

Machine Learning in Point of Care Ultrasound

Momodou L. Sonko, BS; T. Campbell Arnold, BS; Ivan A. Kuznetsov, BS

The Perelman School of Medicine, The University of Pennsylvania

What are Machine learning and Deep Learning?

When a patient presents to the ED, clinicians often turn to medical imaging to better understand their condition. Traditionally, imaging is collected from the patient and interpreted by a radiologist remotely. However, scanning devices are increasingly equipped with analytical software that can provide quantitative assessments at the patient's bedside. These assessments often rely on machine learning algorithms as a means of interpreting medical images.

A machine learning (ML) algorithm is able to utilize presented data to adapt and learn without following explicit instructions. ML is a branch of artificial intelligence (AI) and has garnered a great deal of attention over the past decade, due in large part to substantial advancements in data processing and improvements in model performance. ML has proven to be a powerful method for interpreting complex data. Clinicians may understand all the information necessary to classify a patient's condition, but seldom can they derive an equation that communicates precisely what information is relevant and irrelevant. ML excels at this task and permits scientists to develop solutions without knowing how to explicitly code the answer. In the most common form of ML, called supervised learning, scientists provide data inputs (called features) and corresponding class labels. The machine learning algorithm then determines what input features are relevant to predict the class labels, thus generating a model that can take in novel features and provide a predicted class label as output.

One of the most successful methods for solving medical imaging problems is a subfield of machine learning called deep learning (DL). Deep learning was inspired by the complex neural architecture of the human brain, which is organized into interconnected layers of neurons and can solve incredibly complex problems. In the primate visual cortex, simple photoreceptor input is passed through convolutional layers in the ventral visual stream of the brain. Each successive layer produces increasingly complex representations of the photoreceptor input, which permits humans to classify the objects and interpret the scenes they see. Similarly, deep learning algorithms simulate the ventral visual stream by passing image information through multiple layers of a

convolutional neural network (CNN). These networks process simple pixel information, form new complex representations, and pass those representations on to subsequent layers for eventual image classification [1].

To train a CNN to classify images, pixel values are passed into the initial layer of neurons which are activated by the information (Figure 1). The activations are then fed forward into additional hidden layers which further process the data. The features generated by this process are fed through a final activation function, which provides a classification label in the output layer. To improve model accuracy, algorithm predictions are compared to provided labels. A cost function assesses the difference between model predictions and actual values, awarding a proportional penalty to the model. The goal of the training process is to minimize the cost or penalty awarded to the model. Using an optimization technique called gradient descent, the response weight of each individual neuron in the network is iteratively tuned such that the final classifications better match the expected output, thus reducing the penalty assessed by the cost function. After training on a corpus of images, a novel image can be fed into the algorithm and a predicted classification will be output.

Over the past several years there have been huge advances in the use of ML and DL algorithms to address a number of challenges clinically. DL algorithms have been extensively used within the field of radiology where they are used to perform numerous tasks, including segmentation of anatomical structures or local lesions, detection of probable tumors, and classification of lung and breast nodules [2,3]. A prominent example of how DL is rapidly changing the field of radiology can be seen in chest X-ray advancements. In 2017, with the release of the world's largest publicly available chest X-ray dataset (over 100,000 frontal-view X-ray images) by Stanford and the NIH [4], P. Rajpurkar et al. developed a DL system called CheXNet that could automatically detect and classify 14 different diseases on chest X-ray [5,6]. While there are some concerns surrounding the validity of human to algorithm comparisons [7], their system was able to achieve a comparable detection rate to expert radiologists for most diseases and demonstrated the promise of DL systems within medical imaging [8].

