-expert operators to acquire high quality A4C view videos and compare subsequent clinical metrics between non-expert and expert acquisitions.

6.2 Proposed approach

6.2.1 Study design and participants

This was a prospective study to evaluate the feasibility of using AI-assisted LV segmentation for assisting non-expert operators to acquire high quality 4-chamber view videos. The participants were 30 healthy volunteers, age > 18 years and with no known cardiovascular diseases. All participants provided written informed consent to participate in the study. The study was conducted at HTD and approved by OxTRECT Ethics Committee and HTD Ethics Committee.

The operators were 5 nurses and 10 non-expert clinicians without prior echo experience.

6.2.2 Real-time AI-assisted LV segmentation (RAISEG) system

The RAISEG software provides real-time segmentation of the LV A4C view during scanning, to assist operators to acquire appropriate A4C images from standard transthoracic POC echo. Importantly, the RAISEG tool provides a visually acceptable segmentation of the LV endocardial border, only if the image is at the correct A4C view and the quality is good enough. Therefore, operators can use the quality of the real-time AI segmentation as a surrogate measure of image quality, thus assisting them in the acquisition process. Foreshortened or non-visible parts are considered to be nonstandard, and so the segmentation would not move synchronously with the endocardial border and this discrepancy could be easily visualized by the operator. In addition, a real-time LV volume curve is also used as an additional check on whether a good quality standard view has been reached. If the operator is at the standard A4C view and the image is good quality, the curve will follow a typical physiologically realistic waveform which depicts the amount of

blood volume in the LV, with the peak of maximum blood volume at end diastole (EDV) and the trough is minimum blood volume at end systole (ESV).

The developed AI assistant is described in detail in Appendix E and named RAISEG and was developed by the technical team. We deployed a U-Net model for LV segmentation and trained it with the Echonet-Dynamic and CAMUS datasets.

The ultrasound machine was connected with a laptop via a framegrabber (HDMI capture device) is illustrated in Figure 6.2.

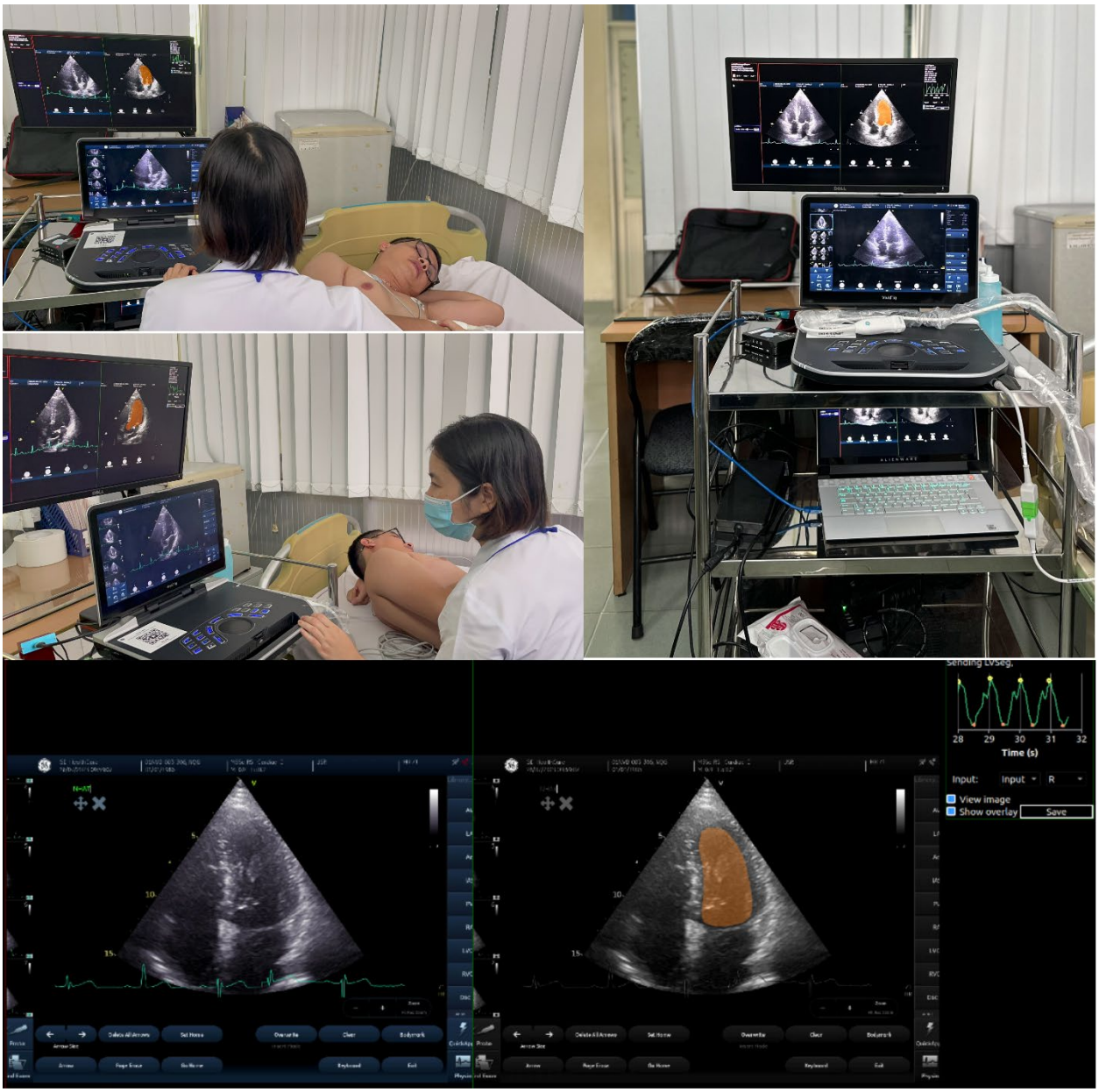


Figure 6.2. The RAISEG framework is operated by a nurse on a healthy volunteer. The segmentation (in orange) is automatically triggered and fits to the LV endocardial border if the operator is in a standard A4C view with good image quality. In addition, the volume curve in green color shows a stable and consistent pattern when in standard view.

6.2.3 Operator training

We provided 1 hour of training for the non-expert operators on how to acquire the A4C view in healthy volunteers. They were first trained theoretically in what is a good A4C standard view and

the 6 standard probe movements (tilt, sweep, rotate, slide, rock and angle) to optimize A4C acquisition. This was followed by a practical session of 30 mins with expert guidance on a volunteer (who was not included in the study participants).

We also provided 30 mins of training to all operators on how to use the RAISEG tool during acquisition.

6.2.4 Study procedures

After written informed consent was obtained, the healthy volunteers were randomly assigned to undergo an echo scan performed by an operator with or without the RAISEG tool. Participants were required to lie in the left lateral decubitus position, with the left arm extended behind the head. This position brings the heart into close contact with the chest wall. The scans were done using a Vivid IQ machine with M5Sc Phased-array transducer (GE Healthcare, London, UK).

The workflow begins with acquisition by the non-expert operators holding the probe and locating the apex. Once they found the apex, they optimized the view by slightly moving the probe (rotating, tilting, rocking, sliding). Once the probe was in the right position and the operator was satisfied with the quality of the image or AI segmentation mask depending on the allocated group, they stored a 5 cardiac cycle video of the A4C view for subsequent LVEF measurement. For the videos acquired by novices, the LVEF metrics were measured using a machine learning algorithm (autoEF, GE Healthcare, London, UK) installed on the ultrasound machine. The resulting LVEF measurements were denoted as $LVEF_{AI}$ and $LVEF_{nonAI}$ for the AI group and the non-AI group respectively.

After that, the experts/radiologists also carried out the echo exam to acquire an A4C video in the same participant within 5 minutes. The expert then manually measured the LVEF ($LVEF_{expert}$) using Simpsons method following the SOP described in Section 3.3.3. To assess whether the autoEF measurements were correlated with the reference measurement, the expert A4C videos were also processed with the autoEF tool resulting in $LVEF_{expert\text{-}autoEF}$.

Image quality assessment of apical 4 chamber view echocardiography video by expert

The image quality of collected videos were assessed by a radiologist based on 4 categories [209] described in detail in **Table 6.1**, which consist of (1) anatomical visibility, (2) anatomical clarity, (3) signal depth-gain, and (4) LV foreshortening. The results were then divided into 3 grades: good (score >36), intermediate (score 25-36), and poor (score < 25). Anatomical visibility represents the visibility of chamber cavities for both A4C and PLAX frames and whether they can be correctly visualized using the correct method of heart's apex slicing, to yield the acceptable clinical projection of images' anatomical structures. This could present a sharp or blurred edge of amplitude structures. Anatomical clarity addresses the degree of distinguishable pixel elements representing the endocardial border cavities or clear distinction between the intraventricular septum, valves, any trabeculated pericardial fluids and endocardial walls. 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's anatomical profile. Apical foreshortening presents as a form of perspective deformation of the LV cavity, especially in the apex region. This deformation occurs as a result of poor image acquisition skills and could effectively alter the chamber's size and renders its volumes geometrically incongruent.

Table 6.1. Image quality assessment criteria including anatomical visibility, anatomical clarity, signal depth-gain, and LV foreshortening.

Image quality assessment criteria	**Maximum mark**
Anatomical visibility	10
Correct axis, Apical segment	6
Interventricular septum visible	2
Interatrial septum visible	2
Anatomical clarity	10
LV cavity clarity, clear edges	4
Distinguishable valves	3
Distinguishable septum wall	3

Signal depth-gain	10
Image sectorial gain	4
No excess gain	3
Minimum artefacts	3
LV foreshorten	10
LV apical segment present	4
Normal-shaped diastole	3
Normal-shaped systole	3
Total quality score	40
Quality Grade	
Good	score > 36
Intermediate	25 <= score <=35
Poor	Score < 25

6.2.5 Statistical analysis

The number of operators able to acquire the A4C standard view, the quality scores and the time taken to carry out the exams were compared between the 2 groups. In addition, the autoEF measurements obtained by non-expert operators were compared to LVEF measurements made manually by an expert using Simpson's rule. Other LV metrics like CO, SV, ESV, EDV were also compared between the two groups.

The distributions of all variables were checked for normality before analysis. The normality test serves an important purpose in statistical analysis, particularly when considering the robustness of certain tests to deviations from normality. In statistical analysis, parametric tests (such as the Student's t-test and ANOVA) assume that the data being analyzed were sampled from a population with a normal distribution. When data deviates significantly from a normal distribution, the results of these tests may become less reliable. Continuous variables were

expressed as mean ± standard deviation or median with interquartile range (IQR) when not normally distributed; categorical variables were presented as counts and/or percentages. Comparison between continuous variables was performed using paired Student's t-test or analysis of variance (ANOVA) with Bonferroni's correction in post-hoc tests, whereas the variables not normally distributed were compared with the non-parametric Kruskal–Wallis test and the Wilcoxon-signed rank test. Categorical variables were compared using the Chi-square test.

Reliability of the measurements was evaluated by intraclass correlations (ICC), where values <0.5 were considered poor, 0.5–0.75 moderate, 0.75–0.9 good and >0.9 excellent [[176]. The reliability of novice LVEF measurement was calculated with a two-way random model defined by absolute agreement in the dataset of average measurements compared with expert manual LVEF measurements. Agreement of LVEF measurements was assessed using Bland–Altman plots [210]. The 95% limits of agreement (LOA) were defined as the range of values between ±1.96 standard deviations from the mean difference. The bias is defined as the average difference between the non-expert using the AI tool with the autoEF measurement and the expert manual LVEF, and the RMSD (root mean square difference) as the square root of the average of the squared differences between the two LVEF measurements. The RMSD is a mathematically robust metric for estimating measurement accuracy.

Pearson's correlation was measured to assess the linear association between novice autoEF measurements and expert reference manual LVEF measurements. The Pearson's correlation coefficients r describes the strength and direction of a linear association between variables, where values <0.3 were considered poor, 0.3–0.6 fair, 0.6–0.7 moderate and >0.8 very strong [211].

For all statistical tests, a two-tailed P value <0.05 was considered statistically significant. Statistical analysis was performed using R 3.0.3.

6.3 Results

Over a 3-week period from July 2023, I enrolled 30 healthy participants at HTD. Overall, 15 (50%) participants were scanned in each group (manual and AI). **Table 6.2** summarises the participants characteristics between the AI and non-AI groups. There are no significant differences between two group in term of age, gender, weight, height, SBP, DBP and HR.

Table 6.2. Demographic characteristics of all participants in the RAISEG study.

Characteristic	non-AI (n=15)	AI (n=15)	p-value
Age	31.0 (29.5, 36.0)	33.0 (28.0, 39.5)	0.6
Sex (female)	7 (47%)	8 (53%)	0.7
Weight (kg)	62 (50, 66)	60 (53, 65)	>0.9
Height (cm)	167 (156, 170)	160 (155, 172)	>0.9
SBP (mmHg)	111 (100, 123)	110 (102, 117)	0.9
DBP (mmHg)	76 (68, 80)	78 (72, 80)	0.5
HR (bpm)	80 (74, 90)	76 (73, 83)	0.4

The results presented in **Table 6.3** outline the median (IQR) image quality for visibility, clarity, depth-gain and foreshortening indicators, total quality score and quality grade respectively. The novices were able to record at least one A4C image in 11/15 (73%) participants in the non-AI group and 12/15 (80%) in the AI group. The total quality score of the AI group is higher than the non-AI group, but this difference was not statistically significant (37 versus 34, p = 0.3). The time to acquire A4C view of the AI group was shorter than the manual group, but this was also not significant (3.0 mins vs AI 3.3 mins, p = 0.5).

Table 6.3. Quality attribute/indicator of A4C acquisitions in the RAISEG study and overall image quality grades.

Image quality	Non-AI (n=15)	AI (n=15)	p-value
A4C acquisition (yes)	11 (73%)	12 (80%)	>0.9

Time to acquire A4C view	3.3 (2.2, 4.1)	3.0 (1.4, 3.5)	0.5
Anatomical visibility	8. (5, 9)	9 (7, 10)	0.2
Anatomical clarity	9 (7, 10)	9 (7, 10)	0.4
Signal depth gain	10 (8, 10)	10 (7, 10)	0.8
LV foreshorten	9 (8, 10)	9 (9, 10)	0.3
Total quality score	34 (27, 37)	37 (32, 39)	0.3
Quality grade			
Good	3 (20%)	5 (33%)	0.8
Intermediate	7 (47%)	7 (47%)	
Poor	4 (33%)	3 (20%)	

Overall, the LV metrics such as HR, EF, CO, SV are very similar between the two groups as illustrated in **Table 6.4**. The difference between novice autoEF versus expert manual measurement is 4 % for both groups (p = 0.8).

Table 6.4. Cardiac metrics from RAISEG study A4C videos including HR, LVEF, CO, SV, EDV, ESV. Results reported for automatic measurement using autoEF and manual measurement by the expert.

Measurement	**Non-AI (n=15)**	**AI (n=15)**	**p-value**
AutoEF of novice acquisition			
HR (bpm)	77 (68, 82)	70 (65, 78)	0.2
LVCO (l/min)	2.65 (2.20, 3.33)	2.60 (2.33, 3.08)	0.8
LVEF (%)	54 (51, 57)	56 (53, 58)	0.4
LVSV (ml)	37 (30, 42)	38 (33, 45)	0.4
LVVED (ml)	68 (57, 77)	77 (63, 83)	0.5
LVVES (ml)	32 (28, 35)	34 (28, 40)	0.7
Manual measurement of expert acquisition by expert			

HR (bpm)	78 (72, 79)	69 (61, 77)	0.1
LVCO (l/min)	3.55 (2.41, 4.67)	3.53 (3.14, 3.84)	0.7
LVEF (%)	59 (55, 62)	58 (57, 63)	0.7
LVSV (ml)	46 (33, 59)	52 (43, 57)	0.7
LVVED (ml)	76 (59, 93)	84 (70, 92)	0.4
LVVES (ml)	29 (26, 31)	30 (25, 39)	0.8
AutoEF of expert acquisition			
HR (bpm)	77 (72, 80)	71 (63, 78)	0.13
LVCO (l/min)	2.80 (2.28, 3.38)	2.70 (2.60, 3.00)	0.9
LVEF (%)	58 (54, 60)	56 (52, 59)	0.4
LVSV (ml)	37 (30, 42)	38 (31, 43)	0.7
LVVED (ml)	68 (56, 74)	72 (56, 82)	0.6
LVVES (ml)	31 (25, 33)	30 (25, 40)	>0.9
Difference in LVEF measurements			
AutoEF novice vs. expert manual	4 (3, 8)	4 (3, 7)	0.8
AutoEF novice vs. AutoEF expert	6 (2, 7)	4 (2, 7)	0.9

Table 6.5. Agreement between RAISEG study autoEF measurements on the non-AI and the AI group of acquisitions by novices with manual LVEF measurements by experts

Reliability	**ICC (95%CI)**	**RMSD (%)**	**r (Pearson)**	**R^2**
$LVEF_{nonAI}$ versus $LVEF_{expert}$	0.419 (<0.001, 0.785)	5.5	0.59	0.35
$LVEF_{AI}$ versus $LVEF_{expert}$	0.615 (<0.001, 0.889)	4.7	0.82	0.67
$LVEF_{expert\text{-}autoEF}$ versus $LVEF_{expert}$	0.660 (0.381, 0.828)	3.7	0.64	0.47
$LVEF_{nonAI}$ vs expert manual EF (same video)	0.362 (<0.001, 0.751)	5.6	0.59	0.15
$LVEF_{AI}$ versus expert manual EF (same video)	0.631 (0.088, 0.876)	4.9	0.66	0.54

As presented in **Table 6.5,** in the AI group, there was a strong correlation with a r of 0.82 ($p<0.001$) and intermediate ICC of 0.615 between the two LVEF measurements ($LVEF_{AI}$ versus $LVEF_{expert}$). Furthermore, the RMSD between the two LVEF measurements was 4.7%.

In the non-AI group, there was a moderate correlation with a r of 0.59 and poor ICC of 0.419 between the two LVEF measurements ($LVEF_{nonAI}$ versus $LVEF_{expert}$). Furthermore, the RMSD between the two LVEF measurements was 5.5%.

