Exploring the utility of assistive artificial intelligence for ultrasound scanning in regional anesthesia

James Simeon Bowness 1,2, Kariem El-Boghdady 3,4, Glenn Woodward 5, J Alison Noble 6, Helen Higham 1,7, David Burckett-St Laurent 8

ABSTRACT

Introduction  Ultrasound-guided regional anesthesia (UGRA) involves the acquisition and interpretation of ultrasound images to delineate sonoanatomy. This study explores the utility of a novel artificial intelligence (AI) device designed to assist in this task (ScanNav Anatomy Peripheral Nerve Block; ScanNav), which applies a color overlay on real-time ultrasound to highlight key anatomical structures.

Methods  Thirty anesthesiologists, 15 non-experts and 15 experts in UGRA, performed 240 ultrasound scans across nine peripheral nerve block regions. Half were performed with ScanNav. After scanning each block region, participants completed a questionnaire on the utility of the device in relation to training, teaching, and clinical practice in ultrasound scanning for UGRA. Ultrasound and color overlay output were recorded from scans performed with ScanNav. Experts present during the scans (real-time experts) were asked to assess potential for increased risk associated with use of the device (eg, needle trauma to safety structures). This was compared with experts who viewed the AI scans remotely.

Results  Non-experts were more likely to provide positive and less likely to provide negative feedback than experts (p=0.001). Positive feedback was provided most frequently by non-experts on the potential role for training (37/60, 61.7%); for experts, it was for its utility in teaching (30/60, 50%). Real-time and remote experts reported a potentially increased risk in 12/254 (4.7%) vs 8/254 (3.1%, p=0.362) scans, respectively.

Discussion  ScanNav shows potential to support non-experts in training and clinical practice, and experts in teaching UGRA. Such technology may aid the uptake and generalizability of UGRA.

Trial registration number  NCT04918693.

INTRODUCTION

The performance of ultrasound-guided regional anesthesia (UGRA) relies on the optimal acquisition and interpretation of ultrasound images. This skill is underpinned by knowledge of the underlying anatomy. However, even experienced anesthesiologists can find this challenging, particularly in the context of anatomical variation or other difficulties in scanning (eg, obesity). Furthermore, sonoanatomic image interpretation is variable, even among experts. In the authors’ experience, though ultrasound technology yields ever-greater image resolution, acumen in ultrasound image acquisition and interpretation has not demonstrated an equal improvement since the introduction of ultrasound guidance in regional anesthesia in 1989.

The implementation of artificial intelligence (AI) in clinical practice continues to grow and regional anesthesia has potential to embrace this technology for patient benefit, such as supporting UGRA training and practice to increase patient access to these techniques, as well as improve patient safety. Broadly speaking, AI includes any technique that enables computers to undertake tasks associated with human intelligence. Common techniques in this field include machine learning (ML) and deep learning (DL). ML uses algorithms, rule-based problem-solving instructions implemented by the computer, to learn: training data are analyzed to identify patterns (statistical correlations) in this information. If the algorithm is informed of the desired endpoints (eg, by labeling the images during the training stage) and then looks for correlations between the raw input data and the endpoints, this is called supervised ML. In unsupervised ML, the algorithm is given no instruction as to what endpoints are desired but looks for clusters of similarity in the dataset. Semisupervised ML uses a mixture of these two approaches. DL is a method of implementing ML. The algorithm encodes a network of artificial neurons (mathematical functions) which are arranged in layers. Successive layers of the network operate on the data as it is passed through the system, extracting progressively more information. DL is a common technique used in computer vision—a branch of AI which allows computers to interact with the visual world—a field that is readily applicable to medical image analysis.