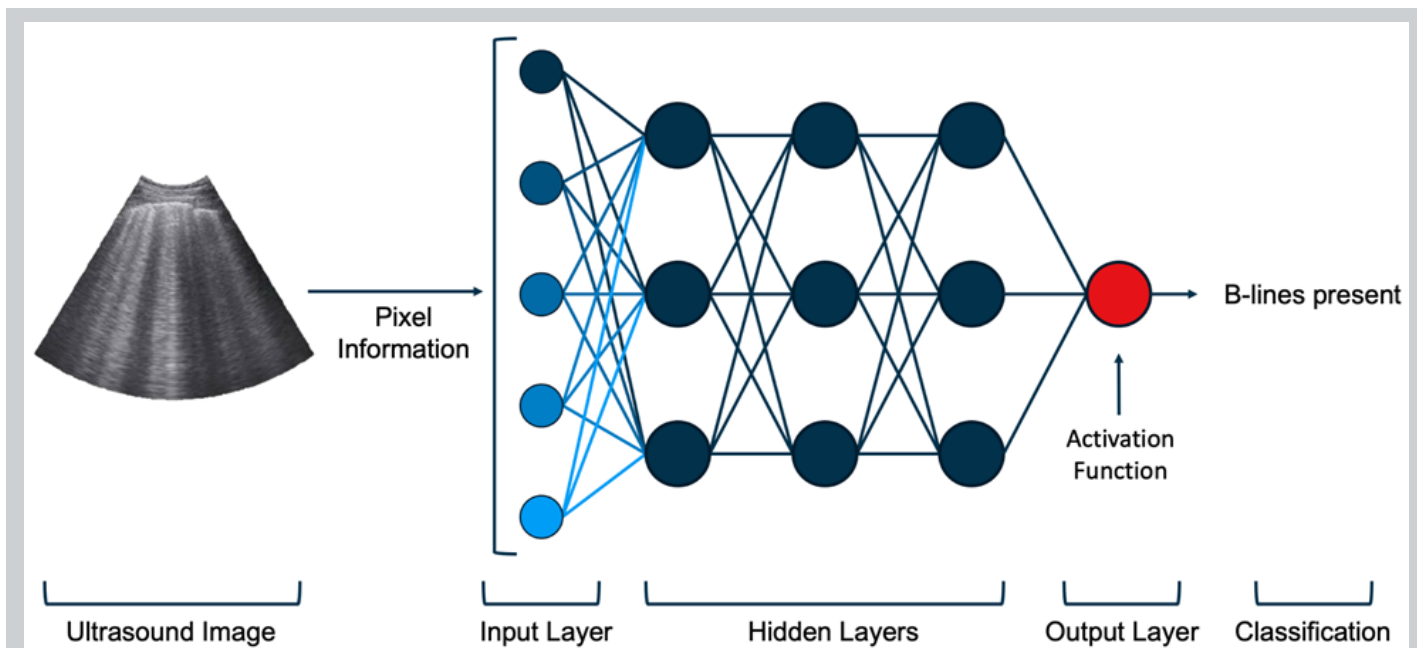


Figure 1. Example convolutional neural network for ultrasound. Here we present a simplified network architecture for a model designed to detect the presence or absence of B-lines in a lung ultrasound image. The model takes in raw data, in this case pixels from the lung ultrasound image. Neurons in the input layer, represented here by circles, are activated to varying degrees by the pixels. These activations are passed forward through several convolutions in the hidden layers. Final activation values are fed into the activation function in the output layer. The output is a binary classification: either presence or absence of B-lines.

Challenges to greater adoption of DL in POCUS

The example above demonstrates the potential of ML systems to improve clinical care for patients as well as assist radiologists with their clinical workload. However, despite rapid advancements in many medical imaging modalities, similar applications of ML algorithms to point of care ultrasound (POCUS) have been slower to arrive. This discrepancy is present for a number of reasons. First, unlike POCUS imaging modalities such as chest X-ray, CT, and MRI, have standardized imaging protocols. Hospital image archiving infrastructure was designed to store and save imaging data for later use. As a result of the persistent imaging infrastructure for these modalities large, organized imaging datasets have been developed that can be more readily interrogated by DL algorithms.

In contrast, images and video acquired at the bedside using POCUS are often used for immediate physician support and not always permanently archived for later analysis. Additionally, the point of care setting inherently introduces variability in data quality even when collected by the same sonographer. Variation in sonographer skill level, image acquisition order, and technique further complicates ultrasound datasets. Even in well-performed scans, imaging distortion and artifacts are often an inescapable reality for POCUS. This results in ultrasound images containing a great deal of “noise” or randomness in the data. Variability is further compounded when

combining images from different scanner manufacturers or academic centers into a single dataset. Additionally, ultrasound images often lack global reference structures, making it difficult to determine exactly where on a patient’s body an image was collected. Finally, as an imaging modality, POCUS is relatively new compared to chest X-ray, CT, and MRI, only achieving widespread use in hospitals in the 1990s [9]. Taken together, these reasons explain why there are relatively few DL applications for POCUS compared to other imaging modalities.

Nonetheless, the last few years have seen an explosion of novel DL applications within POCUS. DL is uniquely suited for analysis of POCUS because it is able to generate high-level abstractions from a wide array of raw imaging data of varying quality. This ability to “cut through the noise” and draw abstractions and note otherwise missed patterns has been one factor leading to greater use of DL within POCUS. Increased interest in DL has come in part due to unique computational approaches to address the obstacles previously mentioned. Both traditional machine learning techniques (i.e. random Forest classifiers, support vector machines) and deep learning methods (i.e recurrent neural networks (RNN), auto-encoders), have been employed on ultrasound datasets with good success. Additionally, researchers have utilized innovative techniques such as transfer

learning to circumvent some issues related to limited and inconsistent datasets. Transfer learning is the process of initializing a DL model with weights derived from another training task and fine-tuning the model to perform a new task with the goal of reducing the number of trials necessary to learn a similar task [10,11]. For instance, a model trained to accurately identify and segment straight lines may be retrained on a carotid ultrasound dataset in order to identify and segment the arterial wall. This approach has the benefit of generally requiring fewer class labels in the training set in order to develop a successful algorithm.

DL algorithms have the potential to further increase the utility and adoption of POCUS. Many important uses of ML applied to POCUS are outside the scope of this review, but also include DL algorithms applied to enable novice sonographers in acquiring the best image [12] and as educational tools for medical students and residents providing procedural training on needle guidance for epidural anesthesia [13]. Herein we will discuss potential and emerging clinical uses of ML approaches applied to POCUS.

Current Clinical Applications of ML within Ultrasound

Here we will briefly highlight some of the clinical ML algorithms that have been developed for POCUS. To date, there are relatively few real-time ML algorithms available in POCUS (Table 1). With the exception of a few commercially available models, the majority of ML algorithms were developed for US applications using datasets captured from retrospective studies. One barrier to greater adoption of ML models within POCUS is the need to implement software in real-time on the ultrasound device hardware itself. Adoption of ML in POCUS is challenging not only because software and hardware must be integrated to enable real time applications, but also because most recently developed ML models lack sufficient clinical validation and FDA approval to be used in the clinical setting. While this regulatory milestone may seem distant, it has already been achieved for similar medical imaging applications. In 2018 the FDA approved a retinal imaging device with onboard artificial intelligence that could make diagnostic decisions, a first of its kind innovation [14]. As researchers continue to develop ML models for ultrasound, it is important to note that given adequate implementation, many of these models can be adapted for POCUS devices in the near future.

As mentioned before, the majority of new ML algorithms within US have been applied using DL architectures. Some of the most significant advances in DL applied to POCUS have taken place within echocardiography. Here, a number of models have been developed for a wide number of classification, segmentation, and detection tasks. A frequently used DL application involving both

segmentation and biometric measurements has been the rapid determination of cardiac ejection fraction (EF). In order to accurately determine EF using echocardiography, determination of cardiac cycles—namely, end-diastole and end-systole—is necessary. Some groups such as Dezaki et al. have successfully used ML models to accurately determine cardiac cycles [15] and a number of other groups have also successfully trained DL models to segment various chambers of the heart using recurrent neural networks (RNN) and CNNs [16–18]. Furthermore, the automation of EF and cardiac volumes using ML has been shown to have excellent agreement between automated and manual approaches, with increased efficiency and reproducibility of measurements [19,20].