As illustrated in **Figure 6.3** the mean difference (bias) between $LVEF_{nonAI}$ and $LVEF_{expert}$ was -4 ± 4% with limits of agreement of − 11 to + 4%. The mean difference (bias) between $LVEF_{AI}$ and $LVE_{Fexpert}$ was -4 ± 3% with limits of agreement of − 9 to + 2%. The $LVEF_{expert\text{-}autoEF}$ versus $LVEF_{expert}$ achieved a mean difference of -1 ± 3 % and limits of agreement of -7 to +5 %.

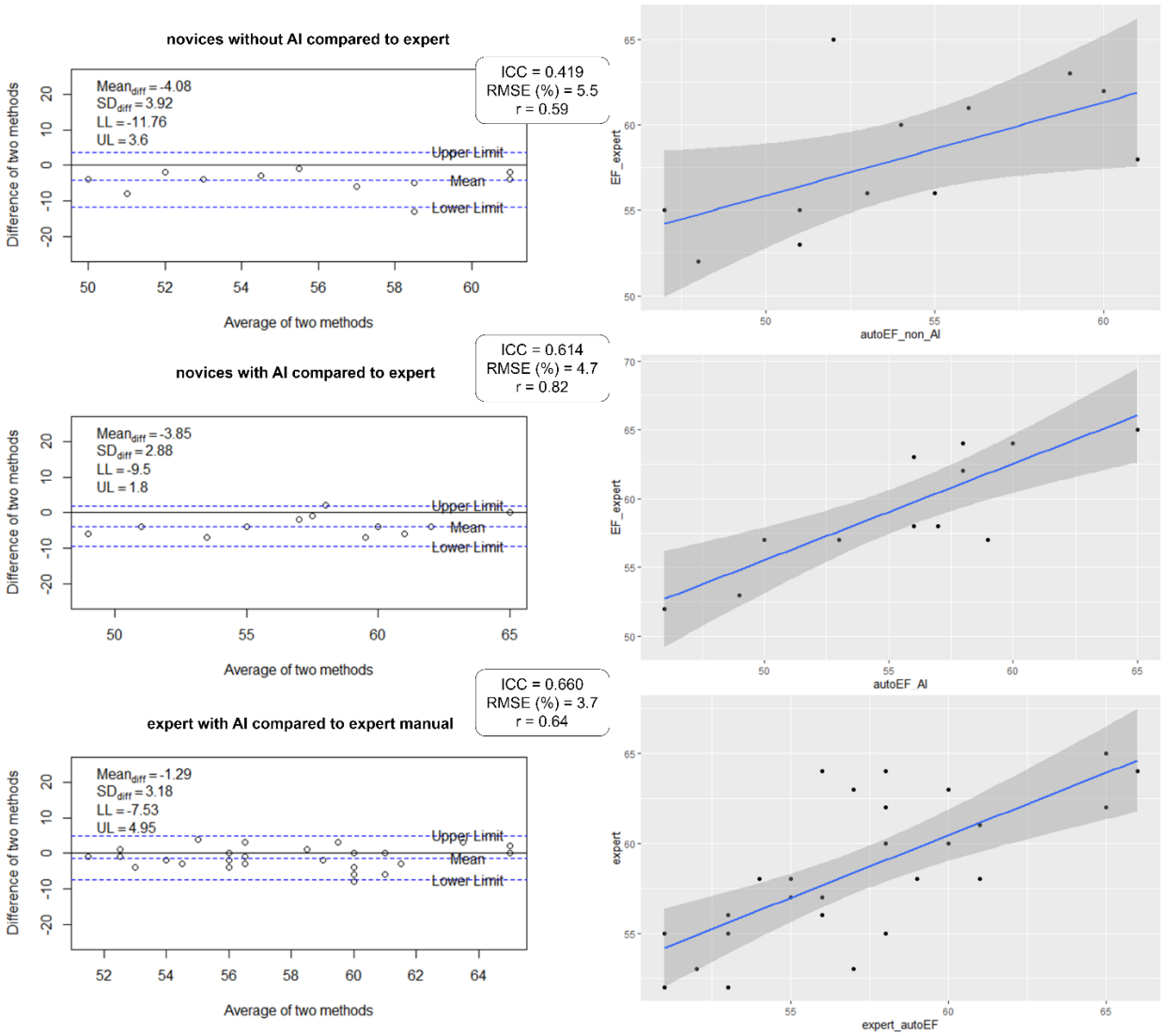


Figure 6.3. Correlation and Bland and Altman comparison between manual LVEF measurements by experts and autoEF measurements by non-expert operators. (a) $LVEF_{nonAI}$ and $LVEF_{expert}$, (b) $LVEF_{AI}$ and $LVEF_{expert}$, (c) $LVEF_{expert\text{-}autoEF}$ versus $LVEF_{expert}$

6.4 Discussion

An increasing number of non-expert clinicians and intensivists have been trained to perform POC echo assessments. However, acquisition of images of diagnostic quality and subsequent quantitative evaluations remains challenging for many, particularly for novices. In the present study, we tested an AI-enabled A4C segmentation tool specifically designed to facilitate high quality image acquisition at the bedside. We demonstrated the feasibility of such a system of

real-time segmentation and provided wider implementation details and how it can be clinically deployed in a unified workflow. Apart from validating an AI based segmentation tool on healthy volunteers, we have also presented a proof of concept for using this tool to assist in operator guidance to acquire high quality A4C images. As shown in the results, when supported with an AI-segmentation tool, the operator could acquire the A4C view with a higher image quality score in a lower time, and subsequent LVEF measurements were better correlated with reference LVEF measurements by expert clinicians. However, these differences were not statistically significant in all the image quality score components as well as time. We speculate that this could be a Type II error and be a consequence of the relatively small sample size used in this preliminary study. Nevertheless, the results are intriguing and provide motivation for future larger-scale studies of this type.

Limitations

This is a single-center study on healthy volunteers with a relatively small sample size. As stated above, the benefits of using the RAISEG tool in ICU patients needs further investigation. Currently, our tool is only able to work well in high quality images, thus the implementation in ICU patients may be limited because it is very challenging to acquire high-quality images in those patients. When compared to other available tools, the autoEF tool used in this study demonstrated a lower correlation coefficient r with expert manual measurement than previous studies (0.82 vs 0.92 in [31] and 0.96 in [17]). Other limitations related to our RAISEG system compared to other systems include firstly the lack of real-time LVEF estimation, secondly, the lack of real-time automated segmentation quality feedback and thirdly, the lack of guidance for acquisition. All the acquisitions and decisions on whether the images were of good quality were made by the operators.

6.5 Conclusions

AI-enabled real-time LVEF segmentation showed the feasibility to assist non-expert operators in capturing higher image quality data for subsequent metric estimation. In addition, there were indications that the reliability and correlation of autoEF measurements from novices' acquisitions were better with the help of the AI-assisted segmentation tool. However, the function of the tool

needs further technical development, and its applicability should be investigated with further clinical validation in different disease states and clinical settings.

7 Conclusions and future directions

Chapter overview

This thesis has presented details of the design, implementation and results of clinical studies to evaluate the clinical benefit of AI-assisted US tools, particularly in the ICU in LMIC settings. This type of research has been relatively neglected in the literature, with a large number of papers proposing and evaluating (offline) new techniques for US image analysis, but very few analysing their impact on clinical workflows. It has been shown that, with appropriate experimental design and data collection, technical development approaches and user training, AI can assist non-expert operators to carry out POCUS lung and muscle examinations in critically ill patients. We also demonstrated a proof-of-concept of AI-assisted echocardiography for acquiring standard A4C views, with promising initial results.

This chapter summarises the main contributions and clinical impact of the work presented in this thesis in Section 7.1. Section 7.2 discusses the limitations of this work and explores opportunities for future research directions.

7.1 Overview of contributions and clinical impact

This thesis has covered the areas of lung, muscle and cardiac US. Although experts have been discussing the use of AI tools in clinical medicine for many years [212–216], real-life examples such as RAILUS, RAIMUS and RAISEG proposed in this thesis provide evidence of concrete scenarios as to how such tools can actually be implemented and embedded in a clinician's workflow. Below we summarise the contributions of each of these three tools, highlighting the validation and implementation of the AI-assisted US imaging tools in the ICU, and specifically in a resource-limited setting.

Clinical benefit of AI-assisted lung ultrasound in resource-limited ICU

We developed and evaluated a real-time AI-assisted LUS system in a resource-limited ICU setting [131]. Our RAILUS software is an open-source tool designed to be easy–to–use, and unlike most

commercial systems, our tool is run through a laptop computer and can be linked to most ultrasound devices. With the backbone of a DL model, RAILUS can assist non expert clinicians in interpreting 5 lung ultrasound features commonly seen in ICU patients. In low resource settings like Vietnam, there are very few experts in LUS and there is a significant gap between clinicians with different expertise in LUS interpretation. This performance gap and the challenging nature of this task may prevent non-expert clinicians from carrying out LUS examinations in clinical practice. Evaluating the impact of an AI tool that can assist inexperienced users will be highly beneficial for improving patient outcomes and quality of care. Our clinical evaluation study has shown that the RAILUS tool assisted non-expert clinicians in improving the accuracy, speed and confidence of bedside LUS, and will therefore help clinicians in ICUs in LMICs to better manage critically ill patients with various lung pathologies [131].

Clinical evaluation of AI-assisted muscle ultrasound for monitoring muscle wasting in ICU patients

We carried out a prospective study to evaluate the impact of an AI tool for real-time RFCSA estimation compared to the traditional manual measurement technique for monitoring muscle mass in the ICU. Building on a UNET model, our RAIMUS tool succeeded in supporting operators in measurement of RFCSA with less variability than the conventional non-AI technique. Furthermore, the time spent on measuring RFCSA using the AI tool was approximately half that of standard measurements. We showed that the scan-rescan variability, intraobserver and interobserver delineation variability were similar which indicates that the main variability of the standard technique is manual delineation [217]. This suggests that using the AI tool may enable a 1-scan measurement instead of the current clinical standard of taking the average of 3-scan measurements, further reducing time. Such a system would significantly assist routine clinical monitoring of muscle changes in ICU patients and help in assessing the effectiveness of interventions such as rehabilitation or nutrition.

Proof-of-concept of AI-assisted cardiac apical four chambers view acquisition in healthy volunteers

We developed and conducted a preliminary pilot study on healthy volunteers of an AI-assisted LV segmentation tool for guiding the operator to a standard view. The UNET model was trained

with two publicly available datasets and deployed with real-time software - RAISEG. We demonstrated that when supported with RAISEG, the operator could acquire the A4C view with a slightly higher image quality score in a lower time, and subsequent EF measurements were better correlated with the reference EF measured by expert clinicians. Therefore, the use of the RAISEG tool can help the operator to acquire images that could be processed better on third party software, suggesting its usefulness as a quality control tool. However, the results were not statistically significant, which suggests that further research and a larger amount of data will be required to validate these intriguing preliminary findings.

7.2 Limitations, remaining challenges, and future work

An increasing number of non-expert clinicians and intensivists have been trained to perform POCUS examinations. However, the acquisition of diagnostic image quality and quantitative evaluations remain challenging for many, particularly for novices.

We believe that the results from this thesis represent an important step towards real-time implementation of AI in LMIC ICUs, but nevertheless some limitations remain. The lung and muscle AI systems were built using datasets from patients in a single hospital so it remains to be seen how the tools would generalize to patients in other settings. Furthermore, while we have designed our system to be agnostic to ultrasound devices, clinical validation of the tool on other types of devices (such as portable devices) is yet to be performed. Although the performance of our AI tools has been shown to be good, future work should more thoroughly assess their robustness and bias, especially in the presence of poor image quality. Further research should also focus on validation of the AI tools in different patient populations, using different machine manufacturers and different settings, for example, monitoring muscle wasting in patients with cancer. Our study focused on 3 clinical applications, namely lung, muscle and cardiac. Clinical practice requires a more nuanced interpretation of variety of sources of clinical information for optimal benefit, for example, whether the serial LUS or echocardiography results are in concordance with clinical guideline criteria to discriminate between cardiogenic or pneumonia pulmonary edema. Future work could investigate multimodal AI techniques to incorporate such information. Moreover, in the case of automatic measurements such as muscle size and ejection

fraction, it is important to emphasize that removing the variability completely by the use of a deterministic AI model makes repeated measurements more reproducible but does not necessarily make them more accurate.

Furthermore, the translation of clinical AI tools in POCUS is unlikely to take place without addressing governance requirements. These challenges with AI applications are not necessarily unique to POCUS and include the need for data quality control, external validation, cultural barriers to technology adoption, regulatory approvals, health economic assessment, data security and the demonstration of superior patient outcomes with the implementation of AI-enabled tools. Although there are barriers to deploying AI-enabled US at scale in LMICs, strategy of promoting local innovation and initiative can accelerate the efficiently sustainable AI-enabled US implementation.

In this thesis, I have focused on application of AI to US imaging in the ICU, demonstraing on the promise, clinical capabilities, research opportunities, gaps and risks of the application of AI to the ultrasound imaging for the diagnosis and management of critically ill patients. Nevertheless, several barriers to the deployment of AI-assisted tools in clinical settings arise during or following the creation, implementation, and maintenance of digital health technologies [12]. During the design phase of any AI tool, it is essential to identify, quantify and analyze its risks and opportunities, enhancing the likelihood of obtaining favorable outcomes and optimizing the chances of success. However, these aspects regarding implementation are not investigated in this thesis. The efficient implementation of AI technologies, characterized by proper implementation of a systematic management approach, including strategic planning, resource allocation, and control and evaluation processes, is fundamental to refining healthcare services, equipment, and technologies. The greatest challenge to AI in these healthcare domains is not whether the technologies will be capable enough to be useful, but rather ensuring their adoption in daily clinical practice. For widespread adoption to take place, AI systems must be approved by regulators, accompanied by standardised processes and AI medical training to clinicians, all of which should be updated over time.

Another significant issue is that the application and implementation of AI in healthcare requires a decision and commitment by governments and a recognition that successful AI applications require an integrated strategy, promotion of the appropriate use of AI interventions for health, and recognition of the urgent need to address the major impediments faced by LMICs implementing AI technologies in healthcare. To protect the privacy of patients and regional communities, health systems are often heavily regulated. Such regulations, especially those dealing with patient data privacy and data sharing, may slow digital technology adoption. Since digital equipment may be relatively easy to operate, there is a risk of unauthorised and unskilled usage, especially where human resources are scarce. Thoughtful modifications to the regulatory environment and AI education for healthcare professionals may be necessary in certain areas to allow the benefits of AI-assisted US to be realised. The medical community has an ethical obligation to expand the quality and scope of AI technologies that have the potential to improve health and to implement such innovations equitably. Access to technology is already inequitably distributed, so we need to ensure that new innovations do not widen existing health inequalities.

7.3 Summary

AI in US is an exciting are of medical innovation and has the potential to optimize existing resources and help overcome workforce shortages, by assisting in the operation of the US system, measurement of biometric parameters from images, diagnostics as well as assessing the severity of disease, and providing other insights that can help patient management. AI has the potential to greatly improve US practice and ultimately patient outcomes in resource-limited settings. Although there are barriers to deploying AI-enabled US at scale in LMICs, a strategy of promoting local innovation and initiatives can accelerate the sustainable implementation of AI-enabled US. Key challenges for the translation of AI-enabled US in routine clinical practice include those intrinsic to the science of AI, logistical difficulties in implementation, and consideration of the barriers to adoption as well as of the necessary sociocultural or pathway changes. Robust peer-reviewed clinical evaluation as part of randomized controlled trials should be viewed as the gold standard for evidence generation, but conducting these in practice may not always be feasible or appropriate. As a result, very few randomized controlled trials or prospective studies

have shown the safety and effectiveness of existing AI algorithms in US, and the lack of real-world evaluation of AI systems can pose a substantial risk to patients and clinicians. Performance metrics should aim to capture real clinical applicability and be understandable to intended users. Regulation that balances the pace of innovation with the potential for harm, alongside thoughtful post-market surveillance, is required to ensure that patients are not exposed to dangerous interventions nor deprived of access to beneficial innovations. Mechanisms to enable direct comparisons of AI systems must be developed, including the use of independent, local and representative test sets. Developers of AI algorithms must be vigilant to potential dangers, including dataset shift, accidental fitting of confounders, unintended discriminatory bias, the challenges of generalisation to new populations, and the unintended negative consequences of new algorithms on health outcomes. At present, most data for current AI development are sourced from HICs, with some contributions from high-resource institutions in LMICs. This imbalanced data sourcing limits AI generalizability because of differing demographic characteristics, diseases, and equipment.

Successful AI adoption in LMICs also requires education of the local healthcare professional leadership in AI validation. Training for local IT personnel and healthcare workers is also critical for sustainable infrastructure and AI utilization.

Though still in their preliminary stages, tools like RAILUS, RAIMUS and RAISEG have great potential to change the way people approach POCUS in the ICU. These are just a few examples of how AI-assisted US is changing the research direction of clinical research. Although a clinician has the unique ability to take into account other considerations such as patient context, a longitudinal view of the patient, and social factors, AI tools like these may still be of value in being able to diagnose and monitor, or act as clinical decision-making support systems. Indeed, as US imaging has become such an integral part of the practice of critical care medicine, this field of innovation can truly impact the world of healthcare delivery.

Bibliography

1. Volpicelli G, Mayo P, Rovida S. Focus on ultrasound in intensive care. Intensive Care Med. 2020;46:1258–60.

2. Brogi E, Bignami E, Sidoti A, Shawar M, Gargani L, Vetrugno L, et al. Could the use of bedside lung ultrasound reduce the number of chest x-rays in the intensive care unit? Cardiovasc Ultrasound. 2017;15.

3. Sistrom CL, McKay NL. Costs, charges, and revenues for hospital diagnostic imaging procedures: Differences by modality and hospital characteristics. Journal of the American College of Radiology. 2005;2:511–9.

4. Lichtenstein DA. Lung ultrasound in the critically ill. 2014.

5. Pivetta E, Goffi A, Nazerian P, Castagno D, Tozzetti C, Tizzani P, et al. Lung ultrasound integrated with clinical assessment for the diagnosis of acute decompensated heart failure in the emergency department: a randomized controlled trial. Eur J Heart Fail. 2019;21:754–66.