ScanNav Anatomy Peripheral Nerve Block (ScanNav, formerly known as AnatomyGuide; Intelligent Ultrasound, Cardiff, UK) is an AI-based device that uses DL to apply a color overlay to real-time B-mode ultrasound. It highlights relevant anatomical structures on the ultrasound image, aiming to assist ultrasound image interpretation (see online supplemental files A–D). Expert evaluation of ultrasound videos has previously considered the color overlay to be helpful in identifying specific structures and confirming an appropriate block view in over 99% of cases. Similar systems include NerveBox (Smart Alfa Teknoloji San. Ve Tic AS, Ankara, Turkey), which also applies a color overlay of anatomical structures on ultrasound in regions commonly scanned for UGRA, and NerveTrack (Samsung Medison, Seoul, South Korea), which applies a bounding box around the median and ulnar nerves when scanning the forearm.
Brief technical report

Though highlighting accuracy data have been published for ScanNav and Nervebloc,14 15 there are limited data from real-world use of these systems—particularly on their utility for operators. We therefore aimed to assess the subjective utility of ScanNav as an aid to identifying relevant structures, teaching/learning UGRA scanning, and increasing operator confidence. We also assessed UGRA experts’ perception of potential risks associated with the use of ScanNav, including increased risk of block failure or unwanted needle trauma to important safety structures (eg, nerves, arteries, and pleura/peritoneum).

METHODS
The study was registered with www.clinicaltrials.gov.

Non-expert participants (ultrasound scanners)
Fifteen non-experts in UGRA were recruited; US postgraduate year 2–4 medical doctors enrolled in the anesthesiology residency training program at OHSU.

Expert participants (ultrasound scanners/real-time assessors)
Fifteen experts in UGRA were recruited: US board-certified anesthesiologists and current or former anesthesiology faculty at OHSU. All were competent to perform UGRA independently and met at least two of the following three characteristics: completed fellowship training in UGRA, regularly delivered direct clinical care using UGRA (including for ‘awake’ surgery), and regularly taught UGRA in the course of their clinical work (including advanced techniques).

Remote expert assessors
Three remote experts were recruited to subsequently review ultrasound videos recorded when a subject was scanned using ScanNav (unmodified videos presented with videos to which the AI color overlay was applied). One was a UK-based consultant anesthetist and two were US-based attending anesthesiologists. All met the criteria defined previously (with the exception that one was a pediatric anesthesiologist, thus their conduct of ‘awake’ surgery is limited by their patient population). With the exception of one participant (GW, who is an author if this article and was not remunerated for work in this study), participants and remote expert assessors did not form part of the investigator group and were compensated for their time.

Subjects
Two healthy volunteers for ultrasound scanning were recruited from a professional modeling agency and compensated for their services. Half of the scans were performed on each subject. The only exclusion criterion was pathology of the areas to be scanned.

Equipment
Ultrasound scanning was performed using the X-Porte or PX SonoSite ultrasound machines (Fujifilm SonoSite, Bothell, Washington, USA). Participants were free to choose from a selection of compatible probes for each machine; X-Porte with HFL50xp/ L38xp linear or C60xp curvilinear probe, or PX with L15-5/ L12-3 linear or C5-1 curvilinear probe.

ScanNav (US V.1.0) was connected to the main ultrasound machine high-definition multimedia interface (HDMI) output. It displayed the same ultrasound image, with color overlay, on a secondary monitor (see figure 1). It is important to note that this device is not intended to replace clinician judgment and has been submitted to the US Food and Drug Administration for regulatory approval as a scanning-only device (for use prior to needle insertion or local anesthetic injection).

Scanning protocol
After gaining informed consent (both scanners and subjects), ultrasound assessment was performed at the typical locations for nine peripheral nerve blocks. The upper limb included the interscalene-level, upper trunk-level, supraclavicular-level and axillary-level brachial plexus regions. The trunk included the erector spinae plane and rectus sheath plane block regions. Lower limb scanning included regions for the suprapinguinal-level fascia iliaca plane, adductor canal and popliteal-level sciatic nerve blocks.