There has also been significant interest in applying ML algorithms to lung ultrasound. Lung ultrasound has gained increased use in the POC setting due to the wide number of clinically useful assessments it provides [21]. The quantitative assessment of B-line score (BLS) has become an important tool for assessing pulmonary congestion using POCUS [22]. B-lines are hyperechoic reverberation artifacts arising from the pleural surface that extend to the bottom of the screen without fading and move in tandem with lung sliding. Total BLS can be used to determine fluid overload (FO) severity score and a number of studies have demonstrated that BLS accurately quantifies pulmonary congestion outperforming the physical exam and chest x-ray [23–25]. Additionally, in the point of care setting, rapid assessment of a patient's volume status can be a crucial tool in guiding clinical interventions. Yet, widespread use of this technique is limited partly due to the tedious nature of the assessment.

A number of groups, including our own, have developed DL models using CNN to automatically quantify B-line scores from POC lung ultrasound video clips. Recently, B. Christiana et al. developed a supervised CNN trained on 400 lung ultrasound clips to calculate total BLS in emergency department patients. They achieved a binary classification (B-lines present versus absent) sensitivity and specificity of 93% and 96% compared to an expert interpreter. In multiclass classification of B-line severity their DL model achieved a linear weighted kappa of 0.65 vs an interrater reliability of 0.87 [26]. Our own group has developed a DL model that uses a transformer block architecture CNN trained on 91 hemodialysis patients with ESRD to calculate total BLS and severity level. In preliminary results, our DL model demonstrated a binary classification (presence versus absence of B-lines) accuracy of 87.3% and a total BLS classification (scored 0-4) accuracy of 60.5% [27].

Point of care lung ultrasound has also shown great

Table 1. Discussed studies on the application of machine learning in ultrasound. Due to differences in study design, inter-study performance cannot be compared. Refer to original studies for details of study design. Acronyms: ED - end diastolic; ES - end systolic; RNN - recurrent neural network; LSTM - long short-term memory network; CNN - convolutional neural network; DSC - Dice score coefficient; EF - ejection fraction; LV - left ventricle; FASP - fetal abdominal standard plane; FFASP - fetal face axial standard plane; FFVSP - fetal four-chamber view standard plane; CKD - chronic kidney disease; DVT - deep venous thrombosis; ICC - interclass correlation coefficient; CI - confidence interval. (con't next page...)

Authors	Application	Model	Performance
Dezaki et al.	Cardiac: cycle phase determination (ED vs ES)	Residual RNNs (ResNet + LSTM)	R^2 score = 0.66 Error ED = 3.7 Error ES = 4.1
Smistad et al.	Cardiac: LV segmentation	CNN (U-Net [64])	DSC = 0.86 ± 0.06
S. Chen et al.	Cardiac: plane detection, LV detection, LV segmentation	PSPNet + temporal affine network (TAN)	DSC = 0.91
Knackstedt et al.	Cardiac: EF	AutoLV	ICC = 0.70-0.83
Thavendiranathan et al.	Cardiac: LV volume & EF	Probabilistic contouring algorithm composed of a Bayesian framework, hierarchical K-means clustering, and probabilistic boosting tree.	Correlation with cardiac magnetic resonance measurements: ED volume = 0.90 ES volume = 0.96 EF = 0.98
F. Dominika et al.	Cardiac: EF	LVivo EF (DiA Imaging Analysis)	Performance relative to calculated EF via 3D echocardiography: Pearson correlation = 0.92 (95% CI 0.87-0.95) Mean difference = 0.61% (95% CI -0.68-1.89%)
Nafee et al.	Extremities: DVT	Ensemble classifier	Concordance statistic = 0.69
Tanno et al.	Extremities: DVT	CNN	F1 score = 90%
H. Chen et al.	Fetal: plane detection (FASP, FFASP, FFVSP)	Transferred RNN (T-RNN): CNN + LSTM	Accuracy: FASP = 0.91 FFASP = 0.87 FFVSP = 0.87
Jang et al.	Fetal: abdominal circumference	CNN + U-Net	Accuracy = 87.1%
Gao et al.	Fetal: anatomy classification	T-CNN	Accuracy = 91.5%
Ravishankar et al.	Kidney: segmentation	Ensemble classifier with gradient boosting	DSC = 0.83
Wu et al.	Kidney: segmentation	Cascaded DenseNet	Mean intersection over union = 0.83
C. Chen et al.	Kidney: CKD detection	Support vector machines	5 stages: Accuracy = 70%
Kuo et al.	Kidney: CKD detection	ResNet	5 stages: Accuracy = 85.6%

Table 1 (con't...). Discussed studies on the application of machine learning in ultrasound. Due to differences in study design, inter-study performance cannot be compared. Refer to original studies for details of study design. Acronyms: ED - end diastolic; ES - end systolic; RNN - recurrent neural network; LSTM - long short-term memory network; CNN - convolutional neural network; DSC - Dice score coefficient; EF - ejection fraction; LV - left ventricle; FASP - fetal abdominal standard plane; FFASP - fetal face axial standard plane; FFVSP - fetal four-chamber view standard plane; CKD - chronic kidney disease; DVT - deep venous thrombosis; ICC - interclass correlation coefficient; CI - confidence interval.