6. Leopold SJ, Ghose A, Plewes KA, Mazumder S, Pisani L, Kingston HWF, et al. Point-of-care lung ultrasound for the detection of pulmonary manifestations of malaria and sepsis: An observational study. PLoS One. 2018;13.

7. Buonsenso D, De Rose C. Implementation of lung ultrasound in low-to middle-income countries: a new challenge global health? Eur J Pediatr. 2021;:1–8.

8. Mitchell C, Rahko PS, Blauwet LA, Canaday B, Finstuen JA, Foster MC, et al. Guidelines for Performing a Comprehensive Transthoracic Echocardiographic Examination in Adults: Recommendations from the American Society of Echocardiography. Journal of the American Society of Echocardiography. 2019;32:1–64.

9. Field LC, Guldan GJ, Finley AC. Echocardiography in the intensive care unit. Semin Cardiothorac Vasc Anesth. 2011;15:25–39.

10. Thavendiranathan P, Popović ZB, Flamm SD, Dahiya A, Grimm RA, Marwick TH. Improved interobserver variability and accuracy of echocardiographic visual left ventricular ejection fraction assessment through a self-directed learning program using cardiac magnetic resonance images. Journal of the American Society of Echocardiography. 2013;26:1267–73.

11. Knackstedt C, Bekkers SCAM, Schummers G, Schreckenberg M, Muraru D, Badano LP, et al. Fully Automated Versus Standard Tracking of Left Ventricular Ejection Fraction and Longitudinal Strain the FAST-EFs Multicenter Study. J Am Coll Cardiol. 2015;66:1456–66.

12. Nhat PTH, Hao N Van, Pisani L, Kerdegari H, Thuy DB, Thu LNM, et al. Role of AI-enabled Ultrasound Imaging in a Resource Limited Intensive Care Unit. Strasbourg: MICCAI-Affordable Healthcare and AI for Resource Diverse Global Health (FAIR) ; 2021.

13. Turner HC, Hao N Van, Yacoub S, Hoang VMT, Clifton DA, Thwaites GE, et al. Achieving affordable critical care in low-income and middle-income countries. BMJ Glob Health. 2019;4:e001675.

14. Thuy DB, Campbell J, Nhat LTH, Hoang NVM, Hao N Van, Baker S, et al. Hospital-acquired colonization and infections in a Vietnamese intensive care unit. PLoS One. 2018;13:e0203600-.

15. Riviello ED, Buregeya E, Twagirumugabe T. Diagnosing acute respiratory distress syndrome in resource limited settings: The Kigali modification of the Berlin definition. Current Opinion in Critical Care. 2017;23.

16. Rajamani A, Knudsen S, Ngoc Bich Ha Huynh K, Huang S, Wong W-T, Ting I, et al. Basic echocardiography competence program in intensive care units: A multinational survey of intensive care units accredited by the College of Intensive Care Medicine. Anaesth Intensive Care. 2020;48:150–4.

17. Fan E, Cheek F, Chlan L, Gosselink R, Hart N, Herridge MS, et al. An official American Thoracic Society Clinical Practice guideline: the diagnosis of intensive care unit–acquired weakness in adults. Am J Respir Crit Care Med. 2014;190:1437–46.

18. Schefold JC, Bierbrauer J, Weber-Carstens S. Intensive care unit—acquired weakness (ICUAW) and muscle wasting in critically ill patients with severe sepsis and septic shock. J Cachexia Sarcopenia Muscle. 2010;1:147–57.

19. Wong A, Galarza L, Forni L, De Backer D, Slama M, Cholley B, et al. Recommendations for core critical care ultrasound competencies as a part of specialist training in multidisciplinary intensive care: a framework proposed by the European Society of Intensive Care Medicine (ESICM). Crit Care. 2020;24:1–6.

20. Boldt J. Clinical review: Hemodynamic monitoring in the intensive care unit. Crit Care. 2002;6:1–8.

21. Laursen CB, Sloth E, Lassen AT, Christensen R dePont, Lambrechtsen J, Madsen PH, et al. Point-of-care ultrasonography in patients admitted with respiratory symptoms: a single-blind, randomised controlled trial. Lancet Respir Med. 2014;2:638–46.

22. Blanco P. A traditional paradigm vs. an ultrasound-supported paradigm in emergency and critical care medicine: a crisis of the mind is needed. Journal of Emergency Medicine. 2015;49:e63–4.

23. Al Deeb M, Barbic S, Featherstone R, Dankoff J, Barbic D. Point-of-care ultrasonography for the diagnosis of acute cardiogenic pulmonary edema in patients presenting with acute dyspnea: a systematic review and meta-analysis. Academic Emergency Medicine. 2014;21:843–52.

24. Baker DE, Nolting L, Brown HA. Impact of point-of-care ultrasound on the diagnosis and treatment of patients in rural Uganda. Trop Doct. 2021;51:291–6.

25. Zieleskiewicz L, Lopez A, Hraiech S, Baumstarck K, Pastene B, Di Bisceglie M, et al. Bedside POCUS during ward emergencies is associated with improved diagnosis and outcome: an observational, prospective, controlled study. Crit Care. 2021;25:1–12.

26. Volpicelli G, Lamorte A, Tullio M, Cardinale L, Giraudo M, Stefanone V, et al. Point-of-care multiorgan ultrasonography for the evaluation of undifferentiated hypotension in the emergency department. Intensive Care Med. 2013;39:1290–8.

27. Expert Round Table on Echocardiography in ICU pmayo@ nshs. edu antoine. vieillard-baron@ apr. aphp. fr. International consensus statement on training standards for advanced critical care echocardiography. Intensive Care Med. 2014;40:654–66.

28. Burk RE, Beesley SJ, Grissom CK, Hirshberg EL, Lanspa MJ, Brown SM. Echocardiography in the Intensive Care Unit. Current Cardiovascular Imaging Reports. 2017;10.

29. Shrestha GS, Weeratunga D, Baker K. Point-of-Care Lung Ultrasound in Critically ill Patients. Rev Recent Clin Trials. 2018;13.

30. Messina A, Robba C, Bertuetti R, Biasucci D, Corradi F, Mojoli F, et al. Head to toe ultrasound: a narrative review of experts' recommendations of methodological approaches. Journal of Anesthesia, Analgesia and Critical Care. 2022;2:1–17.

31. Perera P, Mailhot T, Riley D, Mandavia D. The RUSH exam: Rapid Ultrasound in SHock in the evaluation of the critically lll. Emergency Medicine Clinics. 2010;28:29–56.

32. Lichtenstein DA. BLUE-Protocol and FALLS-Protocol: Two applications of lung ultrasound in the critically ill. Chest. 2015;147.

33. Lazzeri C, Peris A. The Kigali modification of the berlin definition: A new epidemiological tool for ARDS? J Thorac Dis. 2016;8:E443–5.

34. Whitson MR, Mayo PH. Ultrasonography in the emergency department. Crit Care. 2016;20:1–8.

35. Alherbish A, Priestap F, Arntfield R. The introduction of basic critical care echocardiography reduces the use of diagnostic echocardiography in the intensive care unit. J Crit Care. 2015;30:1419-e7.

36. Vetrugno L, Baciarello M, Bignami E, Bonetti A, Saturno F, Orso D, et al. The "pandemic" increase in lung ultrasound use in response to Covid-19: can we complement computed tomography findings? A narrative review. Ultrasound J. 2020;12:1–11.

37. Tenajas R, Miraut D, Illana CI, Alonso-Gonzalez R, Arias-Valcayo F, Herraiz JL. Recent Advances in Artificial Intelligence-Assisted Ultrasound Scanning. Applied Sciences. 2023;13:3693.

38. Ferraz S, Coimbra M, Pedrosa J. Assisted probe guidance in cardiac ultrasound: A review. Front Cardiovasc Med. 2023;10:1056055.

39. Smistad E, Johansen KF, Iversen DH, Reinertsen I. Highlighting nerves and blood vessels for ultrasound-guided axillary nerve block procedures using neural networks. Journal of Medical Imaging. 2018;5:44004.

40. Cai Y, Droste R, Sharma H, Chatelain P, Drukker L, Papageorghiou AT, et al. Spatio-temporal visual attention modelling of standard biometry plane-finding navigation. Med Image Anal. 2020;65:101762.

41. Cai Y, Sharma H, Chatelain P, Noble JA. SonoEyeNet: Standardized fetal ultrasound plane detection informed by eye tracking. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE; 2018. p. 1475–8.

42. Østvik A, Smistad E, Aase SA, Haugen BO, Lovstakken L. Real-Time Standard View Classification in Transthoracic Echocardiography Using Convolutional Neural Networks. Ultrasound Med Biol. 2019;45.

43. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. NPJ Digit Med. 2018;1.

44. Baumgartner CF, Kamnitsas K, Matthew J, Fletcher TP, Smith S, Koch LM, et al. SonoNet: Real-Time Detection and Localisation of Fetal Standard Scan Planes in Freehand Ultrasound. IEEE Trans Med Imaging. 2017;36.

45. Raina D, Ntentia D, Chandrashekhara SH, Voyles R, Saha SK. Expert-Agnostic Ultrasound Image Quality Assessment using Deep Variational Clustering. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE; 2023. p. 2717–23.

46. Abdi AH, Luong C, Tsang T, Allan G, Nouranian S, Jue J, et al. Automatic quality assessment of echocardiograms using convolutional neural networks: feasibility on the apical four-chamber view. IEEE Trans Med Imaging. 2017;36:1221–30.

47. Wu L, Cheng J-Z, Li S, Lei B, Wang T, Ni D. FUIQA: fetal ultrasound image quality assessment with deep convolutional networks. IEEE Trans Cybern. 2017;47:1336–49.

48. Luong C, Abdi A, Jue J, Gin K, Fleming S, Abolmaesumi P, et al. Automatic quality assessment of echo apical 4-chamber images using computer deep learning. Circulation. 2016;134 suppl_1:A17562–A17562.

49. Wang Z, Song Y, Zhao B, Zhong Z, Yao L, Lv F, et al. A Soft-Reference Breast Ultrasound Image Quality Assessment Method That Considers the Local Lesion Area. Bioengineering. 2023;10:940.

50. Perdios D, Vonlanthen M, Besson A, Martinez F, Arditi M, Thiran J-P. Deep convolutional neural network for ultrasound image enhancement. In: 2018 IEEE International Ultrasonics Symposium (IUS). IEEE; 2018. p. 1–4.

51. Abdel-Nasser M, Omer OA. Ultrasound image enhancement using a deep learning architecture. In: Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016 2. Springer; 2017. p. 639–49.

52. Akkus Z, Cai J, Boonrod A, Zeinoddini A, Weston AD, Philbrick KA, et al. A survey of deep-learning applications in ultrasound: Artificial intelligence–powered ultrasound for improving clinical workflow. Journal of the American College of Radiology. 2019;16:1318–28.

53. Han S, Kang H-K, Jeong J-Y, Park M-H, Kim W, Bang W-C, et al. A deep learning framework for supporting the classification of breast lesions in ultrasound images. Phys Med Biol. 2017;62:7714.

54. Byra M, Galperin M, Ojeda-Fournier H, Olson L, O'Boyle M, Comstock C, et al. Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion. Med Phys. 2019;46:746–55.

55. Ko SY, Lee JH, Yoon JH, Na H, Hong E, Han K, et al. Deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. Head Neck. 2019;41:885–91.

56. Diaz-Escobar J, Ordóñez-Guillén NE, Villarreal-Reyes S, Galaviz-Mosqueda A, Kober V, Rivera-Rodriguez R, et al. Deep-learning based detection of COVID-19 using lung ultrasound imagery. PLoS One. 2021;16.

57. Roy S, Menapace W, Oei S, Luijten B, Fini E, Saltori C, et al. Deep Learning for Classification and Localization of COVID-19 Markers in Point-of-Care Lung Ultrasound. IEEE Trans Med Imaging. 2020;39.

58. Arntfield R, Vanberlo B, Alaifan T, Phelps N, White M, Chaudhary R, et al. Development of a convolutional neural network to differentiate among the etiology of similar appearing pathological b lines on lung ultrasound: A deep learning study. BMJ Open. 2021;11.

59. Ahn JC, Qureshi TA, Singal AG, Li D, Yang J-D. Deep learning in hepatocellular carcinoma: Current status and future perspectives. World J Hepatol. 2021;13:2039.

60. Brehar R, Mitrea D-A, Vancea F, Marita T, Nedevschi S, Lupsor-Platon M, et al. Comparison of deep-learning and conventional machine-learning methods for the automatic recognition of the hepatocellular carcinoma areas from ultrasound images. Sensors. 2020;20:3085.

61. Yap MH, Pons G, Marti J, Ganau S, Sentis M, Zwiggelaar R, et al. Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J Biomed Health Inform. 2017;22:1218–26.

62. Azizi S, Bayat S, Yan P, Tahmasebi A, Nir G, Kwak JT, et al. Detection and grading of prostate cancer using temporal enhanced ultrasound: combining deep neural networks and tissue mimicking simulations. Int J Comput Assist Radiol Surg. 2017;12:1293–305.

63. Liu Z, Yang C, Huang J, Liu S, Zhuo Y, Lu X. Deep learning framework based on integration of S-Mask R-CNN and Inception-v3 for ultrasound image-aided diagnosis of prostate cancer. Future Generation Computer Systems. 2021;114:358–67.

64. Cannesson M, Tanabe M, Suffoletto MS, McNamara DM, Madan S, Lacomis JM, et al. A Novel Two-Dimensional Echocardiographic Image Analysis System Using Artificial Intelligence-Learned Pattern Recognition for Rapid Automated Ejection Fraction. J Am Coll Cardiol. 2007;49.

65. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Medical Image Computing

and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer International Publishing; 2015. p. 234–41.

66. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18:203–11.

67. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, et al. Video-based AI for beat-to-beat assessment of cardiac function. Nature. 2020;580.

68. Dangoury S, Sadik M, Alali A, Fail A. V-net performances for 2D ultrasound image segmentation. In: 2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA). IEEE; 2022. p. 96–100.

69. Ashkani Chenarlogh V, Ghelich Oghli M, Shabanzadeh A, Sirjani N, Akhavan A, Shiri I, et al. Fast and accurate U-net model for fetal ultrasound image segmentation. Ultrason Imaging. 2022;44:25–38.

70. Malhotra P, Gupta S, Koundal D, Zaguia A, Enbeyle W. Deep neural networks for medical image segmentation. J Healthc Eng. 2022;2022.

71. Wang R, Lei T, Cui R, Zhang B, Meng H, Nandi AK. Medical image segmentation using deep learning: A survey. IET Image Process. 2022;16:1243–67.

72. Ritsche P, Wirth P, Cronin NJ, Sarto F, Narici M V, Faude O, et al. DeepACSA: Automatic Segmentation of Cross-sectional Area in Ultrasound Images of Lower Limb Muscles Using Deep Learning. Med Sci Sports Exerc. 2022.

73. Akerman AP, Porumb M, Scott CG, Beqiri A, Chartsias A, Ryu AJ, et al. Automated echocardiographic detection of heart failure with preserved ejection fraction using artificial intelligence. JACC: Advances. 2023;:100452.

74. Gu Y, Xu W, Lin B, An X, Tian J, Ran H, et al. Deep learning based on ultrasound images assists breast lesion diagnosis in China: a multicenter diagnostic study. Insights Imaging. 2022;13:124.

75. Motazedian P, Marbach JA, Prosperi-Porta G, Parlow S, Di Santo P, Abdel-Razek O, et al. Diagnostic accuracy of point-of-care ultrasound with artificial intelligence-assisted assessment of left ventricular ejection fraction. NPJ Digit Med. 2023;6:201.

76. Gaube S, Suresh H, Raue M, Lermer E, Koch TK, Hudecek MFC, et al. Non-task expert physicians benefit from correct explainable AI advice when reviewing X-rays. Sci Rep. 2023;13:1383.

77. Bernstein MH, Atalay MK, Dibble EH, Maxwell AWP, Karam AR, Agarwal S, et al. Can incorrect artificial intelligence (AI) results impact radiologists, and if so, what can we do about it? A multi-reader pilot study of lung cancer detection with chest radiography. Eur Radiol. 2023;:1–7.

78. Zeng X, Wen L, Xu Y, Ji C. Generating diagnostic report for medical image by high-middle-level visual information incorporation on double deep learning models. Comput Methods Programs Biomed. 2020;197:105700.

79. Yang S, Niu J, Wu J, Wang Y, Liu X, Li Q. Automatic ultrasound image report generation with adaptive multimodal attention mechanism. Neurocomputing. 2021;427:40–9.

80. Zeng X, Wen L, Liu B, Qi X. Deep learning for ultrasound image caption generation based on object detection. Neurocomputing. 2020;392:132–41.

81. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, et al. Fully Automated Echocardiogram Interpretation in Clinical Practice. Circulation. 2018;138:1623–35.

82. Christensen M, Vukadinovic M, Yuan N, Ouyang D. Multimodal Foundation Models For Echocardiogram Interpretation. arXiv preprint arXiv:230815670. 2023.

83. Lekadir K, Feragen A, Fofanah AJ, Frangi AF, Buyx A, Emelie A, et al. FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare. arXiv preprint arXiv:230912325. 2023.

84. de Hond AAH, Leeuwenberg AM, Hooft L, Kant IMJ, Nijman SWJ, van Os HJA, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. NPJ Digit Med. 2022;5:2.

85. Szabo L, Raisi-Estabragh Z, Salih A, McCracken C, Ruiz Pujadas E, Gkontra P, et al. Clinician's guide to trustworthy and responsible artificial intelligence in cardiovascular imaging. Front Cardiovasc Med. 2022;9:1016032.

86. Dutta A, Zisserman A. The VIA Annotation Software for Images, Audio and Video. In: Proceedings of the 27th ACM International Conference on Multimedia. New York, NY, USA: Association for Computing Machinery; 2019. p. 2276–9.

87. Yushkevich PA, Gao Y, Gerig G. ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2016. p. 3342–5.

88. Kikinis R, Pieper SD, Vosburgh KG. 3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support. In: Jolesz FA, editor. Intraoperative Imaging and Image-Guided Therapy. New York, NY: Springer New York; 2014. p. 277–89.

89. Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, et al. Preparing Medical Imaging Data for Machine Learning. Radiology. 2020;295:4–15.