Participants were free to set the depth and gain settings deemed appropriate for each scan and use the scanning technique they would in their normal clinical practice. The first scan was performed by an expert and performed on both sides, once with the use of ScanNav and once without. Eight participants performed the first scan with the use of ScanNav and seven without (order alternating between participants). Subsequently, a non-expert participant entered the room and was asked to scan for the same block in the same subject, under the supervision of the expert. The expert taught and/or supported the non-expert, as necessary, to achieve an optimal block view. As before, both sides were scanned, once with the use of the AI color overlay and once without (order alternating as above). This same process of four scans was then repeated on the same subject, scanning for a different block. The expert/non-expert pair then repeated the scanning protocol on a second subject (the scanning sequence is summarized in online supplemental file E).

This process of 16 scans (8 by the expert and 8 by the non-expert) was performed by 15 expert/non-expert pairs (total 240 scans). All AI-assisted scans (n=120) were recorded for later analysis by remote experts.

Utility: experts versus non-experts
After scanning for each block, the participants were asked to complete a questionnaire evaluating the AI color overlay for the images acquired (n=120, half by non-experts and half by experts; questionnaires available in online supplemental file E).
Potential risks: real-time expert users versus remote experts

For each scan performed with ScanNav (n=120), experts were asked to report potential for increased risk of block failure or unwanted needle trauma to 'safety critical' structures as relevant to each block (eg, nerves, arteries, pleura, and peritoneum; see online supplemental file E). The risk of complications related to each structure included:

- Nerve injury/postoperative neurological symptoms (PONS, nerves).
- Local anesthetic systemic toxicity (LAST, arteries).
- Pneumothorax (pleura).
- Peritoneal violation (peritoneum).
- Block failure (overall).

The real-time expert assessment was compared with that of the panel of remote experts, who viewed the recorded ultrasound with the unmodified video presented adjacent to the AI-color overlay video. Three remote experts viewed each ultrasound video and a majority view for each question was taken. In cases where no majority was reached, this was classified as ‘no consensus’.

Data analysis

Data were reported descriptively and, where appropriate, statistical evaluation (using R software V4.1.1) was used to assess the relationship between variables. A $\chi^2$-test was used to compare feedback (expert vs non-expert, real-time user vs remote expert, and subject 1 vs subject 2) except for participant confidence, which used a Mann-Whitney U test to compare ordinal data. Statistical significance was deemed as a p value of <0.05.

RESULTS

In total, 240 ultrasound scans were performed, 120 with ScanNav and 120 without. Of the 120 scans performed with the ScanNav, 60 were performed by non-experts in UGRA (under the supervision of experts) and 60 by experts. Both scan subjects were adult men; one was 34 years old with a body mass index (BMI) of 37.2 kg/m$^2$ and the other was 41 years old with a BMI of 28.9 kg/m$^2$. The data are summarized in tables 1–3 and a full breakdown is in online supplemental file E.

Utility: non-experts versus experts

Non-experts provided positive feedback more frequently and provided negative feedback less frequently than experts (p=0.001). The most frequent positive feedback provided by non-experts was on ScanNav’s role in their training (37/60, 61.7%); for experts, it was for ScanNav’s use in teaching (30/60, 50%). Overall, 70% of participants reported that ScanNav aided in the identification of key anatomical structures for a peripheral nerve block, and 63% believed that it assisted in confirming the correct ultrasound view during scanning. Non-experts’ most frequent negative feedback was that it may decrease their confidence in scanning (4/60, 6.7%); for experts, it was that ScanNav may increase the frequency of supervisor intervention (10/60, 16.7%). Non-expert median confidence in their own scanning was 6 (IQR 5–8) without ScanNav and 7 (IQR 5.75–9) with ScanNav (p=0.07). Experts reported median confidence of 10 (IQR 8–10) vs 10 (IQR 8–10) respectively (p=0.57). When supervising a non-expert scanning, their median confidence in the non-expert’s scanning was 7 (IQR 4.75–8) without ScanNav and 8 (IQR 4–9) with ScanNav (p=0.23). Overall, there was no difference in the reporting between subjects scanned (p=0.562).

