Computers in Biology and Medicine 107 (2019) 39–46

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Quantifying lung ultrasound comets with a convolutional neural network: Initial clinical results

Xianglong Wang[a], Joseph S. Burzynski[a], James Hamilton[b], Panduranga S. Rao[c], William F. Weitzel[c,d], Joseph L. Bull[e,*]

[a] Biomedical Engineering, College of Engineering, University of Michigan – Ann Arbor, Ann Arbor, MI, USA
[b] Emerge, Inc, Los Angeles, CA, USA
[c] Division of Nephrology, Internal Medicine, University of Michigan – Ann Arbor, Ann Arbor, MI, USA
[d] Ann Arbor Veteran Affairs Healthcare System, Ann Arbor, MI, USA
[e] Biomedical Engineering, School of Science and Engineering, Tulane University, New Orleans, LA, USA

ARTICLE INFO

Keywords:
Lung
Lung comets
B-Lines
Ultrasound
Machine learning
Neural network

ABSTRACT

Lung ultrasound comets are "comet-tail" artifacts appearing in lung ultrasound images. They are particularly useful in detecting several lung pathologies and may indicate the amount of extravascular lung water. However, the comets are not always well defined and large variations in the counting results exist between observers. This study uses a convolutional neural network to quantify these lung ultrasound comets on a 4864-image clinical lung ultrasound dataset labeled by the authors. The neural network counted the number of comets correctly on 43.4% of the images and has an intraclass correlation (ICC) of 0.791 with respect to human counting on the test set. The ICC level indicates a higher correlation level than previously reported ICC between human observers. The neural network was then deployed and applied to a clinical 6272-image dataset. The correlation between the automated comet counts and the clinical parameters was examined. The comet counts correlate positively with the diastolic blood pressure ($p = 0.047$, $r = 0.448$), negatively with ejection fraction ($p = 0.061$, $r = -0.513$), and negatively with BMI ($p = 0.009$, $r = -0.566$). The neural network can be alternatively formulated as a diagnostic test for comet-positive images with 80.8% accuracy. The results could potentially be improved with a larger dataset and a refined approach to the neural networks used.

1. Introduction

Lung ultrasound comets, or "B-Lines", are "comet-tail" artifacts that emanate from the lung surface [1]. Various lung ultrasound findings are receiving increasing attention for detecting or ruling out several lung pathologies [2–6]. In particular, observations of lung ultrasound comets have been related to the presence of extravascular lung water [1]. Recent evidence further suggests that these comet artifacts may useful in evaluating pulmonary edema [7–10] and have the potential to predict patient outcomes [7]. These ultrasound findings may be particularly helpful in the end stage renal disease setting where patients suffer from fluid overload and pulmonary edema [7,9].

The current practice of identifying lung ultrasound comets, which involves physicians observing still ultrasound images, has two major limitations. The number and location of lung comets can vary during the time span of an ultrasound motion picture (cine) loop [11], meaning that the identification of comets from a single image is likely not representative of the region imaged [9]. In addition, the evaluation of lung comets can vary from observer to observer and between initial interpretation and reinterpretation by a single observer [12].

Initial steps have been taken to address the limitations by adopting an automated computer algorithm with varied degree of clinical success [9,13–16]. These previous approaches assume lung comets are pure line features and apply line detection methods on the ultrasound images. Traditional filtering and feature extraction methods were used on some of these approaches [9,13,14]. These methods do not sufficiently differentiate B-lines from the other line-type artifacts in the images. More recently, B-line detection has been formulated as an inverse problem [15,16]. The approach by Anantrasirichai et al. [16] has shown significant promise with a reported F_1-score of 0.94 for individual comets in a 100-image pediatric ultrasound image set acquired from array transducers, but it is unclear how the approach will perform when extended to adult ultrasound images or when applied to images acquired by single element transducers.

* Corresponding author. 524 Lindy Boggs Center, Department of Biomedical Engineering, Tulane University, New Orleans, 70118, USA.
E-mail address: jbull@tulane.edu (J.L. Bull).

https://doi.org/10.1016/j.compbiomed.2019.02.002
Received 28 November 2018; Received in revised form 3 February 2019; Accepted 3 February 2019

Artificial neural networks are well established and have been successfully adapted for computer vision and classification in machine learning [17]. Unlike traditional image processing methods, neural networks would not assume lung comets as lines and hidden features from an observer may be extracted. With adequate training data support, artificial neural networks can yield a satisfactory level of accuracy in classification of images. Training these networks often require high-performance GPUs but interpretation can be computationally inexpensive. Because the accuracy of a neural network model can be determined by comparison of the output of the neural network with the observed findings by human, the performance of the algorithm can be evaluated using the human observations as the reference.

Our goal is to guide patient diagnosis and quantify lung edema by applying real-time lung comet quantification using machine learning with a portable low-cost USB ultrasound probe in a point-of-care setting. This study is the initial step towards our goal. In this pilot study, we used a convolutional neural network as the primary algorithm to identify lung ultrasound comets to evaluate the use of neural networks in future lung ultrasound comet studies.

2. Methods

The data was collected under our institutional review board (IRB) approved study HUM 79017 in 2013–14 with informed consent from the subjects. A database of lung ultrasound images along with their comet counts were constructed from the collected data. The lung ultrasound measurements were performed using a commercially available USB-based portable transducer with single element mechanical scanning (7.5 MHz, small parts probe, Interson Corporation, Pleasanton, CA). Ultrasound scanning was performed through the intercostal (IC) spaces in the anterior and lateral chest for the right and left hemi-thorax using data collection methods previously reported [7,9]. Uniform gain and power settings were used throughout the ultrasound image acquisition. The database was then used for training and testing the performance of the neural network models. The accuracy of comet count was used as the primary measure of performance of the system; in addition, the ability of the system to classify an ultrasound image as “negative, mild, or severe” was used as another measure of performance. The neural network is also deployed on a small scale of clinical data to establish its clinical significance. The clinical dataset is composed of 152 lung ultrasound loops with 32 frames each, for a total 4864 images. These cine loops were obtained under the protocol documented in our previous paper [9]. The original study involved 20 stable hemodialysis patients, 17 men and 3 women, with two sessions of scans before and after the patients undergo dialysis. Each set of scan contains 12 to 32 loops [9]. The methods documented below covers the construction of the database, the structure of the neural network, the considerations to address overfitting, and the computing environment.