90. Candemir S, Nguyen X V, Folio LR, Prevedello LM. Training Strategies for Radiology Deep Learning Models in Data-limited Scenarios. Radiol Artif Intell. 2021;3:e210014.

91. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014.

92. Ruder S. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:160904747. 2016.

93. Yin J, Ngiam KY, Teo HH. Role of Artificial Intelligence Applications in Real-Life Clinical Practice: Systematic Review. J Med Internet Res. 2021;23:e25759.

94. Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nat Mach Intell. 2021;3:199–217.

95. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. J Med Internet Res. 2021;23:e25759.

96. Kuroda Y, Kaneko T, Yoshikawa H, Uchiyama S, Nagata Y, Matsushita Y, et al. Artificial intelligence-based point-of-care lung ultrasound for screening COVID-19 pneumoniae: Comparison with CT scans. PLoS One. 2023;18:e0281127-.

97. Dave C, Wu D, Tschirhart J, Smith D, VanBerlo B, Deglint J, et al. Prospective Real-Time Validation of a Lung Ultrasound Deep Learning Model in the ICU. Crit Care Med. 2023;51.

98. Pluym ID, Afshar Y, Holliman K, Kwan L, Bolagani A, Mok T, et al. Accuracy of automated three-dimensional ultrasound imaging technique for fetal head biometry. Ultrasound in Obstetrics & Gynecology. 2021;57:798–803.

99. Matthew J, Skelton E, Day TG, Zimmer VA, Gomez A, Wheeler G, et al. Exploring a new paradigm for the fetal anomaly ultrasound scan: Artificial intelligence in real time. Prenat Diagn. 2022;42:49–59.

100. Tiyarattanachai T, Apiparakoon T, Chaichuen O, Sukcharoen S, Yimsawad S, Jangsirikul S, et al. Artificial intelligence assists operators in real-time detection of focal liver lesions during ultrasound: A randomized controlled study. Eur J Radiol. 2023;165:110932.

101. Papadopoulou S-L, Sachpekidis V, Kantartzi V, Styliadis I, Nihoyannopoulos P. Clinical validation of an artificial intelligence-assisted algorithm for automated quantification of left ventricular ejection fraction in real time by a novel handheld ultrasound device. European Heart Journal - Digital Health. 2022;3:29–37.

102. Gohar E, Herling A, Mazuz M, Tsaban G, Gat T, Kobal S, et al. Artificial Intelligence (AI) versus POCUS Expert: A Validation Study of Three Automatic AI-Based, Real-Time, Hemodynamic Echocardiographic Assessment Tools. J Clin Med. 2023;12.

103. Zhai S, Wang H, Sun L, Zhang B, Huo F, Qiu S, et al. Artificial intelligence (AI) versus expert: A comparison of left ventricular outflow tract velocity time integral (LVOT-VTI) assessment between ICU doctors and an AI tool. J Appl Clin Med Phys. 2022;23:e13724.

104. Varudo R, Gonzalez FA, Leote J, Martins C, Bacariza J, Fernandes A, et al. Machine learning for the real-time assessment of left ventricular ejection fraction in critically ill patients: a bedside evaluation by novices and experts in echocardiography. Crit Care. 2022;26:386.

105. He B, Dash D, Duanmu Y, Tan TX, Ouyang D, Zou J. AI-enabled assessment of cardiac function and video quality in emergency department point-of-care echocardiograms. J Emerg Med. 2023. https://doi.org/https://doi.org/10.1016/j.jemermed.2023.02.005.

106. FDA Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices (2023). https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices. 2023.

107. Narang A, Bae R, Hong H, Thomas Y, Surette S, Cadieu C, et al. Utility of a Deep-Learning Algorithm to Guide Novices to Acquire Echocardiograms for Limited Diagnostic Use. JAMA Cardiol. 2021;6.

108. Komatsu M, Sakai A, Dozen A, Shozu K, Yasutomi S, Machino H, et al. Towards Clinical Application of Artificial Intelligence in Ultrasound Imaging. Biomedicines. 2021;9.

109. Chanh HQ, Ming DK, Nguyen QH, Duc TM, Phuoc An L, Trieu HT, et al. Applying artificial intelligence and digital health technologies, Viet Nam. Bull World Health Organ. 101:487–92.

110. Mollura DJ, Culp MP, Pollack E, Battino G, Scheel JR, Mango VL, et al. Artificial Intelligence in Low- and Middle-Income Countries: Innovating Global Health Radiology. Radiology. 2020;297:513–20.

111. Haniffa R, Pubudu De Silva A, Weerathunga P, Mukaka M, Athapattu P, Munasinghe S, et al. Applicability of the APACHE II model to a lower middle income country. J Crit Care. 2017;42:178–83.

112. Wilder-Smith A, Ooi E-E, Horstick O, Wills B. Dengue. The Lancet. 2019;393:350–63.

113. Organization WH. Global tuberculosis report 2022. World Health Organization; 2022.

114. Van Hao N, Yen LM, Davies-Foote R, Trung TN, Duoc NVT, Trang VTN, et al. The management of tetanus in adults in an intensive care unit in Southern Vietnam. Wellcome Open Res. 2021;6.

115. Whitehorn J, Yacoub S. Global warming and arboviral infections. Clinical Medicine. 2019;19:149.

116. Choisy M, McBride A, Chambers M, Quang CH, Quang HN, Chau NTX, et al. Climate change and health in Southeast Asia–defining research priorities and the role of the Wellcome Trust Africa Asia Programmes. Wellcome Open Res. 2021;6.

117. Bélard S, Tamarozzi F, Bustinduy AL, Wallrauch C, Grobusch MP, Kuhn W, et al. Point-of-care ultrasound assessment of tropical infectious diseases—a review of applications and perspectives. Am J Trop Med Hyg. 2016;94:8.

118. Henriquez-Camacho C, Garcia-Casasola G, Guillén-Astete C, Losa J. Ultrasound for the diagnosis of infectious diseases: approach to the patient at point of care and at secondary level. Journal of Infection. 2015;71:1–8.

119. Peiffer-Smadja N, Rawson TM, Ahmad R, Buchard A, Georgiou P, Lescure F-X, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clinical Microbiology and Infection. 2020;26:584–95.

120. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Health. 2018;3:e000798.

121. Wang J, Yang X, Zhou B, Sohn JJ, Zhou J, Jacob JT, et al. Review of machine learning in lung ultrasound in COVID-19 pandemic. J Imaging. 2022;8:65.

122. Cornuéjols A, Murena P-A, Olivier R. Transfer Learning by Learning Projections from Target to Source. In: Berthold MR, Feelders A, Krempl G, editors. Advances in Intelligent Data Analysis XVIII. Cham: Springer International Publishing; 2020. p. 119–31.

123. Alzubaidi L, Bai J, Al-Sabaawi A, Santamaría J, Albahri AS, Al-dabbagh BSN, et al. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. J Big Data. 2023;10:46.

124. Balestroni G, Bertolotti G. EuroQol-5D (EQ-5D): an instrument for measuring quality of life. Monaldi Archives for Chest Disease. 2012;78.

125. Brazier JE, Harper R, Jones NM, O'cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. Br Med J. 1992;305:160–4.

126. Association WM. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013;310:2191–4.

127. Leclerc S, Smistad E, Pedrosa J, Østvik A, Cervenansky F, Espinosa F, et al. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. IEEE Trans Med Imaging. 2019;38:2198–210.

128. Ouyang D, He B, Ghorbani A, Lungren MP, Ashley EA, Liang DH, et al. Echonet-dynamic: a large new cardiac motion video data resource for medical machine learning. In: NeurIPS ML4H Workshop: Vancouver, BC, Canada. 2019.

129. Dutta A, Zisserman A. The VIA Annotation Software for Images, Audio and Video. In: Proceedings of the 27th ACM International Conference on Multimedia. New York, NY, USA: ACM; 2019.

130. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, et al. 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. 2015;16:233–71.

131. Nhat PTH, Van Hao N, Tho PV, Kerdegari H, Pisani L, Thu LNM, et al. Clinical benefit of AI-assisted lung ultrasound in a resource-limited intensive care unit. Crit Care. 2023;27:257.

132. Riaz A, Cambaco O, Ellington LE, Lenahan JL, Munguambe K, Mehmood U, et al. Feasibility, usability and acceptability of paediatric lung ultrasound among healthcare providers and caregivers for the diagnosis of childhood pneumonia in resource-constrained settings: A qualitative study. BMJ Open. 2021;11.

133. Gibson PG, Qin L, Puah SH. COVID-19 acute respiratory distress syndrome (ARDS): clinical features and differences from typical pre-COVID-19 ARDS. Medical Journal of Australia. 2020;213.

134. Graça L, Abreu IG, Santos AS, Graça L, Dias PF, Santos ML. Descriptive Acute Respiratory Distress Syndrome (ARDS) in adults with imported severe Plasmodium falciparum malaria: A 10 year-study in a Portuguese tertiary care hospital. PLoS One. 2020;15:e0235437.

135. Kwizera A, Dünser MW. A global perspective on acute respiratory distress syndrome and the truth about hypoxia in resource-limited settings. American Journal of Respiratory and Critical Care Medicine. 2016;193.

136. Liu L, Lei W, Wan X, Liu L, Luo Y, Feng C. Semi-supervised active learning for COVID-19 lung ultrasound multi-symptom classification. In: 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI). IEEE; 2020. p. 1268–73.

137. Mento F, Perrone T, Fiengo A, Smargiassi A, Inchingolo R, Soldati G, et al. Deep learning applied to lung ultrasound videos for scoring COVID-19 patients: A multicenter study. J Acoust Soc Am. 2021;149.

138. Arntfield R, Wu D, Tschirhart J, Vanberlo B, Ford A, Ho J, et al. Automation of lung ultrasound interpretation via deep learning for the classification of normal versus abnormal lung parenchyma: A multicenter study. Diagnostics. 2021;11.

139. Kerdegari H, Phung NTH, McBride A, Pisani L, Van Nguyen H, Duong TB, et al. B-line detection and localization in lung ultrasound videos using spatiotemporal attention. Applied Sciences (Switzerland). 2021;11.

140. Kerdegari H, Nhat PTH, McBride A, Razavi R, Hao N Van, Thwaites L, et al. Automatic Detection of B-lines in Lung Ultrasound Videos from Severe Dengue Patients. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE; 2021.

141. Born J, Wiedemann N, Cossio M, Buhre C, Brändle G, Leidermann K, et al. Accelerating detection of lung pathologies with explainable ultrasound image analysis. Applied Sciences (Switzerland). 2021;11:1–23.

142. Camacho J, Muñoz M, Genovés V, Herraiz JL, Ortega I, Belarra A, et al. Artificial Intelligence and Democratization of the Use of Lung Ultrasound in COVID-19: On the Feasibility of Automatic Calculation of Lung Ultrasound Score. International Journal of Translational Medicine. 2022;2:17–25.

143. Tan GFL, Du T, Liu JS, Chai CC, Nyein CM, Liu AYL. Automated lung ultrasound image assessment using artificial intelligence to identify fluid overload in dialysis patients. BMC Nephrol. 2022;23:410.

144. Bolya D, Zhou C, Xiao F, Lee YJ. Yolact: Real-time instance segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019. p. 9157–66.

145. He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 2961–9.

146. Gottlieb M, Patel D, Viars M, Tsintolas J, Peksa GD, Bailitz J. Comparison of artificial intelligence versus real-time physician assessment of pulmonary edema with lung ultrasound. Am J Emerg Med. 2023.

147. Russell FM, Ehrman RR, Barton A, Sarmiento E, Ottenhoff JE, Nti BK. B-line quantification: comparing learners novice to lung ultrasound assisted by machine artificial intelligence technology to expert review. Ultrasound J. 2021;13:1–7.

148. Moore CL, Wang J, Battisti AJ, Chen A, Fincke J, Wang A, et al. Interobserver Agreement and Correlation of an Automated Algorithm for B-Line Identification and Quantification With Expert

Sonologist Review in a Handheld Ultrasound Device. Journal of Ultrasound in Medicine. 2022;41:2487–95.

149. Brusasco C, Santori G, Bruzzo E, Trò R, Robba C, Tavazzi G, et al. Quantitative lung ultrasonography: a putative new algorithm for automatic detection and quantification of B-lines. Crit Care. 2019;23:1–7.

150. Gomez A, Zimmer VA, Wheeler G, Toussaint N, Deng S, Wright R, et al. PRETUS: A plug-in based platform for real-time ultrasound imaging research. SoftwareX. 2022;17.

151. Needham DM, Davidson J, Cohen H, Hopkins RO, Weinert C, Wunsch H, et al. Improving long-term outcomes after discharge from intensive care unit: report from a stakeholders' conference. Crit Care Med. 2012;40:502–9.

152. Puthucheary ZA, Rawal J, McPhail M, Connolly B, Ratnayake G, Chan P, et al. Acute skeletal muscle wasting in critical illness. JAMA. 2013;310:1591–600.

153. Turan Z, Topaloglu M, Ozyemisci Taskiran O. Medical Research Council-sumscore: a tool for evaluating muscle weakness in patients with post-intensive care syndrome. Crit Care. 2020;24:1–2.

154. Parry SM, El-Ansary D, Cartwright MS, Sarwal A, Berney S, Koopman R, et al. Ultrasonography in the intensive care setting can be used to detect changes in the quality and quantity of muscle and is related to muscle strength and function. J Crit Care. 2015;30:1151-e9.

155. Trung TN, Duoc NVT, Nhat LTH, Yen LM, Hao N V, Truong NT, et al. Functional outcome and muscle wasting in adults with tetanus. Trans R Soc Trop Med Hyg. 2019;113.

156. Puthucheary ZA, Phadke R, Rawal J, McPhail MJW, Sidhu PS, Rowlerson A, et al. Qualitative ultrasound in acute critical illness muscle wasting. Crit Care Med. 2015;43:1603–11.

157. e Lima KMM, da Matta TT, de Oliveira LF. Reliability of the rectus femoris muscle cross-sectional area measurements by ultrasonography. Clin Physiol Funct Imaging. 2012;32:221–6.

158. Pardo E, El Behi H, Boizeau P, Verdonk F, Alberti C, Lescot T. Reliability of ultrasound measurements of quadriceps muscle thickness in critically ill patients. BMC Anesthesiol. 2018;18:1–8.

159. Sponbeck JK, Frandsen CR, Ridge ST, Swanson DA, Swanson DC, Johnson AW. Leg muscle cross-sectional area measured by ultrasound is highly correlated with MRI. J Foot Ankle Res. 2021;14:1–7.

160. Fazzini B, Märkl T, Costas C, Blobner M, Schaller SJ, Prowle J, et al. The rate and assessment of muscle wasting during critical illness: a systematic review and meta-analysis. Crit Care. 2023;27:1–26.

161. Puthucheary ZA, McNelly AS, Rawal J, Connolly B, Sidhu PS, Rowlerson A, et al. Rectus femoris cross-sectional area and muscle layer thickness: comparative markers of muscle wasting and weakness. Am J Respir Crit Care Med. 2017;195:136–8.

162. van Alfen N, Gijsbertse K, de Korte CL. How useful is muscle ultrasound in the diagnostic workup of neuromuscular diseases? Curr Opin Neurol. 2018;31.

163. Casey P, Alasmar M, McLaughlin J, Ang Y, McPhee J, Heire P, et al. The current use of ultrasound to measure skeletal muscle and its ability to predict clinical outcomes: a systematic review. J Cachexia Sarcopenia Muscle. 2022;13:2298–309.

164. Andrade-Junior MC de, Salles ICD de, de Brito CMM, Pastore-Junior L, Righetti RF, Yamaguti WP. Skeletal muscle wasting and function impairment in intensive care patients with severe COVID-19. Front Physiol. 2021;12:640973.

165. Parry SM, El-Ansary D, Cartwright MS, Sarwal A, Berney S, Koopman R, et al. Ultrasonography in the intensive care setting can be used to detect changes in the quality and quantity of muscle and is related to muscle strength and function. J Crit Care. 2015;30:1151-e9.

166. Zhang W, Wu J, Gu Q, Gu Y, Zhao Y, Ge X, et al. Changes in muscle ultrasound for the diagnosis of intensive care unit acquired weakness in critically ill patients. Sci Rep. 2021;11:18280.

167. Mayer KP, Thompson Bastin ML, Montgomery-Yates AA, Pastva AM, Dupont-Versteegden EE, Parry SM, et al. Acute skeletal muscle wasting and dysfunction predict physical disability at hospital discharge in patients with critical illness. Crit Care. 2020;24:1–12.

168. Sobolewski EJ, Wein LD, Crow JM, Carpenter KM. Intra-rater and inter-rater reliability of the process of obtaining cross-sectional area and echo intensity measurements of muscles from ultrasound images. J Ultrason. 2021;21:7–11.

169. Seymour JM, Ward K, Sidhu PS, Puthucheary Z, Steier J, Jolley CJ, et al. Ultrasound measurement of rectus femoris cross-sectional area and the relationship with quadriceps strength in COPD. Thorax. 2009;64:418–23.

170. Pardo E, El Behi H, Boizeau P, Verdonk F, Alberti C, Lescot T. Reliability of ultrasound measurements of quadriceps muscle thickness in critically ill patients. BMC Anesthesiol. 2018;18:1–8.

171. Puthucheary ZA, McNelly AS, Rawal J, Connolly B, Sidhu PS, Rowlerson A, et al. Rectus femoris cross-sectional area and muscle layer thickness: comparative markers of muscle wasting and weakness. Am J Respir Crit Care Med. 2017;195:136–8.

172. Nijholt W, Scafoglieri A, Jager-Wittenaar H, Hobbelen JSM, van der Schans CP. The reliability and validity of ultrasound to quantify muscles in older adults: a systematic review. J Cachexia Sarcopenia Muscle. 2017;8:702–12.

173. Hammond K, Mampilly J, Laghi FA, Goyal A, Collins EG, McBurney C, et al. Validity and reliability of rectus femoris ultrasound measurements: comparison of curved-array and linear-array transducers. J Rehabil Res Dev. 2014;51:1155.

174. Popović ZB, Thomas JD. Assessing observer variability: a user's guide. Cardiovasc Diagn Ther. 2017;7:317.

175. Kottner J, Streiner DL. The difference between reliability and agreement. J Clin Epidemiol. 2011;64:701.

176. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155–63.

177. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1:30.

178. Ritsche P, Wirth P, Franchi M V., Faude O. ACSAuto-semi-automatic assessment of human vastus lateralis and rectus femoris cross-sectional area in ultrasound images. Sci Rep. 2021;11.

179. Chen X, Xie C, Chen Z, Li Q. Automatic Tracking of Muscle Cross-Sectional Area Using Convolutional Neural Networks with Ultrasound. Journal of Ultrasound in Medicine. 2019;38.

180. Katakis S, Barotsis N, Kakotaritis A, Tsiganos P, Economou G, Panagiotopoulos E, et al. Muscle Cross-Sectional Area Segmentation in Transverse Ultrasound Images Using Vision Transformers. Diagnostics. 2023;13:217.

181. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, Ashrafian H, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Lancet Digit Health. 2020;2:e537–48.

182. Mourtzakis M, Parry S, Connolly B, Puthucheary Z. Skeletal muscle ultrasound in critical care: a tool in need of translation. Ann Am Thorac Soc. 2017;14:1495–503.

183. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med. 1998;17:101–10.

184. Christensen HS, Borgbjerg J, Børty L, Bøgsted M. On Jones et al.'s method for extending Bland-Altman plots to limits of agreement with the mean for multiple observers. BMC Med Res Methodol. 2020;20:1–8.

185. Bergamaschi V, Vignazia GL, Messina A, Colombo D, Cammarota G, Corte F Della, et al. Transthoracic echocardiographic assessment of cardiac output in mechanically ventilated critically ill patients by intensive care unit physicians. Brazilian Journal of Anesthesiology (English Edition). 2019;69.

186. Neskovic AN, Hagendorff A, Lancellotti P, Guarracino F, Varga A, Cosyns B, et al. Emergency echocardiography: the European association of cardiovascular imaging recommendations. European Heart Journal–Cardiovascular Imaging. 2013;14:1–11.

187. Labovitz AJ, Noble VE, Bierig M, Goldstein SA, Jones R, Kort S, et al. Focused cardiac ultrasound in the emergent setting: a consensus statement of the American Society of Echocardiography and American College of Emergency Physicians. Journal of the American Society of Echocardiography. 2010;23:1225–30.

188. Mandavia DP, Hoffner RJ, Mahaney K, Henderson SO. Bedside echocardiography by emergency physicians. Ann Emerg Med. 2001;38:377–82.

189. Hutyra M, Paleček T, Hromádka M. The use of echocardiography in acute cardiovascular care. Summary of the document prepared by the Czech Society of Cardiology. Cor Vasa. 2018;60:e70–88.

190. Kirkpatrick JN, Grimm R, Johri AM, Kimura BJ, Kort S, Labovitz AJ, et al. Recommendations for echocardiography laboratories participating in cardiac point of care cardiac ultrasound (POCUS) and critical care echocardiography training: report from the American Society of Echocardiography. Journal of the American Society of Echocardiography. 2020;33:409–22.

191. Kumar A, Kugler J, Jensen T. Evaluation of trainee competency with point-of-care ultrasonography (POCUS): a conceptual framework and review of existing assessments. J Gen Intern Med. 2019;34:1025–31.

192. Oren-Grinberg A, Talmor D, Brown SM. Concise definitive review: focused critical care echocardiography in the ICU. Crit Care Med. 2013;41:2618.

193. Buesing J, Weng Y, Kugler J, Wang L, Blaha O, Hom J, et al. Handheld Ultrasound Device Usage and Image Acquisition Ability Among Internal Medicine Trainees: A Randomized Trial. J Grad Med Educ. 2021;13:76–82.

194. Kumar A, Weng Y, Bentley J, Almli M, Hom J, Witteles R, et al. Portable ultrasound device usage and learning outcomes among internal medicine trainees: a parallel-group randomized trial. J Hosp Med. 2020;15:154–9.

195. Kosaraju A, Goyal A, Grigorova Y, Makaryus AN. Left ventricular ejection fraction. 2017.

196. Cannesson M, Tanabe M, Suffoletto MS, McNamara DM, Madan S, Lacomis JM, et al. A novel two-dimensional echocardiographic image analysis system using artificial intelligence-learned pattern recognition for rapid automated ejection fraction. J Am Coll Cardiol. 2007;49:217–26.

197. Rahmouni HW, Ky B, Plappert T, Duffy K, Wiegers SE, Ferrari VA, et al. Clinical utility of automated assessment of left ventricular ejection fraction using artificial intelligence–assisted border detection. Am Heart J. 2008;155:562–70.

198. Smistad E, Østvik A, Salte IM, Melichova D, Nguyen TM, Haugaa K, et al. Real-Time Automatic Ejection Fraction and Foreshortening Detection Using Deep Learning. IEEE Trans Ultrason Ferroelectr Freq Control. 2020;67:2595–604.

199. Knackstedt C, Bekkers SCAM, Schummers G, Schreckenberg M, Muraru D, Badano LP, et al. Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain: the FAST-EFs multicenter study. J Am Coll Cardiol. 2015;66:1456–66.

200. Abdi AH, Luong C, Tsang T, Allan G, Nouranian S, Jue J, et al. Automatic quality assessment of echocardiograms using convolutional neural networks: feasibility on the apical four-chamber view. IEEE Trans Med Imaging. 2017;36:1221–30.

201. Liao Z, Girgis H, Abdi A, Vaseli H, Hetherington J, Rohling R, et al. On modelling label uncertainty in deep neural networks: Automatic estimation of intra-observer variability in 2d echocardiography quality assessment. IEEE Trans Med Imaging. 2019;39:1868–83.

202. Nagata Y, Yuichiro K, Takeshi O, Kyoko O, Akemi N, Yutaka O, et al. Impact of image quality on reliability of the measurements of left ventricular systolic function and global longitudinal strain in 2D echocardiography. Echo Res Pract. 2018;5:28–39.

203. Sassaroli E, Crake C, Scorza A, Kim D-S, Park M-A. Image quality evaluation of ultrasound imaging systems: advanced B-modes. J Appl Clin Med Phys. 2019;20:115–24.

204. Ha H, Samuel S, Khalid CA, Nripesh P, Charles C, Randolph M, et al. AI-GUIDED ECHOCARDIOGRAPHY SYSTEM MATCHES THE IMAGE QUALITY ASSESSMENT ABILITY OF CARDIAC SONOGRAPHERS. J Am Coll Cardiol. 2021;77 18_Supplement_1:3240.

205. Pasdeloup D, Olaisen SH, Østvik A, Sabo S, Pettersen HN, Holte E, et al. Real-time echocardiography guidance for optimized apical standard views. Ultrasound Med Biol. 2023;49:333–46.

206. Baum E, Tandel M, Ren C, Weng Y, Pascucci M, Kugler J, et al. Use of Artificial Intelligence for Acquisition of Limited Echocardiograms: A Randomized Controlled Trial for Educational Outcomes. medRxiv. 2023;:2023–4.

207. Cheema BS, Walter J, Narang A, Thomas JD. Artificial Intelligence–Enabled POCUS in the COVID-19 ICU: A New Spin on Cardiac Ultrasound. JACC Case Rep. 2021;3:258–63.

208. Schneider M, Bartko P, Geller W, Dannenberg V, König A, Binder C, et al. A machine learning algorithm supports ultrasound-naïve novices in the acquisition of diagnostic echocardiography loops and provides accurate estimation of LVEF. Int J Cardiovasc Imaging. 2021;37:577–86.

209. Labs RB, Vrettos A, Loo J, Zolgharni M. Automated assessment of transthoracic echocardiogram image quality using deep neural networks. Intelligent Medicine. 2023;3:191–9.

210. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. Journal of the Royal Statistical Society Series D: The Statistician. 1983;32:307–17.

211. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med. 2018;18:91–3.

212. Rajpurkar P, Lungren MP. The Current and Future State of AI Interpretation of Medical Images. New England Journal of Medicine. 2023;388:1981–90.

213. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. New England Journal of Medicine. 2023;388:1201–8.

214. Yoon JH, Pinsky MR, Clermont G. Artificial intelligence in critical care medicine. Annual Update in Intensive Care and Emergency Medicine 2022. 2022;:353–67.

215. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28:31–8.

216. Zeitoun J-D, Ravaud P. Artificial intelligence in health care: value for whom? Lancet Digit Health. 2020;2:e338–9.

217. Nhat PTH, Van HN, Lam MY, Nguyen HA, Dong PK, Kerdegari H, et al. Clinical evaluation of AI-assisted muscle ultrasound for monitoring muscle wasting in ICU patients. 2023.

Appendices

Appendix A. Ethical approval documents

A1. Ethical approval of Oxford Tropical Research Ethics Committee (English)

Oxford Tropical Research Ethics Committee

University of Oxford
Research Services, University Offices
Wellington Square, Oxford OX1 2JD
Tel. +44 (0)1865 (2)82106
E-mail: oxtrec@admin.ox.ac.uk

Dr. Sophie Yacoub
Oxford University Clinical Research Unit
764 Vo Van Kiet
Ward 1, District 5
Ho Chi Minh City
Vietnam

24 March 2020

Dear Dr. Yacoub,

Full Title of Study: Developing automated point of care ultrasound for Cardiac, Lung and Muscle examination in critical illness

OxTREC Reference: 516-20

Thank you for your email of the 16 March 2020, and for your minimal risk application form.

I am pleased to confirm that approval has now been granted for this study. This is valid for the first five years and is subject to receiving the local ethical approval (if this approval has not yet been received).

The documents approved for this study are as follows:

Documents:	Version:	Date:
Minimal risk application form		
PIS/ICF/AF	V2.0	18/11/19
EQ-5D Questionnaire		
SF-36 Survey		

Any subsequent changes to the application must be submitted to the Committee as an Amendment. This should include a letter to give the reasons for the proposed modifications and all revised documents with changes tracked.

Please ensure that you submit a completed Annual Report form on every anniversary of this approval and a final End of Study Report. The relevant forms can be found on the OxTREC website: https://researchsupport.admin.ox.ac.uk/governance/ethics/apply/oxtrec.

Finally, please note the following **important information**:

Data safety—all studies
It is the responsibility of the PI to ensure that all data collected during the course of the study is stored and transferred safely and securely. Further guidance and advice is available from the Research Data Team.

Tel: +44 (0)1865 (2)82106
Email: oxtrec@admin.ox.ac.uk
Web: https://researchsupport.admin.ox.ac.uk/governance/ethics

Studies that will involve storing human tissue samples in Oxford
As you are planning to import the samples into England, you will need to make arrangements before the samples are transferred to store them under the governance of a Human Tissue Authority (HTA) licence. It is a legal requirement that any tissue or fluid made up of or containing human cells to be used for the purpose of research is stored on premises licensed by the HTA unless covered by an exemption. OxTREC approval is not a recognised exemption. Further information may be found on the University's human tissue governance web pages: https://researchsupport.admin.ox.ac.uk/governance/human-tissue.

Yours sincerely

Rebecca C T Bryant

Dr Rebecca Bryant
Research Ethics Manager, OxTREC

A2. Ethical approval of Hospital for Tropical Diseases at Ho Chi Minh city (Vietnamese)

SỞ Y TẾ TP. HỒ CHÍ MINH
BỆNH VIỆN BỆNH NHIỆT ĐỚI

Số : 2016 /QĐ-BVBNĐ

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc

Thành phố Hồ Chí Minh, ngày 19 tháng 11 năm 2020

QUYẾT ĐỊNH
Về việc phê duyệt đề tài khoa học công nghệ cấp cơ sở

GIÁM ĐỐC BỆNH VIỆN BỆNH NHIỆT ĐỚI

Căn cứ Quyết định số 5178/QĐ-SYT ngày 29 tháng 7 năm 2020 của Sở Y tế thành phố Hồ Chí Minh ban hành Quy chế tổ chức và hoạt động của Bệnh viện Bệnh Nhiệt đới trực thuộc Sở Y tế;

Căn cứ nội dung phiên họp ngày 18/11/2020 và Chấp thuận của Hội đồng Đạo đức trong nghiên cứu y sinh học Bệnh viện Bệnh Nhiệt đới số 2043/BVBNĐ-HĐĐĐ ngày 18/11/2020;

Xét đề nghị của Chủ tịch Hội đồng Đạo đức trong nghiên cứu y sinh học Bệnh viện Bệnh Nhiệt đới,

QUYẾT ĐỊNH:

Điều 1: Nay phê duyệt bổ sung, chỉnh sửa, đề tài nghiên cứu khoa học cấp cơ sở:

1. Tên nghiên cứu: Thiết kế tự động hóa trong siêu âm tim, phổi và cơ tại giường ở bệnh nhân hồi sức tích cực (01NVb).
   - Thuyết minh đề tài, phiên bản 4.0 ngày 30/9/2020,
   - Đề cương nghiên cứu, phiên bản 4.0 ngày 30/9/2020.
2. Chủ nhiệm đề tài: TS.BS Sophie Yacoub, TS.BS Nguyễn Văn Hảo.
   Nghiên cứu viên tham gia chính tại Bệnh viện Bệnh Nhiệt đới: ThS.BS Hà Thị Hải Đường, ThS.BS Nguyễn Quốc Việt, BSCKII. Phan Vĩnh Thọ, TS.BS Hồ Đặng Trung Nghĩa, TS.BS Nguyễn Hoan Phú, BSCKII. Nguyễn Thanh Phong, BSCKII. Lương Thị Huệ Tài.
3. Mã số nghiên cứu: CS/BND/19/38.
4. Tổ chức chủ trì đề tài: Bệnh viện Bệnh Nhiệt đới.
5. Địa điểm thực hiện: Bệnh viện Bệnh Nhiệt đới, Bệnh viện Bệnh Nhiệt đới Trung ương.
6. Thời gian nghiên cứu: từ tháng 01/2020 đến tháng 01/2023.
7. Nguồn kinh phí: Dự Án Hỗ trợ kỹ thuật Nghiên cứu những bệnh nhiễm trùng quan trọng đối với sức khỏe cộng đồng tại Việt Nam (giai đoạn 2020-2030)..

Điều 2: Các chủ nhiệm đề tài; các Khoa/ Phòng; bộ phận có liên quan theo nội dung đề cương nghiên cứu được phê duyệt chịu trách nhiệm phối hợp thực hiện nghiên cứu đúng quy định của nhà nước về Nghiên cứu khoa học.

Điều 3: Quyết định có hiệu lực thi hành kể từ ngày ký và thay thế Quyết định số 50/QĐ-BVBNĐ ngày 27 tháng 12 năm 2019 của Bệnh viện Bệnh Nhiệt đới./.

Nơi nhận:
- Chủ nhiệm đề tài;
- Lưu: VT, P.QLCL.

GIÁM ĐỐC

SỞ Y TẾ THÀNH PHỐ HỒ CHÍ MINH
BỆNH VIỆN BỆNH NHIỆT ĐỚI

TS.BS. NGUYỄN VĂN VĨNH CHÂU

A3. Informed consent form for patient (English)

PARTICIPANT INFORMATION SHEET
Title: Developing automated point of care ultrasound for Cardiac, Lung and Muscle examination in critical illness

You are being invited to take part in a research study. Please read this information sheet carefully. You will be given a copy of this form to keep. Please ask the study staff to explain any information that you are not sure about.

What is the reason for doing the study?
Very sick patients need to be cared for in an intensive care unit and carefully monitored. One of the ways of monitoring patients is to do ultrasound scans. This can be of the heart, lungs or muscles and the information lets doctors know how better to treat you. Although ultrasound is very safe, it is very difficult to perform and the doctors doing it need a lot of training. In this study we aim to use new computer-assisted ways of helping doctors do and read ultrasound scans and also for some patients to see how the results of these scans link with how you recover after leaving hospital. This information will help us treat patients better in the future.

What will happen if you participate in the study?
If you agree to participate in the study, your basic clinical and demographic data will be collected.

What are the possible risks of the study?
Ultrasound is very safe and we do not think you will suffer any risks from this study.

What are the possible benefits of the study?
If the doctors find that you are having problems with your health, they will be able to help by either giving you advice or referring you to other people who can.
You will receive 100 000 VND (~ $4) as compensation for your time participating in the study.

Do I have to participate?
Being in a research study is your decision. If you do not want to be in the study or if at any time during the study you decide to stop participating, the doctors will respect your decision.
Will anyone know that I am participating in this study?
All information about you will be kept confidential. Your medical records will be reviewed in strict confidence by those who are working on this study and may also be reviewed by the ethics committees and health authorities reviewing the study. Your name will not be used on any of the study documents or on the stored samples or in any reports or publications about this study.

What if I have more questions?

You are encouraged to ask any questions related to this study during the time of participation. If you have any questions about this program, its procedures, risks and benefits, or alternatives please call Dr Nguyen Van Hao at 0913857025.
If you have any questions about your rights as a subject in this study or if you want to speak to someone outside of the program you may contact the Ethics Committee at the Hospital for Tropical Diseases at (+84)23924 2661.

Data protection and data sharing
The University of Oxford is responsible for ensuring the safe and proper use of any personal information you provide, solely for research purposes.
Thank you for your time and your consideration to participate in this study.

A4. Informed consent form for healthy volunteer (English)

INFORMED CONSENT FORM
(For healthy volunteer, >= 18 year old)

Developing automated point of care ultrasound for Cardiac, Lung and Muscle examination in critical illness

- I have read the information given to me and freely agree to be in this study. I also have had a chance to discuss it with the study staff.
- I have been told about the risks and benefits. I got answers that I could understand to all my questions.
- I understand that I can withdraw from the study at any time. If I stop the study, it will **not** affect my future care. If I decide to stop the study, I agree that the information collected up to the point when I stop, may continue to be used.

PARTICIPANT # 01NVb-[__] [__] [__]-[__] [__] [__]

By signing my name here, I confirm what is written above.

x_____	x_____	___/____/_____
Participant Signature	Print Name	Date of Signature

Investigator/Designee:
I, the undersigned, have fully explained the relevant information of this study to the person named above and will provide her/him with a copy of this signed and dated informed consent form.

x_____	x_____	___/____/_____
Investigator/Designee Signature	Print Name	Date of Signature

A5. Informed consent form for patient (Vietnamese)

THÔNG TIN DÀNH CHO ĐỐI TƯỢNG THAM GIA NGHIÊN CỨU

Tên đề tài: Thiết kế tự động hóa trong siêu âm tim, phổi và cơ tại giường ở bệnh nhân hồi sức tích cực

Bạn được mời tham gia một nghiên cứu khoa học. Xin vui lòng đọc kỹ phiếu thông tin này. Bạn sẽ được giữ một bản phiếu thông tin này. Vui lòng hỏi nhân viên nghiên cứu để được giải thích về các thông tin mà bạn chưa hiểu rõ.