Authors	Application	Model	Performance
Christiana et al.	Lung: B-line score	Custom shallow CNN (10 layers)	Presence vs absence: Sensitivity = 93% Specificity = 96% Severity (0-4): Kappa = 0.65
Sonko et al.	Lung: B-line score	Autoencoder + CNN	Presence vs absence: Accuracy = 87.3% Score (0-4): Accuracy = 60.5%
Correa et al.	Lung: pneumonia	Custom feedforward neural network (3 layers)	Sensitivity = 91% Specificity = 100%
Born et al.	Lung: COVID detection	CNN (VGG-16)	Sensitivity = 0.96 Specificity = 0.79
J. Short et al.	Lung: auto mated B-line counting	Auto B-lines (GE Healthcare Venue Go)	Correlation with expert interpretation: ICC = 0.794 (95% CI 0.736-0.840)

promise in the accurate diagnosing of community acquired pneumonia (CAP). It has demonstrated excellent diagnostic capabilities when performed by a trained sonographer compared to both clinical assessment and chest X-ray, while also avoiding unnecessary radiation exposure in vulnerable patients such as pediatric populations [28–30]. However, lung ultrasound is not included in the diagnostic workup for CAP partly because of inter-operator variability of lung ultrasound and training. DL approaches aimed at reducing these barriers are another emerging trend. In 2018, Correa et al. developed a neural network trained on 1450 CAP-positive ultrasound frames from a hospitalized pediatric population in Peru. The algorithm was successful in correctly identifying pneumonia infiltrates with 90.9% sensitivity and 100% specificity [31].

More recently, amidst the 2020 COVID-19 pandemic, there has been increased attention paid to increasing the diagnostic ability of clinicians to detect the presence of the novel coronavirus in patients. To this end, Born et al. developed a deep CNN trained on over 1100 COVID-19 confirmed lung ultrasound images to achieve a detection sensitivity of 0.96 and specificity of 0.79 and F1-score of 0.92 in a 5-fold cross validation [32]. The authors provide

an open-access web service (POCOVIDScreen) that deploys the predictive model, allowing clinicians to both perform predictions on ultrasound lung images and upload their captured images to add to the database.

Use of POCUS within nephrology has also increased in use over the past several years. In patients with chronic kidney disease (CKD), volume overload plays an important role in the disease pathology by complicating cardiovascular pathophysiology leading to increased cardiovascular morbidity and overall mortality [33,34]. For patients with end-stage renal disease (ESRD) on hemodialysis, it has also been shown that the extent of volume overload correlates with adverse cardiovascular events [35]. Therefore, for the nephrologist, close monitoring of their patient's overall volume status is important in the clinical management of patients. Thus, POC lung ultrasound (and BLS quantification) has also become an important tool in the nephrologist's arsenal. However, other ML advancements within renal ultrasound include the accurate segmentation of the kidneys, yielding rapid and accurate measurement of renal dimensions in patients by groups such as Ravishankar et al. and Wu et al. [36,37], as well as the detection of various stages of CKD by groups such as Chen et al. and

Kuo et al. [38,39].

Fetal measurement is another widely used application in POCUS. In the emergency room, rapid and accurate assessment of fetal parameters such as crown-rump length and classification of the abdominal standard plane, are important to avoid misdiagnosis and guide appropriate interventions [40]. Several groups have developed deep learning models for fetal exams. In a two-step process, Jang et al. first developed a CNN to identify the abdominal standard plane and then trained a model to segment and estimate fetal abdominal circumference from fetal ultrasound images [41,42]. Gao et al. developed a CNN that categorized abdominal freehand sweep images into four categories: fetal abdomen, heart, skull, or other. They trained two models, one using only obstetric ultrasound images and a second that employed transfer learning, using a pretrained ImageNet model and fine-tuning it on obstetric ultrasound images. Transfer learning improved classification accuracy in all categories of fetal anatomical structures compared to their non-transfer learning approach [43].

The final clinical application of ML for POCUS discussed here is deep-venous thrombosis (DVT) screening. POCUS is an important tool for physicians treating potential DVT patients within the emergency room as well as in the inpatient setting. POCUS can guide clinical decision-making for patients at risk for, or suspected of having, a pulmonary embolism [44]. During the exam, the deep veins of the lower extremity are compressed along their course and areas of low compressibility suggest potential thrombus formation at that location. Recently, Nafee et al. sought to evaluate the performance of two ML models they developed versus a validated DVT scoring system in acutely ill patients. Their study demonstrated that both of their ML approaches outperformed the validated manual scoring system in predicting venous thromboembolism (VTE) (c-statistic: ML methods = 0.69 and 0.68, manual scoring system = 0.59) [45].

Other models such as that by Tanno et al. have aimed to increase classification accuracy of DVT scans by automatically detecting the extent of vein compressibility in DVT scans [46]. Researchers proposed a dual-task CNN to predict vein compressibility with an F1 score of 90% when evaluated on 1150 5–10 s compression image sequences from 115 healthy volunteers resulting in a data set size of approximately 200k labelled images. As further development continues, these advancements may greatly increase the accessibility and clinical usage of this already impactful diagnostic study.

Commercially Available products utilizing ML

Of note, companies such as Mindray and GE have

utilized ML and DL based algorithms in commercially available echocardiography products to perform automated tasks such as automated EF calculation, LV border identification, and chamber length calculations (Mindray North America, Mahwah NJ; GE Healthcare, Chicago IL). Newer devices entering the market are now often branded with “AI enabled” capabilities, such as left ventricular outflow tract (LVOT) plane identification and Doppler placement (see: GE Venue and Mindray). These tools are beginning to make their way into newer POCUS devices as well. Companies like Butterfly Network, Bay Labs, and Clarius have released POCUS probes that contain AI-enabled cardiac algorithms for automated EF estimation as well as cardiac chamber segmentation (i.e. Butterfly Network’s IQ probe).

There is also commercial interest and new adoption of automated B-line counting algorithms within POCUS. Notably, GE has incorporated an auto B-line counter within their new suite of GE Venue Go POCUS devices. Their model uses computer vision and DL approaches, including a proprietary CNN, to automatically detect and count B-lines in lung ultrasound scans. A study by J. Short et al. found that automatic counting of lung B-lines was consistent with visual counting, as performed by experts in the field and both systems showed a high intra- and interobserver reliability [47]. Other device manufacturers such as Mindray have similarly developed their own automatic B-line counting algorithms using a mixture of traditional computer vision systems and DL approaches.