2.1. Database construction

The 152 cine loops were randomly selected from a pool of more than 500 loops, composed of the first to thirteenth cine loop from each scan session. Locations #1–13 were chosen because these loops were found to be relatively comet rich. A total of 4864 ultrasound images were obtained during the above process. These images are then randomly pooled into three sets: a training set (3584 images, 73.7%), a cross-validation set (640 images, 13.2%) and a test set (640 images, 13.2%).

The images were imported into an in-house GUI for data collection and were labeled per the number of comets present. The region of interest (ROI), recycled from our previous study [9], defines a 10 mm region just below the lung border, where the lung comets are typically present. The available labels for each image ranges are every integer between 0 and 6, for a total of seven available labels because 6 is the observed maximum number of comets present in the image database. Due to the author's intention of performing a feasibility study, the lung comets were not segmented and labeled individually; instead, the images were labeled with the number of comets present in the ROI stored as the label. The observer had no knowledge of the clinical parameters of the subject, the comet counts from the previous study, or which set the image would fall into. The ROI was stored instead of the whole picture because the ROI is the most comet rich and the number of pixels contained in the ROI (410×256) is much smaller than the whole image (2048×256). The ROI was then rescaled to a resolution of 128×128 for a further reduction in dimensionality of input. The comets are seen in the original image as 5–10 pixels long and hundreds of pixels wide according to the observer, so they can still be seen after resizing. All images are in grayscale. A workflow of the labeling process is available in Fig. 1.

2.2. Structure of the neural network

The architecture of a basic convolutional neural network can contain three types of layers: convolutional layers, pooling layers, and fully connected layers. The convolutional layer uses a small kernel and extracts features from a larger image. The number of convolutional layers tends to increase with the depth of the neural network. The pooling layer effectively down-samples an image and decreases its size. The fully connected layer computes its results from all the parameters from the previous layers and is often used for categorization purposes. An activation function can appear along with these layers above to regulate the output of these layers and apply nonlinearity to their output.

This neural network was constructed with considerations of a relatively high dimension of input and a small size of the training set. Rectified linear units (ReLU) were chosen as the activation functions for the neurons due to a reported faster training speed and the ability to prevent overfitting [18]. Overlapping average pooling with a window size of 3×3 and stride 2 was used across all layers except for the first layer, which uses max pooling of the same window size. This pooling scheme generated the highest accuracy on the validation set.

The final neural network contains four convolutional layers and a fully connected layer. Each of the four convolutional layers has kernels of size 5×5 with non-decreasing data depths of 16, 32, 64, and 128. The output of each convolutional layer is rectified by a ReLU unit and accompanied by a pooling layer. The fully connected layer contains a softmax regression layer for final classification. The input of this neural network is 16384-dimensional (128×128); the output is a single label for the number of comets present in the image. The network contains 5.3×10^5 parameters in total. The overall structure of the neural network can be found in Fig. 2.

During evaluation, the neural network is evaluated five times and the most common result was used as the output.

2.3. Considerations to address overfitting

Although the network is relatively small with 5.3×10^5 parameters, overfitting is still a big concern since the number of parameters overwhelm the number of training samples. Two measures were taken to address the potential of overfitting: data augmentation and dropout.

Data augmentation creates unobserved data from observed data and can reduce the chance of overfitting [19,20]. To artificially increase the size of the dataset, two kinds of data augmentation were used. First, the images are flipped to double the size of the training set. The second kind of data augmentation involves moving the ROI. The ROI of each image is moved up and down by 5% and 10% of the length of the ROI (410 pixels), as shown in Fig. 3. The augmented training set was therefore 10 times the original size of the training set, but the interdependency of the training samples was increased.

Dropout involves dropping out the neuron in the network when the probability output of the neuron decreases below a certain level [21,22]. The effective number of parameters within a neural network is

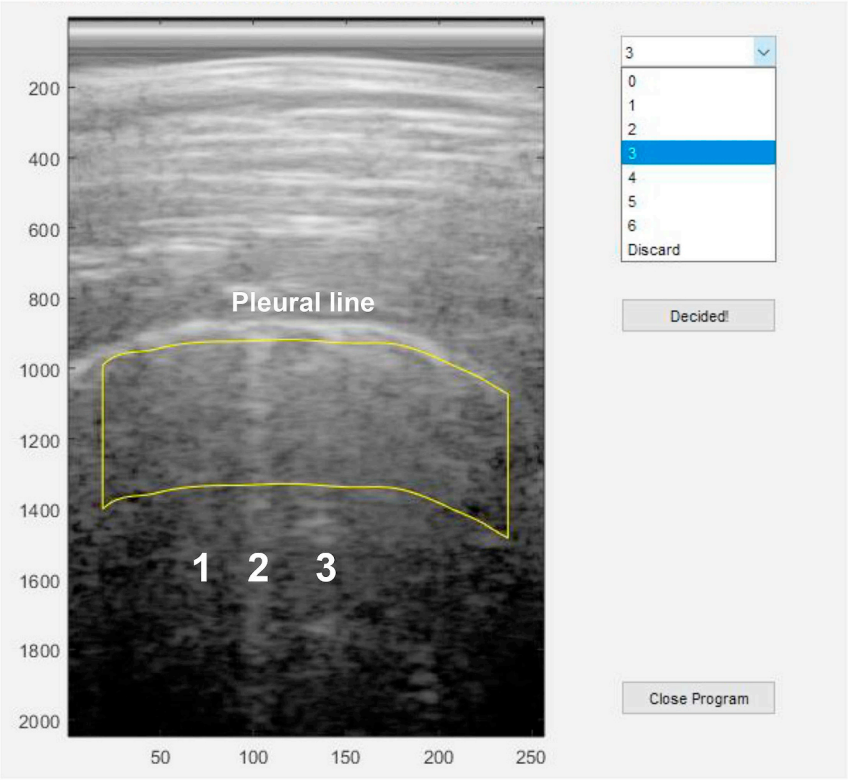

Fig. 1. Dataset Generation. Step 1: Manually count the number of comets present in the ROI of the image (example image: 3 comets). The three comets are shown with a white text overlay. Step 2: Choose the label corresponding to the number of comets present (example image: 3). Step 3: Save the rescaled image with the label.

reduced when dropout is introduced. In this neural network, a dropout level of 0.25 is used across the third and fourth convolutional layers.