Mục đích của nghiên cứu

Bệnh nhân nặng cần được điều trị và chăm sóc tại khoa hồi sức tích cực và cần được theo dõi chặt chẽ. Một trong những phương pháp để theo dõi bệnh nhân là siêu âm. Các thông tin về tim, phổi và cơ sẽ giúp bác sĩ trong công tác điều trị bệnh cho bệnh nhân. Mặc dù siêu âm rất an toàn nhưng việc thực hiện siêu âm khá khó khăn, và các bác sĩ thực hiện cần được đào tạo chuyên sâu. Mục tiêu của nghiên cứu này là sử dụng các phương pháp hỗ trợ từ máy tính để giúp bác sĩ thực hiện và giải thích hình ảnh siêu âm. Nghiên cứu cũng giúp các bệnh nhân biết được sự liên kết giữa kết quả của các lần siêu âm và tình trạng cải thiện của mình sau khi xuất viện. Thông tin này sẽ giúp bác sĩ trong công tác điều trị bệnh nhân tốt hơn trong tương lai.

Điều gì sẽ xảy ra nếu bạn đồng ý tham gia nghiên cứu?

Nếu bạn đồng ý tham gia nghiên cứu, dữ liệu về dấu hiệu lâm sàng (nhịp tim vào huyết áp) và thông tin nhân khẩu học sẽ được thu thập.

Bạn sẽ được siêu âm tim. Quá trình siêu âm sẽ cho một ít gel lên ngực của bạn và hoàn toàn không gây đau. Quá trình siêu âm sẽ kéo dài ít hơn 20 phút. Siêu âm sẽ được 2 lần bởi 2 nhân viên nghiên cứu.

Các nguy cơ có thể có khi tham gia nghiên cứu

Siêu âm rất an toàn và nhóm nghiên cứu dự đoán rằng bạn sẽ không có gặp nguy cơ bất lợi gì trong khi tham gia nghiên cứu.

Các lợi ích có thể có khi tham gia nghiên cứu

Nếu có bất cứ vấn đề gì sẽ được thông báo lại với với bạn và tư vấn nơi thăm khám phù hợp.

Nghiên cứu sẽ bồi dưỡng cho thời gian bạn tham gia nghiên cứu là 100.000 VNĐ.

Tôi có bắt buộc phải tham gia nghiên cứu

Việc tham gia nghiên cứu hoàn toàn phụ thuộc quyết định của bạn. Nếu bạn không muốn tham gia nghiên cứu hoặc trong quá trình nghiên cứu bạn không muốn tiếp tục, bác sĩ sẽ tôn trọng quyết định của bạn.

Có người nào biết tôi đang tham gia nghiên cứu không?

Tất cả thông tin về bạn sẽ được bảo mật. Các thông tin sức khỏe của bạn sẽ được xem xét và bảo mật bởi các thành viên nhóm nghiên cứu và cũng có thể được xem xét bởi hội đồng đạo đức và người quản lý nghiên cứu. Tên của bạn sẽ không sử dụng trong bất cứ tài liệu nghiên cứu hoặc dữ liệu nghiên cứu hoặc báo cáo hay bài báo nào về nghiên cứu.

Nếu tôi có thắc mắc?

- Bạn được khuyết khích đặt các câu hỏi liên quan đến nghiên cứu trong suốt thời gian tham gia nghiên cứu. Nếu bạn có bất kỳ thắc mắc nào về nghiên cứu, quy trình nghiên cứu, nguy cơ và lợi ích, các sự chọn lựa xin vui lòng gọi điện thoại cho Bác sĩ Nguyễn Văn Hảo tại số 0913857025.
- Nếu bạn có bất kỳ thắc mắc nào về quyền của bạn khi tham gia nghiên cứu hoặc nếu bạn muốn hỏi những người ngoài nghiên cứu, bạn có thể liên lạc với Hội đồng Đạo đức Bệnh viện Bệnh Nhiệt đới qua số (+84) 283924 2661.

Bảo mật dữ liệu

Đại học Oxford có trách nhiệm đảm bảo các dữ liệu cá nhân anh/chị đã cung cấp được an toàn và chỉ sử dụng cho mục đích nghiên cứu.
Cảm ơn anh/chị đã dành thời gian cho chúng tôi và đã cân nhắc tham gia nghiên cứu này.

A4. Informed consent form for healthy volunteer (Vietnamese)

PHIẾU CHẤP THUẬN THAM GIA NGHIÊN CỨU

(Người tình nguyện khỏe mạnh, >=18 tuổi)

Thiết kế tự động hóa trong siêu âm tim, phổi và cơ tại giường ở bệnh nhân hồi sức tích cực

- Tôi đã đọc/được nghe đọc các thông tin cung cấp cho tôi và tự nguyện đồng ý tham gia vào nghiên cứu. Tôi cũng đã có cơ hội thảo luận với nhân viên nghiên cứu.
- Tôi đã được tư vấn về nguy cơ và lợi ích. Tôi đã nhận được câu trả lời mà tôi có thể hiểu cho tất cả các câu hỏi của tôi.
- Tôi hiểu rằng tôi có quyền rút khỏi nghiên cứu bất cứ lúc nào tôi muốn. Việc tôi rút khỏi nghiên cứu sẽ không ảnh hưởng đến việc điều trịcủa tôi. Nếu tôi rút khỏi nghiên cứu, tôi đồng ý các thông tin sẽ đã được thu thập đến thời điểm đó sẽ được sử dụng

Mã bệnh nhân: 01NVb-[__][__][__]-[__][__][__]

Người tham gia nghiên cứu:

Bằng việc ký tên vào đây, tôi xác nhân những thông tin được viết ở trên.

x _____	x _____	____/____/_____
Chữ ký người tham gia	Tên	Ngày ký

Nghiên cứu viên/ Người được ủy quyền

Tôi, người ký tên dưới đây, đã giải thích đầy đủ các thông tin liên quan đến nghiên cứu cho người có tên ở trên và sẽ cung cấp cho anh/chị ấy một bản sao Phiếu chấp thuận đã được ký tên và ghi ngày tháng của người chấp thuận.

x _____	x _____	____/____/_____
Chữ ký của nghiên cứu viên/ người được ủy quyền	Tên	Ngày ký

OUTLINE	
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

Appendix B. Case report form (English + patient's questionnaires in Vietnamese)

- Screening
  - Group 1
  - Group 2
- Admission (general)
- Admission
  - Group 1
    - Dengue and Sepsis
    - Tetanus
  - Group 2
    - CNS infection(s)
    - Tetanus
  - Healthy Volunteer
- Cardiac and lung ultrasound (T1, T2, T3) group 1 only
- Muscle ultrasound (T1, T2, ICU discharge) group 2 only
- Hospital discharge
- Questionnaires (discharge, 6months) group 2 only

ADMISSION	**ADM**
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

1. Participant's name: [_____]*	
2. Year of birth: [___][___][___][___] or Age [___][___] years	
3. Sex: ☐ Male ☐ Female	
Inclusion Criteria (All answers must be YES for patients enrolled in the study)	
4. Is the participant ≥16 years old?	☐ YES ☐ No
5. Does the participant have sepsis or dengue or tetanus or tuberculosis meningitis? Group 1: ☐ YES Sepsis ☐ YES Dengue ☐ YES Tetanus ☐ YES Group 2: ☐ YES Tetanus Ablett Grade 3 or 4 ☐ YES Diagnosis as CNS infection(s) ☐ YES Admitted to ICU < 72 hours ☐ YES Duration of ICU stay expected at least 5 days ☐ YES	☐ YES ☐ No
6. Did the participant or their representative consent to participate in the study?	☐ YES ☐ No
7. Are study doctors and equipment available to do ultrasound?	☐ YES ☐ No
Exclusion Criteria (All answers must be NO for patients enrolled in the study)	
8. Contraindication to ultrasound scan	☐ YES ☐ No

ADMISSION	**ADM**
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

1. Hospital Number: [___][___][___][___][___][___][___][___][___][___][___][___]

2. Occupation: [_____]

3. Current place of living: District [_____]
Province [_____]

4. Participant's phone number: [_____]*

5. Representative's phone number: [_____]*

6. Date of admission to HTD: [___][___]/[___][___]/[___][___] *(dd/mm/yy)*

7. Date and time of admission to ICU at HTD:

[___][___]/[___][___]/[___][___] *(dd/mm/yy)* at [___][___]:[___][___] *(use 24 Hr clock)*

8. Was the participant transferred from another hospital? ☐ Yes ☐ No Name: _____

a. If yes, specify date of admission there: [___][___]/[___][___]/[___][___] *(dd/mm/yy)*

9. Self-reported comorbidities prior to admission *(fill in all rows)*

Hypertension	☐ Yes ☐ No ☐ NA	Angina on minimal exertion/rest pain	☐ Yes ☐ No ☐ NA
Myocardial infarction	☐ Yes ☐ No ☐ NA	Congestive cardiac failure (Grade I-III)	☐ Yes ☐ No ☐ NA
Anaemia	☐ Yes ☐ No ☐ NA	Congestive cardiac failure Grade IV	☐ Yes ☐ No ☐ NA
Peripheral vascular disease	☐ Yes ☐ No ☐ NA	Cerebrovascular disease	☐ Yes ☐ No ☐ NA
Chronic pulmonary disease	☐ Yes ☐ No ☐ NA	Severe respiratory deficiency	☐ Yes ☐ No ☐ NA
Connective tissue disease	☐ Yes ☐ No ☐ NA	Peptic ulcer disease	☐ Yes ☐ No ☐ NA
Mild liver disease	☐ Yes ☐ No ☐ NA	Moderate/severe liver disease	☐ Yes ☐ No ☐ NA

ADMISSION	**ADM**
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

Hemiplegia or paraplegia	☐ Yes ☐ No ☐ NA	Moderate/severe kidney disease	☐ Yes ☐ No ☐ NA
Diabetes	☐ Yes ☐ No ☐ NA	Diabetes with chronic complications	☐ Yes ☐ No ☐ NA
Dementia	☐ Yes ☐ No ☐ NA	Metastatic solid tumour	☐ Yes ☐ No ☐ NA
AIDS	☐ Yes ☐ No ☐ NA	Any malignancy	☐ Yes ☐ No ☐ NA
Smoker Number/day	☐ Yes ☐ No ☐ NA _____	Elective surgery within last 30 days	☐ Yes ☐ No ☐ NA
Other (specify)	_____	Immunocompromised	☐ Yes ☐ No ☐ NA

10. Taking any regular medication ☐ Yes ☐ No ☐ NA If yes, specify details _____

11. Taking any herbal/natural remedies ☐ Yes ☐ No ☐ NA
If yes, specify details _____

12. Weight: |__|__|__] Kg Height |__|__|__] cm

13. Admission blood results (if multiple results, record worst result in 1st 24 hrs)

Haemoglobin [__|__].[__] g/dl Platelets [__|__|__] K/µL WCC [__][__].[__]K/µL

Bilirubin [__|__|__] µmol/L Na^+ [__|__|__] mmol/L K^+ [__].[__|__]mmol/L

Creatinine [__|__|__] µmol/L TNT [__|__|__] pg/ml

Albumin [__|__|__] g/L Lactate [__].[__|__] mmol/L pH [__].[__|__]

HbA1C [__|__|__] mmol/mol PT [__|__] secs APTT [__|__] secs

HCT [__|__]% **Specify:** Point of care ☐ Laboratory ☐

DENGUE AND SEPSIS	ADM
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

ADMISSION (complete within 24 hours of enrolment)

Participant's location: Emergency ward ○ ICU ○ Ward C ○ Ward D ○

14. Date of illness/fever onset: [___][___]/[___][___]/[___][___] *(dd/mm/yy)*

15. Date and time of shock onset (if known):

16. [___][___]/[___][___]/[___][___] *(dd/mm/yy)* at [___][___]:[___][___] *(use 24 Hr clock)*

17. Treatment received for this illness prior to HTD admission:

Antibiotics ☐ Yes ☐ No Details: _____

Intravenous fluid/infusions ☐ Yes ☐ No Details: _____

Other: ☐ Yes ☐ No Details: _____

18. Haemodynamics

Clinical shock ☐ Yes ☐ No

Dengue: re-shock? ☐ Yes ☐ No

Min BP [__|__|__]/[__|__|__]mmHg

Max HR [__|__|__] bpm

Vasopressors ☐ Yes ☐ No

If yes, Name : _____

Dose _____ µg/24hrs

Peripheral oedema ☐ Yes ☐ No

Tmax [__|__|.|__]°C

Urine output[__|__|__|__]mL/24hr

19. Respiratory

Respiratory distress Yes ☐ No ☐

Max Resp. rate [__|__] /min

Oxygen requirement Yes ☐ No ☐

Max. FiO_2 [__|__|__]%

Lowest SpO_2 [__|__|__]%

Lowest PaO_2 [__|__|__]mmHg

CPAP/BiPAP Yes ☐ No ☐

Intubated Yes ☐ No ☐

20. Neurological

GCS [__|__]/15

RASS score + - [__]

Sedation Yes ☐ No ☐ If yes: Name:

_____

Dose: _____ µg/24hrs

Name: _____

Dose: _____ µg/24hrs

Convulsions Yes ☐ No ☐

Localising neurology Yes ☐ No ☐

21. Bleeding Yes ☐ No ☐

Clinically severe? Yes ☐ No ☐

Bruising/petechiae Yes ☐ No ☐

Gum/nose bleeding Yes ☐ No ☐

Hematemesis/melaena Yes ☐ No ☐

Hematuria Yes ☐ No ☐

Other places:_____

22. Acute Kidney Injury Yes ☐ No ☐

Urine output <0.5ml/kg/hr ☐

Creatinine >1.5 normal for age ☐

Haemofiltration Yes ☐ No ☐

Acute liver injury Yes ☐ No ☐

TETANUS	ADM
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

ADMISSION (complete within 24 hours of enrolment)	
22. Time from first symptom to hospital admission [__|__] days	
23. Incubation period [__|__] days N/A ○ Period of onset [__|__|__] hours	
24. Wound: Internal or injection ○ Other (or unknown) ○	
25. Difficulty breathing on admission to hospital ☐ Yes ☐ No	
26. ○ I *(no spasm)* ○ II *(spasm not interfering with respiration)* ○ III *(spasm interfering with respiration* ○ IV *(III plus ANSD)*	
27. ASA	Fit and well ○ Minor illness or injury ○ Moderately severe disease ○ Severe illness not immediately life-threatening ○ Life-threatening illness ○
28. SOFA/ APACHE/ TSS	Worst value in 1st 24 hours ICU Highest temperature [__|__|.|__]°C GCS [__|__] RR [__|__]bpm FiO_2 [__|__|% SpO_2 [__|__|% PaO_2 [__|__|__]mmHg pH [__|. [__|__] Highest HR [__|__|__] bpm Lowest HR [__|__|__]bpm Highest SBP [__|__|__] mmHg Worst MAP [__|__|__]/[__|__|__]mmHg Vasopressors: Yes○ No ○ If yes name …… dose …….. µg/kg/min Urine output [__|__|__|__]mL/24hr Acute renal failure Yes○ No ○ Immunocompromised Yes○ No ○

CNS INFECTION(S)	ADM
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

ADMISSION (complete within 24 hours of enrolment)

29. GCS score upon admission to the study?
- ○ GCS 15, no neurological deficiency
- ○ GCS 11- 14 or GCS 15 and neurological deficiency
- ○ GCS <= 10

30. Was the patient BCG vaccinated? (or does the patient have a BCG scar)
○ Yes ○ No ○ Unknown

31. Did the patient receive TB treatment in the past? ○ Yes ○ No ○ Unknown

32. Has the patient ever received IPT? ○ Yes ○ No ○ Unknown

33. Temperature (maximum recorded in last 24hrs) [__|__] . [__] ^{o}C

34. Blood pressure [__|__|__] / [__|__|__] mmHg

35. Weight loss: ○ Yes ○ No ○ Unknown

36. Night sweats ○ Yes ○ No ○ Unknown

37. Cough > 2 weeks ○ Yes ○ No ○ Unknown

38. Date of first neurological symptom (e.g. headache, focal signs, convulsions)
[__|__]/[__|__]/[__|__] (dd/mm/yy)

If date uncertain, estimate number of days of symptoms: [__|__] days

39. Fever	○ Yes	○ No	○ Unknown	Duration [__|__] days
40. Headache:	○ Yes	○ No	○ Unknown	Duration [__|__] days
41. Vomiting:	○ Yes	○ No	○ Unknown	Duration [__|__] days
42. Altered consciousness:	○ Yes	○ No	○ Unknown	Duration [__|__] days
43. Irritability:	○ Yes	○ No	○ Unknown	Duration [__|__] days
44. Lethargy:	○ Yes	○ No	○ Unknown	Duration [__|__] days
45. Neck stiffness:	○ Yes	○ No	○ Unknown	Duration [__|__] days
46. Seizures:	○ Yes	○ No	○ Unknown	Duration [__|__] days
47. Cranial nerve palsy:	○ Yes	○ No	○ Unknown	
48. Hemiplegia/paresis:	○ Yes	○ No	○ Unknown	
49. Paraplegia/paresis:	○ Yes	○ No	○ Unknown	
50. Urinary retention:	○ Yes	○ No		
51. Tetraplegia/paresis:	○ Yes	○ No	○ Unknown	
52. Papilloedema:	○ Yes	○ No	○ Unknown	
53. Machine ventilated:	○ Yes	○ No		
54. Patient sedated:	○ Yes	○ No		