Interestingly, the clinical ML software market has grown to now support firms whose business models almost entirely center around developing novel algorithms for clinical use intended for device manufacturers. DiA Imaging Analysis Ltd. is one notable firm in this category. They partner with ultrasound device manufacturers and large academic medical centers to develop AI-enabled solutions for ultrasound. Currently, as mentioned previously, much of the development for these solutions has focused on POC echocardiography, but additional interest has been shown in the development of AI-enabled abdominal algorithms as well. The company has additionally partnered with GE to offer the first AI-based solution for automated EF analysis on handheld ultrasound through the “LVivo EF” on GE’s Vscan Extend, which has been shown to yield similar EF values as 3D echocardiography [48]. As interest in DL applications within POCUS continues to blossom, it is likely that additional firms similar to DiA will emerge, outsourcing much of the ML innovation once developed in-house by ultrasound device manufacturers to specialized image analysis companies.

Future Steps & Upcoming Advancements

The dynamic and real-time nature of POCUS provides a major advantage over other imaging modalities such as CT and MRI. Yet, this also represents a major challenge for researchers developing ML algorithms for POCUS. A trained sonographer will rarely examine a single image frame to make a clinical assessment of a patient; rather, data from multiple frames are assessed simultaneously together to inform the clinician of a proper course of action. Within the broader context of deep learning, a known issue is that most state-of-the-art architectures are optimized for single image classification and that impressive performance does not necessarily generalize to video-type data, such as POCUS.

A variety of methods have been applied to try to generalize methods used for image classification to video classification. Perhaps the most direct implementation of this has been the use of 3D CNNs (as opposed to the 2D ones used for single image classification). For example, Hara et al. extended the state-of-the-art ResNet architecture to 3D by adjusting the original 3x3 kernels to 3x3x3 [49]. However, introducing 3D convolutions leads to significantly increased computational overhead and increases network complexity, hence yielding longer training times and increased likelihood of overfitting models [50]. Progress on this front has been made by mixing 2D and 3D convolutions and using $R(2+1)D$ convolutions (wherein 3D convolutions are factorized into spatial and temporal convolutions) [51,52]. Such architectures show great promise for application in POCUS, but the complexity of such networks leads to requirements for large amounts of data, which are often unavailable.

Another approach, first proposed by Simonyan et al., involves processing video data as two separate streams: a spatial and temporal stream [53]. The spatial stream is designed to classify still video frames and typically consists of a 3D CNN or a conventional 2D CNN which sequentially processes frames. The temporal stream is meant to capture inter-frame changes and is created by combining optical flow data from several frames. Generally, these two-stream CNNs outperform both conventional 2D and 3D CNNs for video classification. Howard et al. applied such a two-stream CNN to automatically determine the scan view from echocardiography data. Such two-stream CNNs can potentially lower the computational overhead for POCUS analysis and classification [54].

A major area of interest in our group and others has been the application of attention-gated networks to DL ultrasound. Attention mechanisms attempt to better mimic human perception by using surrounding local information in the data to contextualize a specific target. Attention models have been heavily used for natural language

processing (NLP) tasks, where integrating information from potentially distant parts of a sentence is necessary to accurately translate a given word [55]. Here, transformer block architectures have been used with success [56]. Attention mechanisms were first used by Mnih et al. in a recurrent neural network (RNN) for image classification [57], but has since been applied to a variety of ultrasound image analysis including in fetal ultrasound scan plane detection [58]. Attention models could prove useful in a variety of POCUS models including B-line score (BLS) determination from lung ultrasound. Accurate determination of BLS often depends on assessing adjacent frames rather than relying on a single frame. Attention models have the additional advantage of giving insight into which video time frames and what image content the algorithm is attending to for deriving its classification, thereby potentially improving interpretability.

Additionally, researchers have developed alternative approaches to identifying optimal network architectures through neural architecture search. Generally, network architectures are designed by data scientists using some *a priori* hypothesis of underlying data structure. This time-consuming task leaves the entirety of alternative network architectures largely unexplored. To address this issue, scientists and Google's AI division developed a neural architecture search, where machine learning techniques are used to optimize the network architecture while training the network itself [59,60]. This approach has been successful for improving the architecture of conventional (image-classification) CNNs and is now being applied to video CNNs. Piergiovanni et al. designed EvaNet, wherein they used an evolutionary algorithm to explore different layer types and combinations that could optimally represent the relationships between spatial and temporal aspects of videos [61]. Ryoo et al. designed AssembleNet, a network composed of multiple sub-network blocks that interprets input videos as multiple input streams sampled at different levels of temporal resolution [62]. AssembleNet is able to optimize the connectivity between both the sub-network blocks as well as the connectivity between the multiple variable-resolution streams. Such techniques are already being applied for medical image analysis. For example, Yan et al. developed MS-NAS (Multi-Scale Neural Architecture Search for Medical Image Segmentation) and applied it to outperform several state-of-the-art algorithms used for segmentation of CT images [63]. Given the temporal dynamics and acquisition complexities of ultrasound data, *a priori* hypotheses are unlikely to arrive at efficient network structures. Neural architecture search techniques, such as MS-NAS, will permit data-driven approaches to developing optimized algorithms that can address the broad range of ultrasound image processing

problems faced by clinicians.