The training of the neural network was conducted using MATLAB 2016b (Natick, MA) with MatConvNet [23] beta 20. A standard mini-batch stochastic gradient descent with momentum was used as the minimization algorithm. The batch size was 256. The final network was trained on an Intel Xeon E3-1241 v3 with NVIDIA Quadro K620 GPU with 2 GB graphical memory. The training speed on GPU was about 2 h. The output of the neural network was exported as plain text files and analyzed using OriginPro (OriginLab, Northampton, MA), SPSS (IBM, Armonk, NY) and R [24].

3. Results

Training and testing of the neural network was performed on a 4864-image dataset obtained by the authors under IRB approval. Overfitting was well controlled with the measures taken as written in the methods section. The neural network counted the number of comets the same as the observer ("true value") in 43.4% of the images. Beyond the correctly identified images, there are another 40.8% of images with an identified comet number of true value ± 1. The measurement of intraclass correlation (ICC) for observer difference reveals that the agreement between the human-identified comet counts and the output of the neural network is excellent [25] (ICC = 0.791). Our reported ICC value is based on absolute agreement. This ICC value is superior to the ICC (0.586–0.676) of a single person identifying a single set of lung ultrasound images twice as reported by Gullett et al. [12]. A Bland-Altman plot for comparing the observer against the neural network for the test set is available in Fig. 4. The Bland-Altman plot confirms a relatively low variation between the comet counts identified by hand and by the neural network.

The results are then rearranged for assessing the diagnostic value of the neural network. Researchers have used different standards for categorizing the severity of the presence of the lung ultrasound comets. Frassi et al. [2] have reported that the presence of 30 comets (at 28 scan sites, average 1.07 comets/image) indicates a severe condition; Gullett

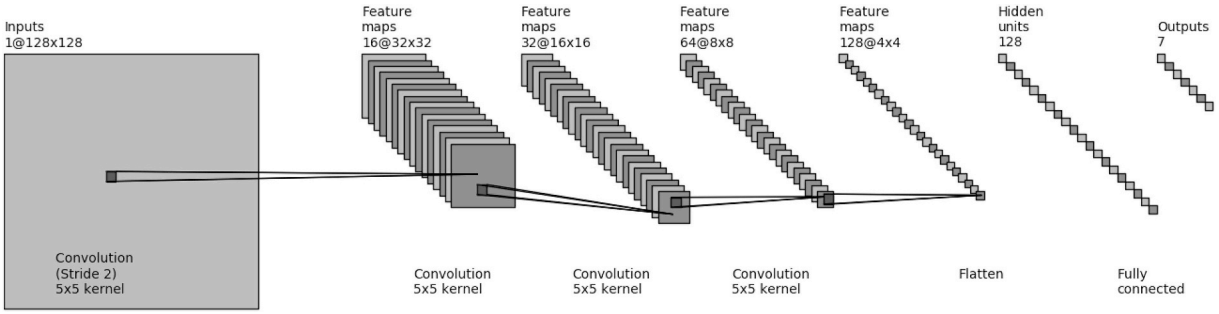

Fig. 2. Structure of the Convolutional Neural Network. The neural network consists of four convolutional layers and a fully connected layer. All convolutional kernels are 5 × 5. Each convolutional layer is accompanied by an overlapping average pooling layer. The fully connected layer routes to one of the possible comet counts in 0–6.

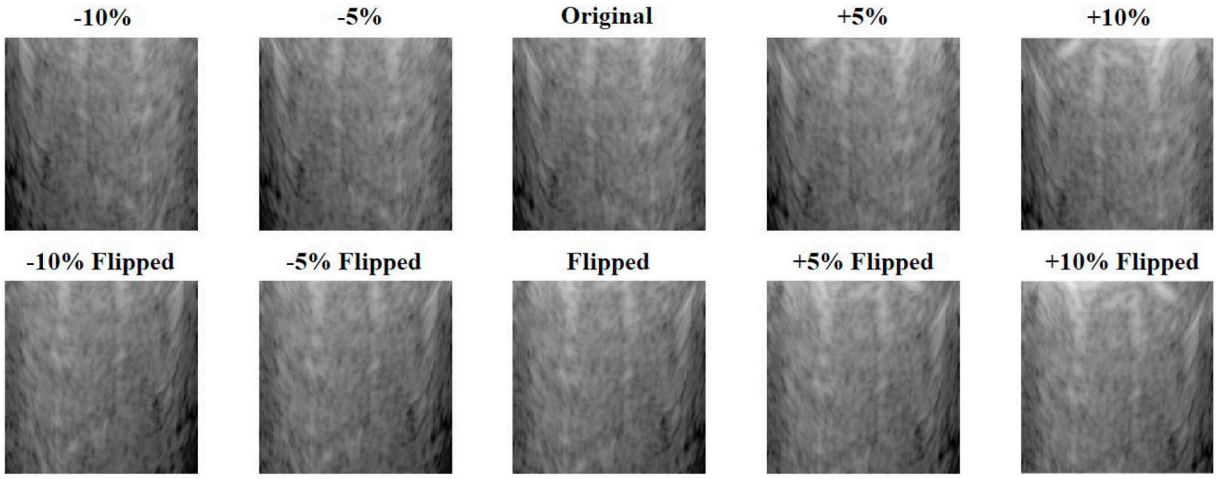

Fig. 3. Data Augmentation by Moving ROI and Flipping. "+" denotes moving down and "-" denotes moving up. Images shown are extracted from the ROI in Fig. 1.

et al. [12] used 3 comets per image as the threshold for a comet-positive image. We will use the latter threshold to determine if an image is comet positive because of our own associated clinical findings. Using this designation, the neural network can categorize 80.8% of images correctly on the test set with a retrained network. From the confusion matrix shown in Table. 1, we achieved a precision of 73.1% and recall of 62.7%. The corresponding F1-score is 0.675.