CARDIAC AND LUNG	**ULS**
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

Complete after cardiac and lung ultrasound examination		
55. T1/T2/T3 [___][___]/[___][___]/[___][___] *(dd/mm/yy)* at [___][___]:[___][___] *(use 24 Hr clock)*		
56. Clinical shock c Yes c No		
57. Vasopressors c Yes c No If yes, Name : _____ Dose _____		
58. BP : __ __ __/__ __ mmHg	**59.** Heart rate [__|__|__] bpm	
60. Respiratory distress Yes c No c	**61.** SpO_2 [__|__|%	**62.** FiO_2 [__|__|%
63. Ventilation mode ○ Non-invasive ventilation (CPAP / BiPAP) ○ Intubated / tracheostomy ○ High Flow Nasal Prongs ○ Native airway (excluding High Flow Nasal Prongs) ○ Other: **PEEP (if applicable):** |__|__] cmH_2O		
64. Cardiovascular drugs ☐ Yes ☐ No If yes, Name : _____ Dose _____		
CARDIAC ULTRASOUND		
65. Pericardial effusion ☐ Yes ☐ No If YES, specific view: _____		
66. Left ventricle (LV)		

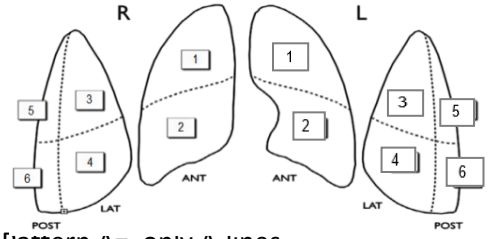
PLAX view		A4C/ A2C view	A5C view
LVIDd [_____]	EF [_____] %	LV volume ed [_____]	LVOT VTI [_____]
LVIDs [_____]	LVOT diameter [_____]	LV volume es [_____]	LV ET [_____]
LVPWd [_____]	AV diameter [_____]	EF (Simpsons') [_____] %	CO [_____]
IVS end-diastole thickness [_____]	LA diameter [_____]		SV [_____]

CARDIAC AND LUNG	**ULS**
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

67. Right ventricle TAPSE [_____] mm

68. IVC IVC max [_____] (cm) IVC min [_____] (cm)

LUNG ULTRASOUND



Pattern A= only A-lines
Pattern B1 = >2 separated B-lines
Pattern B2 = confluent B-lines
Pattern C = Consolidation
Pattern E = Pleural Effusion
UTS = Unable To Score

Record: Pattern| Number of B-lines (if B1)

L1 [___] | [___] R1 [___] | [___]

L2 [___] | [___] R2 [___] | [___]

L3 [___] | [___] R4 [___] | [___]

L4 [___] | [___] R4 [___] | [___]

L5 [___] | [___] R5 [___] | [___]

L6 [___] | [___] R6 [___] | [___]

L base pl. effusion c Yes c No If yes, d=___cm

R base pl. effusion c Yes c No If yes, d=___cm

69. Estimate time to complete ultrasound examination – Training and Validation phase ☐ Yes ☐ No

	Sonographer's estimation	Extract from recorded video
Cardiac ultrasound	[_____] min	[_____] min
Lung ultrasound	[_____] min	[_____] min

Did you do some quantification after examination? ☐ Yes ☐ No

If YES, How long does it take? [_____] minutes

70. Estimate time to complete cardiac ultrasound examination – Testing phase ☐ Yes ☐ No

	Non-expert with software		Expert	
	Sonographer's estimation	Extract from recorded video	Expert's estimation	Extract from recorded video
Cardiac ultrasound	[_____] min	[_____] min	[_____] min	[_____] min

MUSCLE	**ULS**
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

Complete after muscle ultrasound examination

71. T1/T2/ICU discharge
[___][___]/[___][___]/[___][___] *(dd/mm/yy)* at [___][___]:[___][___] *(use 24 Hr clock)*

72. Autonomic nervous system dysfunction (**tetanus only**) ☐ Yes ☐ No

73. Glasgow Coma Score [__|__] /15

74. Sedated ☐ Yes ☐ No

75. Skeletal muscle relaxants (Pipercuronium) ☐ Yes ☐ No

76. Mechanical ventilation ☐ Yes ☐ No If YES, [___][___]/[___][___]/[___][___] *(dd/mm/yy)*

77. Hospital acquired infection ☐ Yes ☐ No If YES, ☐ VAP ☐ UTI ☐ BSI

78. Is the patient doing rehabilitation exercises ☐ Yes ☐ No

If YES, day start exercise: [___][___]/[___][___]/[___][___] *(dd/mm/yy)*

Muscle ultrasound

Rectus femoris muscle X-section area	N1	N2	N3	
Right leg (cm^2)				
Left leg (cm^2)				

79. Estimate time to complete muscle ultrasound examination

	Sonographer's estimation	Extract from recorded video
Muscle ultrasound	[_____] min	[_____] min

DISCHARGE	**DIS**
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

No.	Intervention/Event	Yes	Start time	Stop time	Note
1	Tracheostomy	☐			
2	Mechanical ventilation	☐			
3	ANSD (Tetanus only)	☐			
4	HAIs	☐			
5	Midazolam or Diazepam	☐			
6	Pipercuronum	☐			
7	$MgSO_4$	☐			
8	Steroid	☐			Study code if patient was enrolled to RCT _____
9	Other				

DISCHARGE	DIS
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

1. Date of HTD ICU discharge:	[___][___]/[___][___]/[___][___] (dd/mm/yy)
2. Date of hospital discharge:	[___][___]/[___][___]/[___][___] (dd/mm/yy)
3. Final outcome: ☐ Death ☐ Home/transfer to Die ☐ Hospital transfer ☐ Home	
Confirmation of diagnosis	
4. Sepsis patients only: Microbiological confirmation of causative bacteria If **yes**, specify details: Culture source: Bacteria isolated:	Yes ☐ No ☐ _____ _____
If **no**, specify presumed source of infection: ☐ Respiratory ☐ Central Nervous System ☐ Skin/soft tissue ☐ Bone/joint ☐ Abdominal ☐ Urinary tract ☐ Source unknown	
5. Dengue patients only: NS1 antigen (Place measured: HTD ☐ other hospital ☐) Dengue PCR (Date sent: _ _ / _ _ / _ _) In-hospital dengue IgM (Date sent: _ _ / _ _ / _ _) In-hospital dengue IgG (Date sent: _ _ / _ _ / _ _)	Positive ☐ Negative ☐ Not done ☐ Positive ☐ Negative ☐ Not done ☐ Positive ☐ Negative ☐ Not done ☐ Positive ☐ Negative ☐ Not done ☐
6. CNS infection(s) patients only: **The confirmed diagnosis at hospital discharge** ☐ Encephalitis ☐ Meningitis Laboratory confirmation of causative bacteria/virus/fungus Yes ☐ No ☐ If **yes**, specify details: Culture source: _____ Bacteria/virus/fungus isolated: _____	
Summary of endpoints	
6. Total number of days in ICU	[___][___]days
7. Total number of days admitted to HTD	[___][___]days
8. Total number of days in hospital (if transferred)	[___][___]days
9. Total number of days of mechanical ventilation	[___][___] days
10. Total number of days of non-invasive ventilation	[___][___] days
11. Total number of days requiring vasopressors	[___][___]days
12. Total number of days requiring intravenous fluids	[___][___]days
13. Total volume of intravenous fluid received	[___][___][___][___][___] ml
14. Total number of days requiring haemofiltration	[___][___]days
15. Total dose of midazolam	[___][___]mg
16. Total dose of diazepam	[___][___]mg
17. Total dose of pipercuronium	[___][___]mg
18. Total dose of $MgSO_4$	[___][___]mg
19. Total number of days of ANSD	[___][___]days
20. Total number of days doing rehab exercise in ICU	[___][___]days

SF36	QUE
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

Discharge/Follow-up [___][___]/[___][___]/[___][___] *(dd/mm/yy)* at [___][___]:[___][___] *(use 24 Hr clock) (Vietnamese, answer by patients or their relatives)*

1. Nhìn chung, bạn cảm thấy sức khỏe của mình là				
Tuyệt vời □1	Rất tốt □2	Tốt □3	Hơi kém □4	Kém □5

2. Nhìn chung, so với thời điểm cách đây một năm, bạn đánh giá sức khỏe hiện nay của mình thế nào				
Bây giờ tốt hơn nhiều so với thời điểm cách đây một năm □1	Bây giờ tốt hơn một chút so với thời điểm cách đây một năm □2	Gần giống như thời điểm cách đây một năm □3	Bây giờ kém hơn một chút so với thời điểm cách đây một năm □4	Bây giờ kém hơn nhiều so với thời điểm cách đây một năm □5

3. Sau đây là những câu hỏi việc các sinh hoạt mà có thể bạn sẽ thực hiện trong một ngày binh thường. Sức khỏe hiện tại của bạn có làm hạn chế bạn trong những sinh hoạt này không? Nếu có, mức độ hạn chế là như thế nào.			
	Có, hạn chế nhiều	Có, hạn chế một ít	Không, chẳng hạn chế gì cá
Các hoạt động dùng nhiều sức như chạy, nâng vật nặng, tham gia các môn thể thao minh	□1	□2	□3
Các hoạt động đòi hỏi sức lực vừa phải như di chuyển một cái bàn, quét nhà, bơi lội, hoặc chạy xe đạp	□1	□2	□3
Nâng hoặc mang vác đồ thực phẩm linh tinh	□1	□2	□3
Leo lên vài tầng lầu	□1	□2	□3
Leo lên một tầng lầu	□1	□2	□3

SF36	QUE
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

Uốn người, qùy gối hoặc khom lưng và gập gối	□1	□2	□3
Đi bộ hơn 1km	□1	□2	□3
Đi bộ vài trăm mét	□1	□2	□3
Đi bộ một trăm mét	□1	□2	□3
Tắm rửa hoặc thay quần áo cho chính bạn	□1	□2	□3

4. Trong suốt 4 tuần vừa qua, do ảnh hưởng của sức khỏe thể chất, bạn có thường gặp phải bất kỳ khăn nào sau đây trong công việc hoặc các sinh hoạt thường ngay khác của bạn?

	Có	Không
Làm giảm thời lương bạn tiện hành công việc hoặc sinh hoạt khác	□1	□2
Hoàn thành công việc ít hơn bạn muốn	□1	□2
Bị giới hạn trong một loại công việc nào đó hoặc sinh hoạt	□1	□2
Gặp khó khăn trong việc thực hiện công việc hoặc các sinh hoạt khác – vd phải mất nhiều công sức hơn	□1	□2

5. Trong suốt 4 tuần vừa qua, do ảnh hưởng của yếu tố cảm xúc (chẳng hạn như cảm thấy buồn phiền hoặc lo lắng), bạn có thường gặp phải bất kỳ khăn nào sau đây trong công việc hoặc các sinh hoạt thường ngay khác của bạn?	Có	Không
Làm giảm thời lượng bạn tiện hành công việc hoặc sinh hoạt khác	□1	□2
Hoàn thành công việc ít hơn bạn muốn	□1	□2
Làm việc hoặc tiến hành các sinh hoạt khác kém cẩn thận so với bình thường	□1	□2

6. Trong suốt 4 tuần vừa qua, sức khỏe thể chất hoặc yếu tố cảm xúc có gây trở ngại cho bạn trong các hoạt động xã hội thông thương mà bạn tham gia với gia đình, bạn bể, hàng xom hoặc các nhóm hội không, và ở mức độ nào

Không hề	Một chút	Vừa phải	Hơi nhiều	Rất nhiều
□1	□2	□3	□4	□5

SF36	QUE
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

7. Trong suốt 4 tuần vừa qua, bạn cảm thấy cơ thể đau nhức ở mức độ nào					
Không cảm thấy đau □1	Đau rất nhẹ □2	Đau nhẹ □3	Đau vừa phải □4	Đau trầm trọng □5	Đau rất trầm trọng □6

8. Trong suốt 4 tuần vừa qua, cảm giác đau đớn đã gây trở ngại cho công việc bình thường của bạn ở mức độ nào (bao gôm công việc bên ngoài cũng như việc nội trợ)				
Không hề □1	Một chút □2	Vừa phải □3	Hơi nhiều □4	Rất nhiều □5

9.Những câu hỏi này liên quan đến việc bạn cảm thấy ra sao và mọi việc như thế nào với bạn trong suốt 4 tuần vừa qua. Đối với mỗi câu hỏi, xin vui lòng chọn một câu trả lời đúng với cảm nhận của bạn nhất. Trong suốt 4 tuần vừa qua bạn có thường cảm thấy...						
	Luôn luôn	Rất thường xuyên	Tốt thỉnh thoảng	Thỉnh thoảng	Ít khi	Không bao giờ
Bạn đá từng cảm thấy tràn đầy sinh lực?	□1	□2	□3	□4	□5	□6
Bạn có cảm thấy rất lo lắng?	□1	□2	□3	□4	□5	□6
Bạn có cảm thấy quá đâu buồn và thất vọng đến độ không có gì thể làm bạn vui lên dược?	□1	□2	□3	□4	□5	□6
Bạn có cảm thấy bình tĩnh và thanh thản?	□1	□2	□3	□4	□5	□6
Bạn đã từng cảm thấy dồi dào năng lượng?	□1	□2	□3	□4	□5	□6
Bạn có cảm thấy buốn và nản lòng?	□1	□2	□3	□4	□5	□6
Bạn đã từng cảm thấy kiệt sức?	□1	□2	□3	□4	□5	□6
Bạn có cảm thấy hạnh phúc?	□1	□2	□3	□4	□5	□6
Bạn đã từng cảm thấy mệt mỏi	□1	□2	□3	□4	□5	□6

SF36	**QUE**
Participant code 01NVb [___][___][___] - [___][___][___]	Participant's initial [___][___][___][___][___]

10. Trong suốt 4 tuần vừa qua bạn có thường vì sức khỏe thể chất hoặc các yếu tố cảm xúc của bạn cản trở đến các hoạt động xã hội mà bạn thực hiện – chẳng hạn như đi thăm bạn bè, họ hàng, vv				
Luôn luôn □1	Rất thường xuyên □2	Thỉnh thoảng □3	Ít khi □4	Không bao giờ □5

11. Mỗi nhận xét sau đây có mức độ đúng hay sai như thế nào đối với bạn					
	Hoàn toàn dúng	Hầu như dúng	Không biết	Hầu như sai	Hoàn toàn sai
Đường như tôi hơi dễ bị bệnh hơn những người khác	□1	□2	□3	□4	□5
Tôi khỏe mạnh như bất kì người nào mà tôi biết	□1	□2	□3	□4	□5
Tôi nghĩ răng sức khỏe của tôi sẽ trở nên tệ hơn	□1	□2	□3	□4	□5
Sức khỏe của tôi tuyệt vời	□1	□2	□3	□4	□5

Discharge/Follow-up [___][___]/[___][___]/[___][___] *(dd/mm/yy)* at [___][___]:[___][___] *(use 24 Hr clock)*

Dưới mỗi đề mục, xin đánh dấu vào MỘT ô diễn tả chính xác nhất tình trạng sức khoẻ của anh/chị NGÀY HÔM NAY.

SỰ ĐI LẠI

Tôi đi lại không khó khăn	❑
Tôi đi lại hơi khó khăn	❑
Tôi đi lại khá khó khăn	❑
Tôi đi lại rất khó khăn	❑
Tôi không thể đi lại được	❑

TỰ CHĂM SÓC

Tôi thấy không khó khăn gì khi tự tắm rửa hay khi tự mặc quần áo	❑
Tôi thấy hơi khó khăn khi tự tắm rửa hay khi tự mặc quần áo	❑
Tôi thấy khá khó khăn khi tự tắm rửa hay khi tự mặc quần áo	❑
Tôi thấy rất khó khăn khi tự tắm rửa hay khi tự mặc quần áo	❑
Tôi không thể tự tắm rửa hay không thể tự mặc quần áo	❑

SINH HOẠT THƯỜNG LỆ *(ví dụ: làm việc, học hành, làm việc nhà, các hoạt động trong gia đình, vui chơi giải trí)*

Tôi thấy không khó khăn gì khi thực hiện các sinh hoạt thường lệ của tôi	❑
Tôi thấy hơi khó khăn khi thực hiện các sinh hoạt thường lệ của tôi	❑
Tôi thấy khá khó khăn khi thực hiện các sinh hoạt thường lệ của tôi	❑
Tôi thấy rất khó khăn khi thực hiện các sinh hoạt thường lệ của tôi	❑
Tôi không thể thực hiện các sinh hoạt thường lệ của tôi	❑

ĐAU / KHÓ CHỊU

Tôi không đau hay không khó chịu	❑
Tôi hơi đau hay hơi khó chịu	❑
Tôi khá đau hay khá khó chịu	❑
Tôi rất đau hay rất khó chịu	❑
Tôi cực kỳ đau hay cực kỳ khó chịu	❑

LO LẮNG / U SẦU

Tôi không lo lắng hay không u sầu ❑

Tôi thấy hơi lo lắng hay hơi u sầu một chút ❑

Tôi thấy khá lo lắng hay khá u sầu ❑

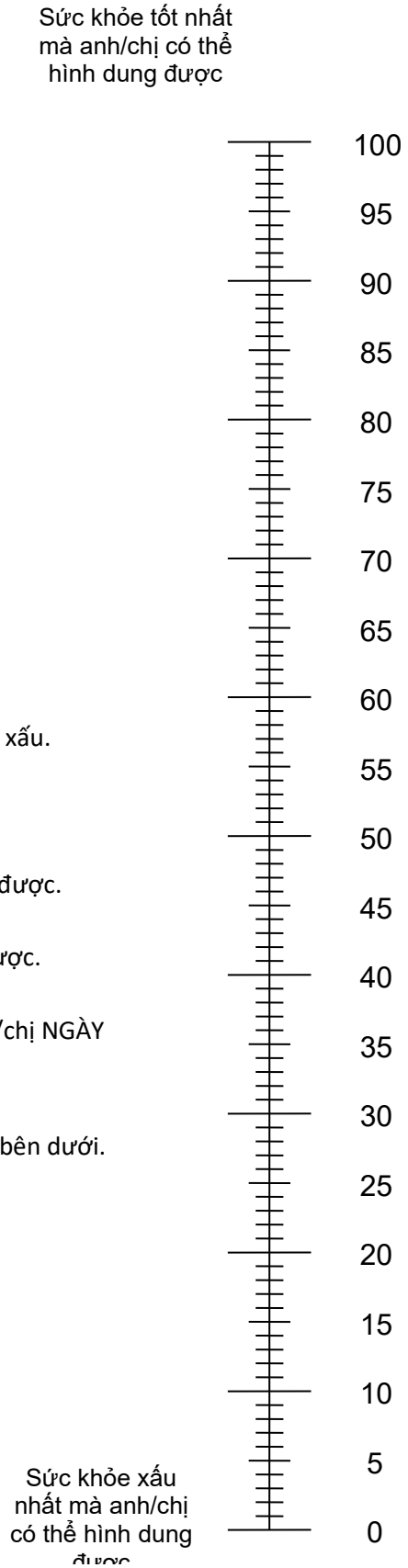
Tôi thấy rất lo lắng hay rất u sầu ❑

Tôi thấy cực kỳ lo lắng hay cực kỳ u sầu ❑

SỨC KHOẺ ANH/CHỊ NGÀY HÔM NAY = ☐

- Chúng tôi muốn biết sức khoẻ của anh/chị NGÀY HÔM NAY tốt hay xấu.
- Thang điểm này được đánh số từ 0 đến 100.
- 100 tương ứng với sức khỏe tốt nhất mà anh/chị có thể hình dung được.