Conclusion

In conclusion, we have introduced the concepts of machine learning and deep learning, reviewed current applications of these powerful tools in POCUS, discussed available commercial products utilizing machine learning, and explored promising future directions for machine learning on POCUS research. The utility of POCUS is largely derived from its capability for real-time inference and portability. While these factors present initial hurdles to the early adoption of machine learning in POCUS, they may also serve as the modalities greatest assets. Machine learning demands increasingly large datasets, sometimes needing millions of training images. POCUS is uniquely positioned to provide large datasets of video frames that could potentially be used for real-time algorithm training. Additionally, the portability of POCUS has the potential to provide a platform for rolling out machine learning applications in medical imaging to the entire world.

Acknowledgements & Funding

T. Campbell Arnold received support from the HHMI-NIBIB Interfaces Initiative (5T32EB009384-10). Ivan A. Kuznetsov acknowledges the fellowship support of the Paul and Daisy Soros Fellowship for New Americans and the NIH/National Institute of Mental Health (NIMH) Ruth L. Kirchstein NRSA (F30 MH122076-01). Momodou Lamin Sonko received support from the NIH/National Heart Lung and Blood Institute (R25-HL084665). We also would like to acknowledge the support of the University of Pennsylvania Emergency Medicine department.

Disclosures

The authors have no conflicts of interest to declare.

References

- Cadieu CF, Hong H, Yamins DLK, Pinto N, Ardila D, Solomon EA, et al. Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. *PLoS Comput Biol* 2014;10:1003963. <https://doi.org/10.1371/journal.pcbi.1003963>.
- Shen W, Zhou M, Yang F, Yang C, Tian J. Multi-scale convolutional neural networks for lung nodule classification. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9123, Springer Verlag; 2015, p. 588–99. https://doi.org/10.1007/978-3-319-19992-4_46.
- Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, et al. Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans. *Sci Rep* 2016;6:1–13. <https://doi.org/10.1038/srep24454>.
- Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, et al. CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. 33rd AAAI Conf. Artif. Intell. AAAI 2019, 31st Innov. Appl. Artif. Intell. Conf. IAAI 2019 9th AAAI Symp. Educ. Adv. Artif. Intell. EAAI 2019, vol. 33, AAAI Press; 2019, p. 590–7. <https://doi.org/10.1609/aaai.v33i01.3301590>.
- Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning 2017.
- Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLOS Med* 2018;15:e1002686. <https://doi.org/10.1371/journal.pmed.1002686>.
- Kulkarni S, Seneviratne N, Baig MS, Khan AHA. Artificial Intelligence in Medicine: Where Are We Now? *Acad Radiol* 2020;27:62–70. <https://doi.org/10.1016/j.acra.2019.10.001>.
- Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. *Proc IEEE Conf Comput Vis Pattern Recognit* 2017:2097–106.
- Liao SF, Chen PJ, Chaou CH, Lee CH. Top-cited publications on point-of-care ultrasound: The evolution of research trends. *Am J Emerg Med* 2018;36:1429–38. <https://doi.org/10.1016/j.ajem.2018.01.002>.
- Taylor ME, Stone P. *Transfer Learning for Reinforcement Learning Domains: A Survey*. vol. 10. 2009.
- Raina R, Battle A, Lee H, Packer B, Ng AY. Self-taught learning: Transfer learning from unlabeled data. *ACM Int. Conf. Proceeding Ser.*, vol. 227, New York, New York, USA: ACM Press; 2007, p. 759–66. <https://doi.org/10.1145/1273496.1273592>.
- Baribeau Y, Sharkey A, Chaudhary O, Krumm S, Fatima H, Mahmood F, et al. Handheld Point-of-Care Ultrasound Probes: The New Generation of POCUS. *J Cardiothorac Vasc Anesth* 2020;34:3139–45. <https://doi.org/10.1053/j.jvca.2020.07.004>.
- Pesteie M, Lessoway V, Abolmaesumi P, Rohling RN. Automatic Localization of the Needle Target for Ultrasound-Guided Epidural Injections. *IEEE Trans Med Imaging* 2018;37:81–92. <https://doi.org/10.1109/TMI.2017.2739110>.
- FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. *Food Drug Adm* 2018. <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye> (accessed August 12, 2021).
- Dezaki FT, Dhungel N, Abdi AH, Luong C, Tsang T, Jue J, et al. Deep residual recurrent neural networks for characterisation of cardiac cycle phase from echocardiograms. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2017. https://doi.org/10.1007/978-3-319-67558-9_12.
- Smistad E, Ostvik A, Haugen BO, Lovstakken L. 2D left ventricle segmentation using deep learning. *IEEE Int. Ultrason. Symp. IUS, IEEE Computer Society*; 2017. <https://doi.org/10.1109/ULTSYM.2017.8092573>.
- Chen H, Dou Q, Ni D, Cheng JZ, Qin J, Li S, et al. Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9349, Springer Verlag; 2015, p. 507–14. https://doi.org/10.1007/978-3-319-24553-9_62.
- Chen S, Ma K, Zheng Y. TAN: Temporal Affine Network for Real-Time Left Ventricle Anatomical Structure Analysis Based on 2D Ultrasound Videos 2019.
- Knackstedt C, Bekkers SCAM, Schummers G, Schreckenber 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. <https://doi.org/10.1016/j.jacc.2015.07.052>.
- Thavendiranathan P, Liu S, Verhaert D, Calleja A, Nitinunu A, Van Houten T, et al. Feasibility, accuracy, and reproducibility of real-time full-volume 3D transthoracic echocardiography to measure LV volumes and systolic function: A fully automated endocardial contouring algorithm in sinus rhythm and atrial fibrillation. *JACC Cardiovasc Imaging* 2012;5:239–51. <https://doi.org/10.1016/j.jcmg.2011.12.012>.
- Volpicelli G, Elbarbary M, Blaivas M, Lichtenstein DA, Mathis G, Kirkpatrick AW, et al. International evidence-based recommendations for point-of-care lung ultrasound. *Intensive Care Med.*, vol. 38, Springer; 2012, p. 577–91. <https://doi.org/10.1007/s00134-012-2513-4>.
- Covic A, Siriopol D, Voroneanu L. Use of Lung Ultrasound for the Assessment of Volume Status in CKD. *Am J Kidney Dis* 2018;71:412–22. <https://doi.org/10.1053/j.ajkd.2017.10.009>.
- Enghard P, Rademacher S, Nee J, Hasper D, Engert U, Jörres A, et al. Simplified lung ultrasound protocol shows excellent prediction of extravascular lung water in ventilated intensive care patients. *Crit Care* 2015. <https://doi.org/10.1186/s13054-015-0756-5>.