The neural network was then deployed on a larger level of analysis using the exact same dataset as in our previous paper [9]. This 196-loop clinical dataset includes scan location 1–4 of all subjects. A summary of the clinical parameters and demographics is available in Table. 2. This

Table 1
Breakdown of accuracy in categorization of degree of comets present. 80.8% images were categorized correctly. The precision is 73.1% and recall is 62.7%. The F1-score is 0.675.

True	Test	
	-	+
–	389	47
+	76	128

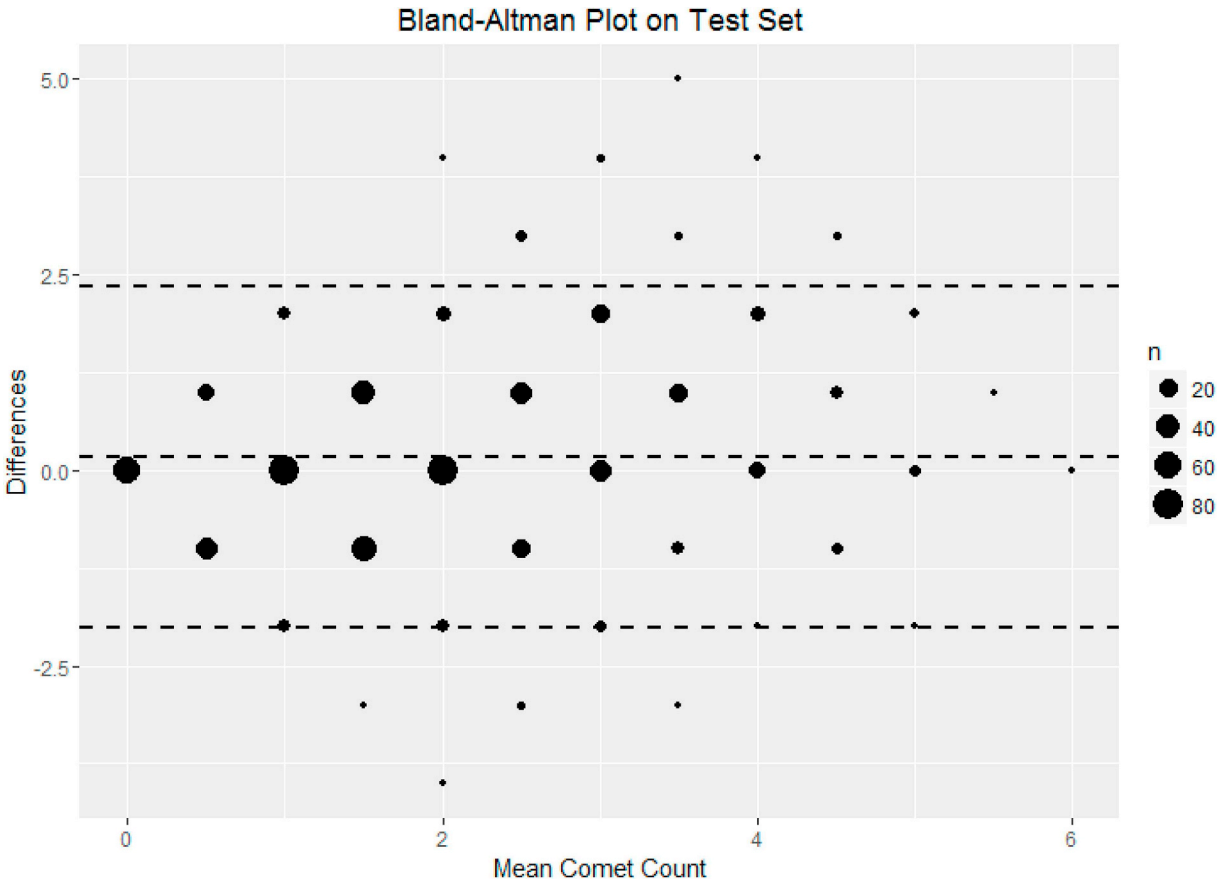

Fig. 4. Bland-Altman Plot for Test Set. Size of circle denotes sample count. Most samples are centered around the Differences = 0 line, meaning that the human observer agrees well with the neural network.

Table 2
Clinical parameters and demographics of the patients.

Clinical Parameter	Range (Mean ± SD)
Age (years)	53 ± 14
Weight (kg)	96 ± 24
BMI	31.5 ± 7.6
Male Sex%	85% (17/20)
Diabetes%	55% (11/20)
Hypertension Documented%	95% (19/20)
Congestive Heart Failure Documented%	5% (1/20)
Coronary Artery Disease Documented%	20% (4/20)
Diastolic Blood Pressure (mmHg)	75 ± 11
Systolic Blood Pressure (mmHg)	134 ± 29
Ejection Fraction%	51 ± 17

dataset has a partial overlap with the trained data. The neural network saw 67 out of 196 loops (34.2%) in training, validation, or test set, while the rest 129 (65.8%) are new to the network. We selected this dataset to serve as cross-verification with respect to our previous results despite an optimal clinical set would have no overlap with the data used in the neural network.

The comet counts by the neural network, averaged to patient level, are analyzed against a series of clinical parameters using linear regression. The comet count exhibits a positive correlation ($p = 0.047$, $r = 0.448$) with respect to diastolic blood pressure and a near-significant negative correlation ($p = 0.061$, $r = -0.513$) with the ejection fraction, as both were confirmed by Weitzel et al. [9]. sIn the meantime, the patients with a higher BMI ($p = 0.009$, $r = -0.566$) were found to have fewer comets. The systolic blood pressure ($p = 0.371$), final blood volume change ($p = 0.468$), and age ($p = 0.575$) were not found to be significant contributors to comet count. We found similar relationships with respect to the number of positive comet images for each patient, especially in diastolic blood pressure ($p = 0.041$, $r = 0.460$) and ejection fraction ($p = 0.090$, $r = -0.470$). A summary of the results is available in Table. 3 and Fig. 5.

Relationship between comet counts and achieved amount of ultrafiltration (UF) is reported in Fig. 6. Achieved UF, which is positively correlated with body weight, can be seen to be firstly positively correlated with comet count. The correlation then reaches a plateau and drops negative. Our hypothesis with respect to this behavior can be seen in the discussion section.