Potential risks: real-time user expert versus remote expert

Of the 120 scans performed with ScanNav, real-time user expert data were collected on 103 scans (17 lost), and remote expert assessment was thus compared for the same 103 scans. Real-time and remote expert reported a potential increase in risk: 12/254 (4.7%) vs 8/254 (3.1%), respectively (p=0.362, table 3).

DISCUSSION

This study explores the potential role of assistive AI in the acquisition and interpretation of ultrasound scans for UGRA, with real-world users in a simulation setting. Positive sentiment
The data show a more critical view of real-time versus remote identification alone does not ensure safe or effective UGRA nor systems, eg, color Doppler). Furthermore, correct structure/block drawing focus away from other relevant structures. The current technology of AI devices to support ultrasound scanning in UGRA. We have demonstrated that ScanNav may support non-experts in their teaching of UGRA. It may help by drawing attentional focus to the area of interest to aid in confirmation of the correct ultrasound view and the identification of sonoanatomical structures on that view. We believe that AI-augmented ultrasound scanning, through devices such as ScanNav, may support the uptake and generalizability of ultrasound-guided techniques in the future.

CONCLUSION
The few studies conducted in this field so far report little in the way of contemporaneous feedback by users or evaluation of the clinical utility of AI devices to support ultrasound scanning in UGRA. We have demonstrated that ScanNav may support non-experts in their training and clinical practice, and experts in their teaching of UGRA. It may help by drawing attentional focus to the area of interest to aid in confirmation of the correct ultrasound view and the identification of sonoanatomical structures on that view. We believe that AI-augmented ultrasound scanning, through devices such as ScanNav, may support the uptake and generalizability of ultrasound-guided techniques in the future.

Twitter James Simeon Bowness @bowness_james, Kariem El-Boghdaydy @elboghdaydy, Glenn Woodworth @woodworgMD, J Alison Noble @AlisonNoble_OU, Helen Higham @HelenHigham and David Buckett-St Laurent @davidbosl

Acknowledgements The authors thank Drs Boyne Bellew, David Maduram and Bryant Tran for their remote review of artificial intelligence-overlay ultrasound videos in the study.

Contributors Study concept and design: JSB and DBSL; data collection: GW; manuscript preparation: JSB; manuscript editing, review and approval: all authors.

Funding This work was funded by Intelligent Ultrasound Limited (Cardiff, UK). Data from this study were included in medical device regulatory approval submissions in the USA.

Competing interests JSB is a Senior Clinical Advisor for Intelligent Ultrasound, receiving research funding and honoraria. KEB is an Editor for Anaesthesia and has received research, honoraria and educational funding from Fisher and Paykel Healthcare Ltd, GE Healthcare and Ambu. DBSL is a Clinical Advisor for Intelligent Ultrasound, receiving honoraria. JAN is a Senior Scientific Advisor for Intelligent Ultrasound.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and ethical approval was granted by the Oregon Health & Science University Institutional Review Board.