24. Maw AM, Hassanin A, Ho PM, McInnes MDF, Moss A, Juarez-Colunga E, et al. Diagnostic Accuracy of Point-of-Care Lung Ultrasonography and Chest Radiography in Adults With Symptoms Suggestive of Acute Decompensated Heart Failure: A Systematic Review and Meta-analysis. *JAMA Netw Open* 2019. <https://doi.org/10.1001/jamanetworkopen.2019.0703>.
25. Zoccali C, Torino C, Tripepi R, Tripepi G, D'Arrigo G, Postorino M, et al. Pulmonary congestion predicts cardiac events and mortality in ESRD. *J Am Soc Nephrol* 2013;24:639–46. <https://doi.org/10.1681/ASN.2012100990>.
26. Baloescu C, Toporek G, Kim S, McNamara K, Liu R, Shaw MM, et al. Automated Lung Ultrasound B-Line Assessment Using a Deep Learning Algorithm. *IEEE Trans Ultrason Ferroelectr Freq Control* 2020;67:2312–20. <https://doi.org/10.1109/TUFFC.2020.3002249>.
27. Sonko LM, Arnold TC, Kuznetsov I, Dean AJ, Panebianco NL, Reisinger N. Machine Learning for Rapid Interpretation of Lung Ultrasound. *Am J Kidney Dis* 2019;73:724. <https://doi.org/10.1053/j.ajkd.2019.03.311>.
28. Pereda MA, Chavez MA, Hooper-Miele CC, Gilman RH, Steinhoff MC, Ellington LE, et al. Lung ultrasound for the diagnosis of pneumonia in children: A meta-analysis. *Pediatrics* 2015;135:714–22. <https://doi.org/10.1542/peds.2014-2833>.
29. Parlamento S, Copetti R, Di Bartolomeo S. Evaluation of lung ultrasound for the diagnosis of pneumonia in the ED. *Am J Emerg Med* 2009;27:379–84. <https://doi.org/10.1016/j.ajem.2008.03.009>.
30. Caiulo VA, Gargani L, Caiulo S, Fisicaro A, Moramarco F, Latini G, et al. Lung ultrasound characteristics of community-acquired pneumonia in hospitalized children. *Pediatr Pulmonol* 2013;48:280–7. <https://doi.org/10.1002/ppul.22585>.
31. Correa M, Zimic M, Barrientos F, Barrientos R, Román-Gonzalez A, Pajuelo MJ, et al. Automatic classification of pediatric pneumonia based on lung ultrasound pattern recognition. *PLoS One* 2018;13:e0206410. <https://doi.org/10.1371/journal.pone.0206410>.
32. Born J, Brändle G, Cossio M, Disdier M, Goulet J, Roulin J, et al. POCOVID-Net: Automatic Detection of COVID-19 From a New Lung Ultrasound Imaging Dataset (POCUS) 2020.
33. Tsai YC, Chiu YW, Tsai JC, Kuo HT, Hung CC, Hwang SJ, et al. Association of fluid overload with cardiovascular morbidity and all-cause mortality in stages 4 and 5 CKD. *Clin J Am Soc Nephrol* 2015;10:39–46. <https://doi.org/10.2215/CJN.03610414>.
34. Tai R, Ohashi Y, Mizuiri S, Aikawa A, Sakai K. Association between ratio of measured extracellular volume to expected body fluid volume and renal outcomes in patients with chronic kidney disease: A retrospective single-center cohort study. *BMC Nephrol* 2014;15:189. <https://doi.org/10.1186/1471-2369-15-189>.
35. Zoccali C, Moissi U, Chazot C, Mallamaci F, Tripepi G, Arkossy O, et al. Chronic fluid overload and mortality in ESRD. *J Am Soc Nephrol* 2017;28:2491–7. <https://doi.org/10.1681/ASN.2016121341>.
36. Ravishanker H, Annangi P, Washburn M, Lanning J. Automated kidney morphology measurements from ultrasound images using texture and edge analysis. In: Duric N, Heyde B, editors. *Med. Imaging 2016 Ultrason. Imaging Tomogr.*, vol. 9790, SPIE; 2016, p. 97901A. <https://doi.org/10.1117/12.2216802>.
37. Wu Z, Hai J, Zhang L, Chen J, Cheng G, Yan B. Cascaded Fully Convolutional DenseNet for Automatic Kidney Segmentation in Ultrasound Images. 2019 2nd Int. Conf. Artif. Intell. Big Data, ICAIBD 2019, Institute of Electrical and Electronics Engineers Inc.; 2019, p. 384–8. <https://doi.org/10.1109/ICAIBD.2019.8836994>.
38. Chen CJ, Pai TW, Hsu HH, Lee CH, Chen KS, Chen YC. Prediction of chronic kidney disease stages by renal ultrasound imaging. *Enterp Inf Syst* 2020;14:178–95. <https://doi.org/10.1080/17517575.2019.1597386>.
39. Kuo C-C, Chang C-M, Liu K-T, Lin W-K, Chiang H-Y, Chung C-W, et al. Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning. *Npj Digit Med* 2019;2:1–9. <https://doi.org/10.1038/s41746-019-0104-2>.
40. Dudley NJ, Chapman E. The importance of quality management in fetal measurement. *Ultrasound Obstet Gynecol* 2002;19:190–6. <https://doi.org/10.1046/j.0960-7692.2001.00549.x>.
41. Kim B, Kim KC, Park Y, Kwon JY, Jang J, Seo JK. Machine-learning-based automatic identification of fetal abdominal circumference from ultrasound images. *Physiol Meas* 2018;39:105007. <https://doi.org/10.1088/1361-6579/aae255>.
42. Jang J, Park Y, Kim B, Lee SM, Kwon JY, Seo JK. Automatic Estimation of Fetal Abdominal Circumference from Ultrasound Images. *IEEE J Biomed Heal Informatics* 2017. <https://doi.org/10.1109/JBHI.2017.2776116>.