As a proof of concept for clinical use, two *de novo* ultrasound loops are randomly selected from the rest of the 500 loops detailed above and fed into the neural network. The comet counts were also identified by hand. The categorized severity of comets was compared between the output of the neural network and hand counting and the results can be seen in Fig. 7. The agreement is 68.8% on the first loop and 100% on the second loop. On a loop level, hand counting indicated 11 (−) and 21 (+) frames compared to 9 (−) and 23 (+) frames from the neural network for the first loop, both confirming that the first loop is comet dense. The second loop has 32 (−) and 0 (+) frames when counted by hand and by neural network, suggesting a comet negative loop. The clinical outcome predicted by the neural network closely resembles the outcome identified by hand.

Table 3
Linear regression of comet count (Count) and number of comet positive images (Comet+) versus relevant clinical parameters. BP: Blood pressure. %BV: Percent blood volume. EF: Ejection fraction. BMI: body mass index.

Clinical Parameter	Unit	Range	R-value	p-value (Count)	Sig.	p-value (Comet+)	Sig.
Diastolic BP	mmHg	75 ± 11	0.448	0.047	*	0.041	*
Systolic BP	mmHg	134 ± 29	0.211	0.370	NS	0.446	NS
Final %BV	Percent	−8.5 ± 4.7	−0.172	0.468	NS	0.511	NS
EF %	Percent	51 ± 17	−0.513	0.061	NS	0.090	NS
Subject Age	years	53 ± 14	−0.133	0.575	NS	0.441	NS
Subject BMI	kg/m2	31.5 ± 7.6	−0.566	0.009	**	0.149	NS

Significance levels (Sig.): NS - not significant. * - $p < 0.05$. ** - $p < 0.01$. *** - $p < 0.001$.

4. Discussion

The lung comet dataset labeled in this study is the largest to date by an order of magnitude compared to the 50–100 images used in existing studies [13,16]. The database contains images collected from a mechanically scanned single-element transducer, which is representative in a point-of-care and telehealth ultrasound setting, although low cost array imaging systems capable of higher image quality were introduced by Phillips in 2015 and have been evolving. This neural network is a promising approach since it has reasonable accuracy and a low number of parameters. The lower number of parameters leads to increased practicality of implementation since this translates into less computational time and computing power requirements as discussed below. The 43.4% absolute accuracy means that in 43.4% of the images, the neural network identified the same number of lung comets as the observer. Despite an apparently low 43.4% absolute accuracy in the test set, the neural network reached a 0.791 absolute ICC, which indicates a low deviation from the true comet count if the comet count may be incorrect in absolute measures. This ICC suggests a higher level of agreement of comet counting between human observers in this study and neural network than that of a single observer counting the comets twice as reported in the literature [12]. The excellent ICC level reveals the potential of using neural network as the “gold standard” in identifying lung comets. In addition, the neural network can still use more training samples and the accuracy and ICC can be further increased with more data. We decided to limit the current study to 4896 images because of time limitations for the investigators in labeling the images to train the algorithm. With increased data, and additional time resources, a larger neural network may be used to train on the available data, which may further increase the accuracy. However, the accuracy will be limited by the intrinsic variations in comet counting by human. We conjecture that further studies using a combination of lung ultrasound and other available clinical information, such as CT information and patient's medical history, may allow greater pace to be made in improving accuracy within diagnostic tolerance.

Categorizing the ultrasound images into comet negative and positive adds a layer of supplemental information that may be diagnostically relevant in assessing pulmonary edema. In clinical practice, the presence or absence of comets may be helpful in determining when a patient it at the ideal fluid status (aka “dry weight”), or these categories may relate to how far a patient is from their dry weight.

In the clinical dataset, the clinical findings agree well with our previous results. As in our previous study [9], these neural network findings also suggest that a higher diastolic blood pressure and a lower ejection fraction are correlated with higher comet counts. Patients with fluid overload generally present with a higher blood pressure, and the excess fluid may manifest itself with increased fluid in the lungs and corresponding higher number of comets seen on ultrasound scans. Likewise, a low ejection fraction may be accompanied by excess lung edema and more comets. We hypothesize that the observation of higher BMI being associated with lower comet counts may be attributed to attenuation of the ultrasound signal by adipose tissue. This hypothesis requires further study but is supported by the following. The number of comets is expected to increase with the obtained ultrafiltration (UF) amount, i.e. water obtained during dialysis sessions; this behavior can be seen initially but quickly decreases with increased UF amount, which is positively correlated with the body weight of the subjects. This observation suggests difficulty for the neural network, and potentially for the observer, to see comets in overweight patients. All these