(36.4%–45.8%) about ScanNav was reported by users more commonly than negative sentiment (2.5%–7.5%), while the majority (63%–70%) reported that it aided in the identification of key anatomical structures on ultrasound and in confirmation of the correct ultrasound view. Non-experts derived most benefit from ScanNav; ≥50% reported it to aid in the identification of sonoanatomical structures, learning the scanning for a block, benefits to training and improving their confidence in scanning. These data suggest a role for ScanNav in the training of non-experts in UGRA. Training can be in the form of non-patient-facing activities such as formal teaching and educational courses. However, teaching during clinical practice plays a fundamental role in medical training. Thus, although it requires additional financial resources, ScanNav may be used for educational purposes in both settings, supporting widespread adoption of UGRA and patient access to these techniques. The areas receiving the highest frequency of negative feedback (frequency of supervisor intervention and decreasing confidence of performing/supervising ultrasound scanning) are perhaps unsurprising. This is a new medical device, based on AI technology with many clinicians are unfamiliar, and the participants in this study had not used it prior to participating in this study. Initial use of any new technology may be associated with more frequent intervention and lower confidence, which may improve with time and increased familiarity with the device. There was a low perception of increased risk associated with AI highlighting by ScanNav (3.1%–4.7%), though potential complications considered may be clinically important (eg, nerve injury/PONS and LAST). Potential causes of error include those related to device performance, such as incorrect highlighting, which may cause anesthesiologists to misinterpret the ultrasound image(s). Block failure or unwanted needle trajectory to safety critical structures may therefore be more likely if the anesthesiologist is falsely reassured by the presence or absence of color on the screen. Others may be associated with use of the device, such as correct highlighting causing distraction by drawing focus away from other relevant structures. The current technology therefore provides additional information for the anesthesiologist rather than a decision-making system on ultrasound image interpretation (much as is the case for other image augmentation systems, eg, color Doppler). Furthermore, correct structure/block view identification alone does not ensure safe or effective UGRA nor does the device guide needle placement. It is therefore the practitioner’s responsibility to identify hazards and undertake safe practice. The data show a more critical view of real-time users as compared with those viewing the video remotely. Real-time users have a richer source of information as they performed the dynamic scan; however, they had to subsequently give their assessment from their memory of the preceding few minutes. Remote users had the benefit of scrolling back and forth through the video to carefully scrutinize the unmodified ultrasound video. In addition, three remote experts assessed each video compared with one real-time user expert. It is not possible to determine, with these limited data, whether one cohort or another is correct—but to be aware of the range of opinion and that more work must be done to explore this facet of the data. Nevertheless, a more cautious view adopted by real-time users is perhaps a welcome inadvertent safety feature, showing a desire to maintain the use of clinical judgment which has been developed over years of training.

This study has limitations. These data are subjective, based on a limited use of the device, and may not necessarily reflect clinical practice. Over time, clinical data may support or refute the conclusions drawn here and could include studies in other settings (eg, UGRA in the emergency department or prehospital emergency care). Outcome data, including patient outcomes and resource-use metrics, will be crucial in validating the clinical utility of ScanNav. This study included only 2 scan subjects and 30 anesthesiologist participants; these data may need to be replicated with a larger number of participants and subjects with different demographics, such as body habitus, comorbid status or anatomical abnormalities. Also, all participants were from a single institution, while UGRA practices and experiences may vary between institutions, regions or countries. In addition, a three-point scale was given for participants to provide their assessment, whereas a five-point scale may have allowed greater discrimination of subjective assessment. Finally, only two ultrasound machines were assessed; further work is required to ensure generalizability of these data.

Table 3

<table>
<thead>
<tr>
<th>Real-time expert</th>
<th>Remote expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nerve injury/PONS (67)</td>
<td>5 (7.5%)</td>
</tr>
<tr>
<td>LAST (71)</td>
<td>3 (4.2%)</td>
</tr>
<tr>
<td>Pneumothorax (16)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Peritoneum violation (8)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Block failure (92)</td>
<td>4 (3.4%)</td>
</tr>
<tr>
<td>Total (254)</td>
<td>12 (4.7%)</td>
</tr>
</tbody>
</table>

N.B. Despite 103 block feedback sheets return, not all feedback elements were completed by participants for all blocks (hence, n=103 for block failure).

*8/92 of all block views considered (7/8 of these views considered ‘inadequate’ by expert panel).

LAST, local anesthetic systemic toxicity; PONS, postoperative neurological symptoms.
Participants gave informed consent to participate in the study before taking part.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, an indication of whether changes were made, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

**ORCID iDs**

James Simeon Bowness http://orcid.org/0000-0002-8665-1984
Kariem El-Boghdady http://orcid.org/0000-0002-9912-717X
Glenn Woodworth http://orcid.org/0000-0002-1924-801X
J Alison Noble http://orcid.org/0000-0002-3060-3772
Helen Higham http://orcid.org/0000-0001-5696-0377
David Burckett-St Laurent http://orcid.org/0000-0002-0772-5851

**REFERENCES**