43. Gao Y, Maraci MA, Noble JA. Describing ultrasound video content using deep convolutional neural networks. *Proc. - Int. Symp. Biomed. Imaging*, vol. 2016- June, IEEE Computer Society; 2016, p. 787–90. <https://doi.org/10.1109/ISBI.2016.7493384>.
44. Kline JA, O'Malley PM, Tayal VS, Snead GR, Mitchell AM. Emergency Clinician-Performed Compression Ultrasonography for Deep Venous Thrombosis of the Lower Extremity. *Ann Emerg Med* 2008;52:437–45. <https://doi.org/10.1016/j.annemergmed.2008.05.023>.
45. Nafee T, Gibson CM, Travis R, Yee MK, Kerneis M, Chi G, et al. Machine learning to predict venous thrombosis in acutely ill medical patients. *Res Pract Thromb Haemost* 2020;4:230–7. <https://doi.org/10.1002/rth2.12292>.
46. Tanno R, Makropoulos A, Arslan S, Oktay O, Mischkewitz S, Al-Noor F, et al. AutoDVT: Joint real-time classification for vein compressibility analysis in deep vein thrombosis ultrasound diagnostics. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11071 LNCS, Springer Verlag; 2018, p. 905–12. https://doi.org/10.1007/978-3-030-00934-2_100.
47. Short J, Acebes C, Rodriguez-De-Lema G, La Paglia GMC, Pavón M, Sánchez-Pernaute O, et al. Visual versus automatic ultrasound scoring of lung B-lines: Reliability and consistency between systems. *Med Ultrason* 2019;21:45–9. <https://doi.org/10.11152/mu-1885>.
48. Filipiak-Strzecka D, Kasprzak JD, Wejner-Mik P, Szymczyk E, Wdowiak-Okrojek K, Lipiec P. Artificial Intelligence-Powered Measurement of Left Ventricular Ejection Fraction Using a Handheld Ultrasound Device. *Ultrasound Med Biol* 2021;47:1120–5. <https://doi.org/10.1016/j.ultrasmedbio.2020.12.003>.
49. Hara K, Kataoka H, Satoh Y. Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet? *Proc IEEE Conf Comput Vis Pattern Recognit* 2018:6546–55.
50. Mahadevan S, Athar A, Oşep A, Hennen S, Leal-Taixé L, Leibe B. Making a Case for 3D Convolutions for Object Segmentation in Videos 2020.
51. Xie S, Sun C, Huang J, Tu Z, Murphy K. Rethinking Spatiotemporal Feature Learning: Speed-Accuracy Trade-offs in Video Classification. *Proc Eur Conf Comput Vis* 2018:305–21.
52. Tran D, Wang H, Torresani L, Ray J, Lecun Y, Paluri M. A Closer Look at Spatiotemporal Convolutions for Action Recognition. *Proc IEEE Conf Comput Vis Pattern Recognit* 2018:6450–9.
53. Simonyan K, Zisserman A. Two-Stream Convolutional Networks for Action Recognition in Videos. *Adv Neural Inf Process Syst* 2014;1:568–76.
54. Howard JP, Tan J, Shun-Shin MJ, Mahdi D, Nowbar AN, Arnold AD, et al. Improving ultrasound video classification: an evaluation of novel deep learning methods in echocardiography. *J Med Artif Intell* 2020;3:4–4. <https://doi.org/10.21037/jmai.2019.10.03>.
55. Vaswani A, Brain G, Shazeer N, Parmar N, Uszkoreit J, Jones L, et al. Attention Is All You Need. *Proc Neural Inf Process Syst* n.d.
56. Cheng J, Dong L, Lapata M. Long short-term memory-networks for machine reading. *EMNLP 2016 - Conf Empir Methods Nat Lang Process Proc* 2016:551–61.
57. Recurrent models of visual attention | Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2. *Proc 27th Int Conf Neural Inf Process Syst* 2014:2204–12.
58. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, et al. Attention gated networks: Learning to leverage salient regions in medical images. *Med Image Anal* 2019;53:197–207. <https://doi.org/10.1016/j.media.2019.01.012>.
59. Real E, Moore S, Selle A, Saxena S, Suematsu YL, Tan J, et al. Large-Scale Evolution of Image Classifiers. *Proc 34th Int Conf Mach Learn PMLR* 2017:2902–11.
60. Zoph B, Le Q V. Neural Architecture Search with Reinforcement Learning. 5th Int Conf Learn Represent ICLR 2017 - Conf Track Proc 2016.
61. Piergiovanni AJ, Angelova A, Toshev A, Ryoo MS, Brain G. Evolving Space-Time Neural Architectures for Videos. *Proc IEEE/CVF Int Conf Comput Vis* 2019:1793–802.

62. Ryoo MS, Piergiovanni A, Tan M, Angelova A. AssembleNet: Searching for Multi-Stream Neural Connectivity in Video Architectures 2019.
63. Yan X, Jiang W, Shi Y, Zhuo C. Ms-nas: Multi-scale neural architecture search for medical image segmentation. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12261 LNCS, Springer Science and Business Media Deutschland GmbH; 2020, p. 388–97. https://doi.org/10.1007/978-3-030-59710-8_38.
64. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation, Springer, Cham; 2015, p. 234–41. https://doi.org/10.1007/978-3-319-24574-4_28.