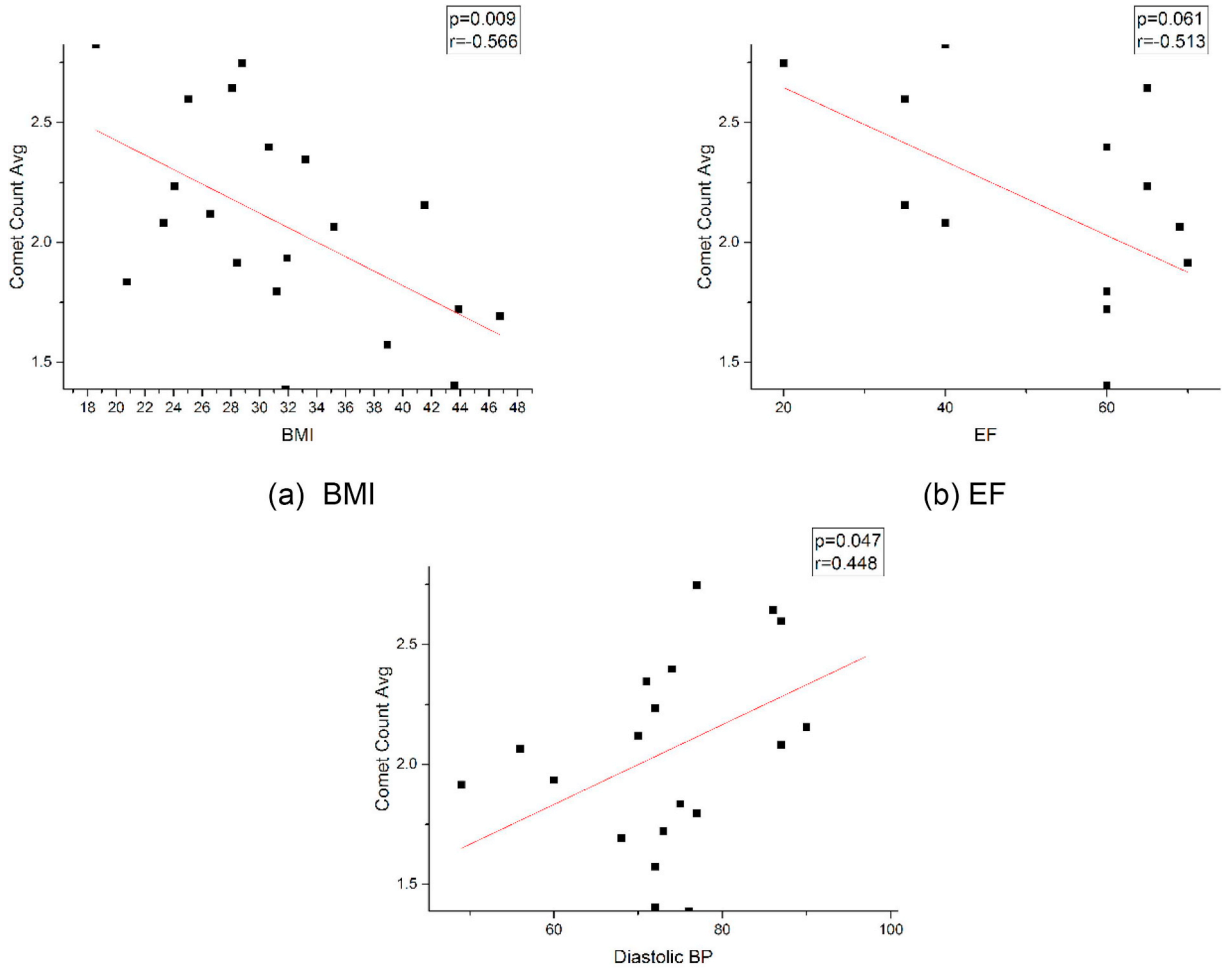

Fig. 5. (a) Linear Regression of BMI ($p = 0.009$, $r = -0.566$), (b) Diastolic Blood Pressure ($p = 0.047$, $r = 0.448$), and (c) Ejection Fraction ($p = 0.061$, $r = -0.513$) against Comet Count.

observations are suggestive and require further investigation.

Potential future deployment will likely benefit from the relatively low number of parameters that this neural network possesses. With the current structure and its 5.3×10^5 parameters, this neural network requires little computing power. A test on an Intel i5-6260U computer reveals that the comet identification and clinical classification of a full loop (32 frames) takes 0.11–0.12 s. The requirement for memory and storage is on the order of megabytes. These low requirements ensure

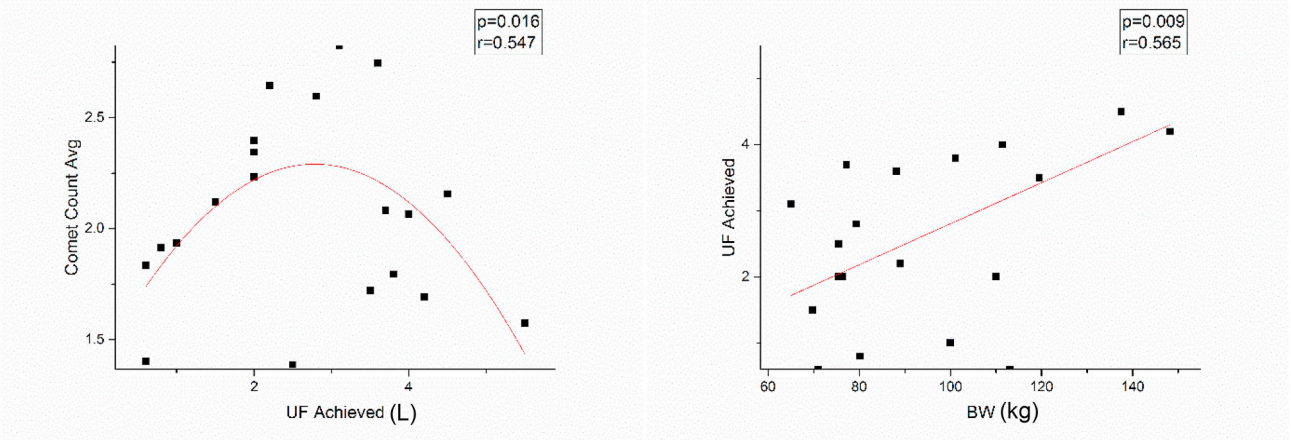

Fig. 6. Left: a quadratic fit of the amount of ultrafiltration (UF), in liters, achieved with respect to the comet count. They are initially positively correlated until it quickly drops off. The p-value shown is the p-value of the leading quadratic term. The quadratic fit is statistically significant. Right: The amount of UF achieved positively correlates with the body weight of the subject in kilograms ($p = 0.009$).