  0 tương ứng với sức khỏe xấu nhất mà anh/chị có thể hình dung được.
- Xin đánh một dấu X trên thang điểm để thể hiện sức khoẻ của anh/chị NGÀY HÔM NAY.
- Bây giờ, xin viết số mà anh/chị đã đánh dấu trên thang điểm vào ô bên dưới.



Appendix C. Design and development of AI-assisted LUS video classification

C1. Dataset

In this study, the dataset was collected from 65 patients with dengue shock or septic shock admitted to the HTD in Ho Chi Minh city, Vietnam between June 2019 and June 2020. The research was approved by the Oxford Tropical Research Ethics Committee (OxTREC) and the HTD Institutional Review Boards. The dataset was split into training, validation and test set as follows: 90% (3079 videos, four seconds each) was used for training and validation and the remaining 10% of the patients (322 videos, four seconds each) were left out for the test set. LUS examinations were carried out using a Sonosite M-Turbo machine (Fujifilm Sonosite, Inc., Bothell, WA) with a low medium frequency (3.5-5 MHz) convex probe by qualified sonographers. LUS was performed using a standardised operating procedure based on the Kigali ARDS protocol [11]: assessment for B-lines, consolidation and pleural effusion, performed at 6 points on each side of the chest (2 anterior, 2 lateral and 2 posterolateral).

These videos were annotated by expert sonographers using the VGG annotator tool. Five lung patterns were selected for multi-class classification. Class imbalance was addressed during training by weighting each class contribution to the loss with the inverse of the number of samples in each class.

AVI-format videos were cropped and masked to remove text and information outside of the scanning sector. The 640x480 pixel videos were downsampled using OpenCV into 64x64 pixels. For training, each four-second clip was converted into shorter clips of one second with an overlap of 20% between consecutive frames in the video.

C2. Model architecture

Our proposed model is an extension of the model from [139] that was used for B-line classification and localization in LUS videos. The model architecture is depicted in Figure C1. It consists of three parts: a convolutional neural network (CNN) to extract frame-wise spatial features, a bidirectional long short-term memory (LSTM) network to extract temporal features from the video and a temporal attention mechanism to increase the weighting of frames that carry more useful information for the classification task. In this paper, we replaced the

classification subnet, after the temporal attention mechanism, with a fully connected layer (with ReLU non-linear activation and dropout) and a 5-element final layer that produces a 1-hot 5D vector for 5 class classification.

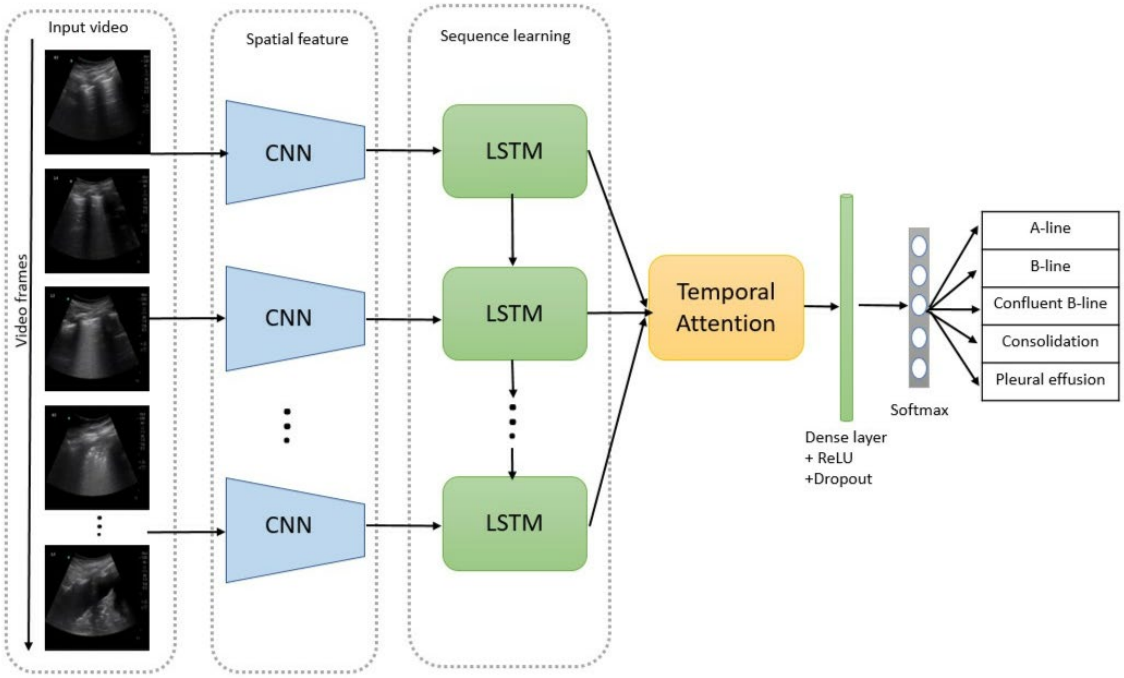


Figure C1. Lung US classification model architecture

C3. Training

The model was implemented in Python 3 using the Keras library with a Tensorflow backend. It was trained using the Adam optimizer with a learning rate of 0.001. A batch size of 16 and batch normalization were applied for both CNN and LSTM network layers. Dropout of 0.2 and L2=10^{-5} for regularization were utilized. These parameters were found to give the best result on the validation set. LUS video data were augmented by adding horizontally-flipped frames to the training set. We used 5-fold cross validation and trained the network for 100 epochs. The trained models were evaluated on the independent test set.

C4. Real-time AI-assisted LUS framework (RAILUS) system standard operating procedure

Figure C2. Real-time AI-assisted LUS framework (RAILUS)

The ultrasound machine HDMI output was connected with the laptop via a USB framegrabber. Instead of looking at the ultrasound machine screen, the user can look at the laptop's screen. The code for RAILUS is made publicly available on a github repository (https://github.com/vital-ultrasound/lung/tree/main/PRETUS_Plugins).

On the left of the screen is the video resolution and scale. Depending on the video output of the ultrasound machine (e.g., 1280x1024, 1920x1080), the user can choose the right setting. On the right of the screen is the widget of the RAILUS software, where the model provides continuous

real-time prediction (as the green bar) of the class of corresponding clips. The user can enable or disable the model prediction. The clinician interacts with the laptop using the mouse to select the lung zones (12 lung zones) and labels. After choosing the label and submitting, the lung zone in the lung diagram changes color (e.g. from green-normal lung to yellow-pleural effusion) according to the class of the label. For the purpose of testing and comparison between clinicians with the AI tool and without the AI tool, the bottom right shows the user prediction, model's prediction, user's confidence and total time taken for interpretation. When the user finishes the lung ultrasound examination, the tool will generate the result in the text file and the corresponding clips will be saved to the same folder.

Appendix D. Design and development of AI-assisted RF muscle ultrasound segmentation

D1. Dataset

The dataset used to train the model included 600 ultrasound images from 112 patients diagnosed with severe tetanus and central nervous infection at HTD. We randomly split the dataset into 80% for training, 10% for validation and 10% for testing. The muscle scans and manual annotations were done by three radiologists and contouring of the RF was performed using the built-in tracing feature of the GE Venue Go, GE Vivid IQ (GE) and handheld Phillips Lumify ultrasound machines. The characteristics of the patients are shown in Table D1.

Table D1. Characteristics of patients in the training dataset (N = 112)

	N = 112
Age	56 (42, 64)
Sex (female)	25 (22%)
Diagnosis	
Tetanus	78 (70%)
Central Nervous System Infection	34 (30%)
Comorbidities (1 or more)	67 (75%)
Sedative	112 (100%)
Muscle relaxion	112 (100%)
Length of ICU stay	24 (17, 33)
Length of hospital stay	30 (23, 42)
MV days	18 (7, 26)
Nutrition support	112 (100%)
RF CSA D1 (cm^2)	4.70 (3.16, 6.32)

RF CSA D7 (cm^2)	4.45 (3.05, 6.07)
RFCSA Discharge (cm^2)	3.67 (2.48, 4.97)
Muscle loss during ICU (%)	17 (6, 32)
Outcomes	
Home	91 (81%)
Hospital Transfer	100 (8.9%)
Transfer To Die	11 (9.8%)

D2. Model architecture

We employed a U-net architecture [65] for the RFCSA semantic segmentation task (Figure D1) as it has been shown to perform well for image segmentation in several medical imaging modalities even if the datasets were small. The main architecture consists of contracting and expanding paths. The contracting path (or encoder) consists of convolution and max pooling layers whereas the expanding path (or decoder) uses transposed convolutional layers. Skip connections were employed between layers of the same resolution in the two paths. The model outputs a pixelwise binary label (background and RFCSA). The input images for the model were resized to 256 x 256 pixels. Augmentation was applied on-the-fly during training including rotation, horizontal flipping, zoom in/out and increase/decrease contrast gain.

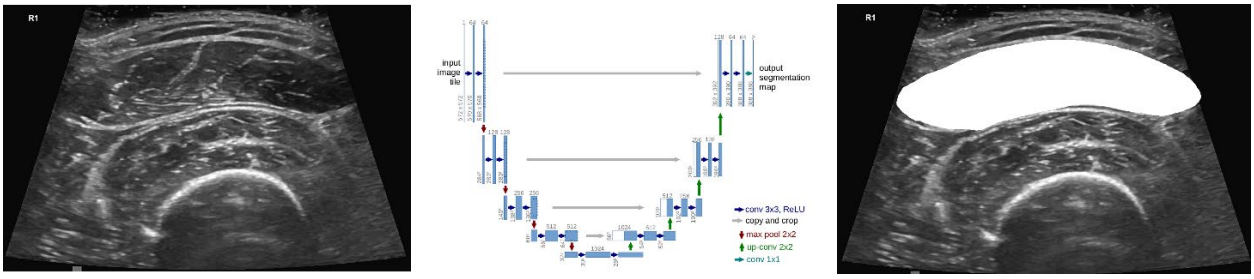


Figure D1. Muscle US segmentation model architecture

D3. Training

The model was implemented in Python 3 using a PyTorch backend. It was trained using the Adam optimizer with a learning rate of 0.001. A batch size of 16 and batch normalization were applied for both CNN and transposed CNN layers. Muscle ultrasound data were augmented using horizontal flipping and random rotations of ±10 degrees to the training set. The models were trained for 100 epochs and evaluated on the independent test set.

D4. Image pre- and post-processing

In pre-processing, input images were resized to 128 x 128 pixels, by first resampling to isotropic pixel size (using the smallest pixel spacing as reference), followed by a padding operation to make the image square, a centered crop operation to keep the ultrasound sector in the middle and finally a resize to 128 x 128. These operations were undone in post-processing on the output segmentation mask to be able to overlay the muscle shape onto the original B-mode image.

After the pre-processing described above, input images were normalized to the range [0, 1] by dividing by the maximum feasible intensity value, 255. Output masks are in the range [0, 1] so they were scaled up to [0, 255] after postprocessing.

Appendix E. Design and development of AI-assisted cardiac US A4C Left Ventricular segmentation

E1. Dataset

The model was trained using publicly available datasets including Echonet-Dynamic and CAMUS datasets for model development. The full description and annotation protocol of Echonet-dynamic [128] can be found and downloaded in https://echonet.github.io/dynamic/ and CAMUS [127] in https://www.creatis.insa-lyon.fr/Challenge/camus/databases.html. The former dataset has 2D A4C and two-chamber view sequences of 500 patients. For each sequence, the manual annotation for the ED and ES frames of the left ventricle structures - endocardium, epicardium, and left atrium are provided as the ground-truth for 450 patients. The Echonet-dynamic dataset consists of 10030 different A4C echocardiography videos with corresponding number of ED and ES frames. For each video, two tracings from experts are provided of both the ED and ES. The US images for ED and ES stages are extracted from the video and frame information, and the ground truth is created from the expert tracings of the left ventricle.

E2. Model architecture

We also employed a U-net architecture [65] for the LV segmentation task similar to the model architecture described in section D2.

E3. Training

The model was implemented in Python 3 using a PyTorch backend. It was trained using the Adam optimizer with a learning rate of 0.05. A batch size of 32 and batch normalization were applied for both CNN and transposed CNN layers. Echo data were augmented using horizontal flipping, random rotations of ±10 degrees, random translations of ±20% both horizontally and vertically, and random scaling from 80% to 120% to the training set. The models were trained for 200 epochs and evaluated on the independent test set.

E4. Image pre- and post-processing

In pre-processing, input images were resized to 128 x 128 pixels, by first resampling to isotropic pixel size (using the smallest pixel spacing as reference), followed by a padding operation to make the image square, a centered crop operation to keep the ultrasound sector in the middle and

finally a resize to 128 x 128. These operations were undone in post-processing on the output segmentation mask to be able to overlay the muscle shape onto the original B-mode image.

After the pre-processing described above, input images were normalized to the range [0, 1] by dividing by the maximum feasible intensity value, 255. Output masks are in the range [0, 1] so they were scaled up to [0, 255] after postprocessing.

Appendix F. Usability questionnaire

Part 1 – Demographics (10 questions)

Level of training	1. Ultrasound doctor 2. Intensivist 3. General Doctor 4. Resident 5. Student 6. Other
What is your experience in any ultrasound modality?	1. Beginner 2. Intermediate 3. Advanced 4. Expert
What kind of ultrasound you have used and currently using?	1. Cardiac 2. Abdominal 3. Blood vessel 4. Lung ultrasound 5. Intervention guidance 6. Other
What is your experience in lung ultrasound	1. Beginner 2. Intermediate 3. Advanced 4. Expert
Is lung ultrasound available in your unit/apartment	1. Yes 2. No
What do you think are the main challenges in doing lung ultrasound	1. Finding the right view 2. Image interpretation 3. Image quantification 4. Identify landmark 5. Instruction for user 6. Physical environment 7. Writing report 8. Other
What kind of ultrasound machine you are currently using?	1. Handheld devices (tablet, phone-based devices) 2. Point-of-care (ICU, ED) 3. “Full feature-trolley” machine

Part 2 – Lung ultrasound video interpretation (10 standard videos and 10 videos with AI tool)

All videos of the test can be found in this link: video_test_LUS

Part 3 – Usability (20 questions)

The answers of the questionnaire were valued by five–point Likert scale. The questionnaire was developed and delivered by Google Forms. Because the interaction between AI-based tools and clinicians is poorly understood, the survey aimed at investigating how our LUS tool should be

deployed in clinical settings, the clinician's trust in the tool, and the level of concern about its utilisation.

Usability questionnaire	**[1] Strongly disagree -> [5] Strongly agree**				
	1	2	3	4	5
1. I think that I would like to use this tool frequently					
2. I found the tool unnecessarily complex					
3. I thought the tool was easy to use					
4. I think that I would need the support of a technical person to be able to use this tool					
5. I found the various functions in this tool were well integrated					
6.I thought there was too much inconsistency in this tool					
7. I would imagine that most people would learn to use this tool very quickly					
8. I found the tool very cumbersome to use					
9. I felt very confident using the tool					
10. I needed to learn a lot of things before I could get going with the tool					
11. I feel that such a tool may improve the differential diagnostic process in the ICU					
12. To what extent do you agree with the following statement: "Seeing the model prediction bar will improve your confidence in interpret lung ultrasound videos"					
13. To what extent do you agree with the following statement: "The model prediction bar distract me from interpret lung ultrasound videos"					
14. To what extent do you agree with the following statement: "Artificial intelligence assisted tool can help improving quality of lung ultrasound"					
15. Have you used other AI-assisted tools before? If yes, is it useful? • Yes • No					
16. When do you think the tool will be most useful? • Real-time (when carrying out the exam) • Offline (collect the video and interpret afterward) • Both					
17. What level of error is acceptable for AI-assisted LUS interpretation in that are used for the patients? • Equivalent to the worst performing advanced user • Equivalent to the average performing advanced user • Superior to the average performing advanced user • Equivalent to the best performing advanced user • Superior to the best performing advanced user					

18. Would you consider using the following clinical workflow? Patients clinical images undergo AI-assisted LUS interpretation tool. A specialist subsequently reviews both the image and the artificial intelligence findings.

- Yes
- No

19. Which of the following do you perceive as the greatest potential advantage of the use AI-assisted LUS interpretation tool? (rank the top 3 preferences where 1=greatest advantage)

- Assist non-expert clinician to carry out acceptable quality lung ultrasound
- Improved patient access to disease screening
- Improved diagnostic confidence
- Reduced time spent by specialists on monotonous tasks
- Greater uniformity in diagnosis and management decisions
- Improved prediction of disease outcomes
- People with less expertise can collect acceptable quality lung ultrasound examination
- Other

20. Which of the following do you perceive as the concern to the utilisation of AI-assisted LUS interpretation tool? (rank the top 3 preferences where 1=greatest concerns)

- Concerns over the divestment of health care to large technology and data companies
- Data security & privacy concerns
- Concerns over medical liability due to machine error
- Lack of confidence or trust in 'black-box' diagnosis
- Decreasing reliance on medical specialists for diagnosis and treatment advice
- Challenge to the fiduciary relationship between patient and doctor
- Concerns over benchmarking clinicians against machines
- Impact on workforce needs
- Other