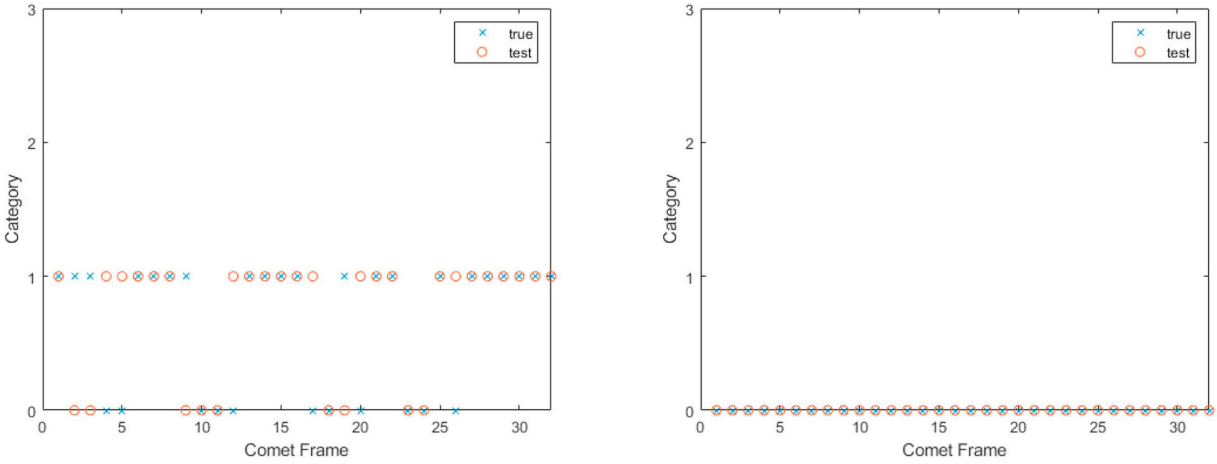

Fig. 7. Comparison of Categorized Comet Severity for New Data. 0 = Negative, 1 = Positive. Both measures agree on the comet-dense nature of the first loop with 22/32 (68.8%) agreement on frames. Both measures find the second loop to be comet negative, with 32/32 (100.0%) agreement on frames. No severe misidentification happened in either loop. On a loop level:
Loop 1: Hand 11 (−)/21 (+) vs Neural network 9 (−)/23 (+).
Loop 2: Hand 32 (−)/0 (+) vs Neural network 32 (−)/0 (+).

that most current computing systems will be able to handle this neural network and real-time comet identification is feasible. If deployed on a cloud level, this neural network will require a very small amount of network data transmission, which makes it suitable for potential use in telehealth and Internet applications. A fully automated comet identification system can incorporate another module with automatic ROI identification using either image processing or neural network methods. A fully-automated system adds little tie and effort to the current clinical practice of lung comet identification, potentially be less burdensome than manual comet counting, and has great potential to be a clinical diagnostic tool for lung water assessment.

While this neural network exhibits many promising characteristics, there are some limitations exist that require further study.

First, the size of the database contains 4864 ultrasound images. It is the largest lung comet database to date but is still comparatively small compared to a big database such as ImageNet. The number of parameters still overwhelms the number of training samples. The distribution of the comet counts in the database is also somewhat uneven. The distribution of the number of images with respect to its labels ("comet counts") are attached in Table 4.

Second, the neural network has performed reasonably well on the images with low or medium comet counts (0–3, 48.3% correct) but disagreed with reference measures on the images with high comet counts (4–6, 19.4% correct). This is mainly due to the scarcity of supporting samples in the training set for comet-rich samples. In fact, it is very difficult for an artificial neural network to learn from a small pool of samples [26]. Collecting and using more data in a larger database can mitigate the low accuracy on comet-rich samples. With a much larger database, there would be enough comet-rich samples to train from; the samples can also be randomly pooled at a loop-level instead of an image-level during data split to make the training set and the test set more independent of each other. In this way, loop-level information can be generated and compared to clinical data. A further expansion for future studies would be to obtain lung ultrasound scans from a wide range of patients and pool the data on a patient level.

Table 4
Distribution of comet count in the comet database.

Label (Count)	# Images	Percentage (%)
0	748	15.4
1	1248	25.7
2	1225	25.2
3	883	18.2
4	471	9.7
5	201	4.1
6	88	1.8

Third, the database was labeled by a single observer. Thus, the neural network behaves like this observer and contains this observer's subjective tendencies. The comet counting is also subject to observer fluctuations as suggested in a prior publication [12]. Because of the inherent difficulties and visual variations in identifying comets, a training set with input from multiple observers will help reduce subjectivity. Ideally, a pool of physicians will be recruited to perform comet identification in future studies so that the inter-observer variation can be minimized; this too will be the subject of future studies.

Finally, the approach to neural network learning, in this case, can also be refined. Instead of labeling the ultrasound image, individual comets can be labeled along with information on their positions. The neural network can then be tuned to learn from individual comets and identify the comets individually when deployed. This approach can potentially reveal information on the position of the comets and track their movement. A subset of neural networks called U-Net [27] have been applied to biomedical imaging and we are actively pursuing this path.

5. Conclusion

Although increasing evidence suggests that lung ultrasound comets may be clinically valuable in assessing pulmonary edema, the objective quantification of these comets has been very challenging. We performed a pilot study to test the feasibility of using a neural network system to perform lung ultrasound counting. A corresponding database of 4896 ultrasound images was built for this system as the learning source. This machine learning system achieved a 43.4% accuracy in absolute measures and a 0.791 absolute ICC. The apparent low absolute accuracy is reflective of the low inter-measurement agreement of the comet count by a single observer re-reading the same frames. The high ICC level indicates substantial agreement of the neural network algorithm with the comet quantity as counted by a human. With the same sets of parameters, the system categorized 80.8% of the images correctly for the severity of the lung ultrasound comets when divided into a positive and a negative group. With a relatively low number of parameters, this neural network approach shows significant potential in medical decision support systems while requiring only modest computing power. The potential clinical utility of the neural network for lung comet

assessment was illustrated by showing correlation between comet count and diastolic blood pressure as well as ejection fraction. Future work on this neural network will involve using more data to improve accuracy, tracking the movement of the comets, testing inter-observer variation, and assessing clinical parameters, to further test the capabilities of neural networks in quantifying lung ultrasound comets.

Conflicts of interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgment

The authors would like to thank all the patients who participated in the data collection; Xinchen Yan for help in establishing the neural network; Dr. John Pitre, Dr. Robinson Seda, and Jonah Harmon for proofreading this paper. This work is supported in part by a grant from the Renal Research Institute (RRI). The authors declare no conflict of interest.

References

[1] Z. Jambrik, S. Monti, V. Coppola, et al., Usefulness of ultrasound lung comets as a nonradiologic sign of extravascular lung water, Am. J. Cardiol. 93 (10) (2004) 1265–1270, https://doi.org/10.1016/j.amjcard.2004.02.012.

[2] F. Frassi, L. Gargani, P. Tesorio, et al., Prognostic value of extravascular lung water assessed with ultrasound lung comets by chest sonography in patients with dyspnea and/or chest pain, J. Card. Fail. 13 (10) (2007) 830–835, https://doi.org/10.1016/j.cardfail.2007.07.003.

[3] L. Gargani, F. Frassi, G. Soldati, et al., "Ultrasound lung comets for the differential diagnosis of acute cardiogenic dyspnoea: a comparison with natriuretic peptides☆, Eur. J. Heart Fail. 10 (1) (2014) 70–77, https://doi.org/10.1016/j.ejheart.2007.10.009.

[4] D. Lichtenstein, G. Mezière, P. Biderman, et al., The comet-tail artifact: an ultrasound sign ruling out pneumothorax, Intensive Care Med. 25 (4) (1999) 383–388, https://doi.org/10.1007/s001340050862.

[5] E. Picano, F. Frassi, E. Agricola, et al., Ultrasound lung comets: a clinically useful sign of extravascular lung water, J. Am. Soc. Echocardiogr. 19 (3) (2006) 356–363, https://doi.org/10.1016/j.echo.2005.05.019.

[6] D. Litchenstein, G. Mezière, P. Biderman, et al., The comet-tail artifact: an ultrasound sign of alveolar-interstitial syndrome, Am. J. Respir. Crit. Care Med. 156 (5) (1997) 1640–1646, https://doi.org/10.1164/ajrccm.156.5.96-07096.

[7] C. Zoccali, C. Torino, R. Tripepi, et al., Pulmonary congestion predicts cardiac events and mortality in ESRD, J. Am. Soc. Nephrol: JASN. 24 (4) (2013) 639–646, https://doi.org/10.1681/ASN.2012100990.

[8] F. Mallamaci, F.A. Benedetto, R. Tripepi, et al., Detection of pulmonary congestion by chest ultrasound in dialysis patients, Cardiovasc. Imag. 3 (6) (2010) 586–594, https://doi.org/10.1016/j.jcmg.2010.02.005.

[9] W.F. Weitzel, J. Hamilton, X. Wang, et al., Quantitative lung ultrasound comet measurement: method and initial clinical results, Blood Purif. 39 (1–3) (2015) 37–44, https://doi.org/10.1159/000368973.

[10] V. Panuccio, G. Enia, R. Tripepi, et al., Chest ultrasound and hidden lung congestion in peritoneal dialysis patients, Nephrol. Dial. Transplant.:Off. Publ. Eur. Dial. Transpl. Assoc. Eur. Ren. Assoc. 27 (9) (2012) 3601–3605, https://doi.org/10.1093/ndt/gfs116.

[11] G. Soldati, R. Copetti, S. Sher, Can lung comets be counted as "objects"? Cardiovasc. Imag. 4 (4) (2011) 438–439, https://doi.org/10.1016/j.jcmg.2010.09.025.

[12] J. Gullett, J.P. Donnelly, R. Sinert, et al., Interobserver agreement in the evaluation of B-lines using bedside ultrasound, J. Crit. Care 30 (6) (2015) 1395–1399, https://doi.org/10.1016/j.jcrc.2015.08.021.

[13] L.J. Brattain, B.A. Telfer, A.S. Liteplo, et al., Automated B-line scoring on thoracic sonography, J. Ultrasound Med.: Off. J. Am. Inst. Ultrasound Med. 32 (12) (2013) 2185–2190, https://doi.org/10.7863/ultra.32.12.2185.

[14] R. Moshavegh, K. Lindskov Hansen, H. Møller Sørensen, et al., Novel automatic detection of pleura and B-lines (comet-tail artifacts) on in vivo lung ultrasound scans, in: N. Duric, B. Heyde (Eds.), SPIE Medical Imaging, SPIE, 201697900K.

[15] N. Anantrasirichai, M. Allinovi, W. Hayes, et al., "Automatic B-line detection in paediatric lung ultrasound," in 2016, IEEE International Ultrasonics Symposium (IUS) (2016) 1–4.

[16] N. Anantrasirichai, W. Hayes, M. Allinovi, et al., Line detection as an inverse problem: application to lung ultrasound imaging, IEEE Trans. Med. Imag. 36 (10) (2017) 2045–2056, https://doi.org/10.1109/TMI.2017.2715880.

[17] D.D. Cox, T. Dean, Neural networks and neuroscience-inspired computer vision, Curr. Biol. 24 (18) (2014) R921–R929, https://doi.org/10.1016/j.cub.2014.08.026.

[18] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: F. Pereira, C.J.C. Burges, L. Bottou, et al. (Eds.), Advances In Neural Information Processing Systems, vol. 25, Curran Associates, Inc, 2012, pp. 1097–1105.

[19] M.A. Tanner, W.H. Wong, The calculation of posterior distributions by data augmentation, J. Am. Stat. Assoc. 82 (398) (1987) 528–540.

[20] A. Karpathy, G. Toderici, S. Shetty, et al., Large-scale video classification with convolutional neural networks, The IEEE Conference On Computer Vision And Pattern Recognition (CVPR), 2014.

[21] G.E. Hinton, N. Srivastava, A. Krizhevsky, et al., Improving Neural Networks by Preventing Co-adaptation of Feature Detectors, (2012) arXiv preprint arXiv:1207.0580.

[22] N. Srivastava, G.E. Hinton, A. Krizhevsky, et al., Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.

[23] A. Vedaldi, K. Lenc, MatConvNet - convolutional neural networks for MATLAB, in: X. Zhou, A.F. Smeaton, Q. Tian, et al. (Eds.), in Proceeding Of the {ACM} Int. Conf. On Multimedia, 2015, pp. 689–692.

[24] R Core Team, "R: A Language and Environment for Statistical Computing", https://www.R-project.org.

[25] D.V. Cicchetti, Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology, Psychol. Assess. 6 (4) (1994) 284–290, https://doi.org/10.1037/1040-3590.6.4.284.

[26] N. Pinto, D.D. Cox, J.J. DiCarlo, Why is real-world visual object recognition hard? PLoS Comput. Biol. 4 (1) (2008), https://doi.org/10.1371/journal.pcbi.0040027 e27.

[27] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, International Conference On Medical Image Computing And Computer-Assisted Intervention, 2015, pp. 234–241.