Ultrasound in Med. & Biol., Vol. 00, No. 00, pp. 1–11, 2019

0301-5629/$ - see front matter

https://doi.org/10.1016/j.ultrasmedbio.2019.04.009

● *Original Contribution*

POINT SHEAR WAVE ELASTOGRAPHY USING MACHINE LEARNING TO DIFFERENTIATE RENAL CELL CARCINOMA AND ANGIOMYOLIPOMA

Hersh Sagreiya MD* Alireza Akhbardeh PhD* Dandan Li, MD, PhD* Rosa Sigrist MD*
Benjamin I. Chung MD† Geoffrey A. Sonn MD*,† Lu Tian PhD‡
Daniel L. Rubin MD, MS*,§,¶ and Jürgen K. Willmann, MD*

* Department of Radiology, Stanford University School of Medicine, Stanford, CA, USA; † Department of Urology, Stanford University School of Medicine, Stanford, CA, USA; ‡ Department of Health, Research & Policy, Stanford University, Stanford, CA, USA; § Department of Biomedical Data Science, Stanford University, Stanford, CA, USA; and ¶ Department of Medicine (Biomedical Informatics Research), Stanford University, Stanford, CA, USA

(*Received* 5 *October* 2018; *revised* 31 *March* 2019; *in final from* 3 *April* 2019)

Abstract—The question of whether ultrasound point shear wave elastography can differentiate renal cell carcinoma (RCC) from angiomyolipoma (AML) is controversial. This study prospectively enrolled 51 patients with 52 renal tumors (42 RCCs, 10 AMLs). We obtained 10 measurements of shear wave velocity (SWV) in the renal tumor, cortex and medulla. Median SWV was first used to classify RCC versus AML. Next, the prediction accuracy of 4 machine learning algorithms—logistic regression, naïve Bayes, quadratic discriminant analysis and support vector machines (SVMs)—was evaluated, using statistical inputs from the tumor, cortex and combined statistical inputs from tumor, cortex and medulla. After leave-one-out cross validation, models were evaluated using the area under the receiver operating characteristic (ROC) curve (AUC). Tumor median SWV performed poorly (AUC = 0.62; $p = 0.23$). Except logistic regression, all machine learning algorithms reached statistical significance using combined statistical inputs (AUC = 0.78–0.98; $p < 7.1 \times 10^{-5}$). SVMs demonstrated 94% accuracy (AUC = 0.98; $p = 3.13 \times 10^{-6}$) and clearly outperformed median SWV in differentiating RCC from AML ($p = 2.8 \times 10^{-4}$). (E-mail: dlrubin@stanford.edu)

Key Words: **Ultrasound, Point shear wave elastography, Renal cell carcinoma, Angiomyolipoma, Machine learning.**

INTRODUCTION

Frequent imaging has increased the detection and need for characterizing solid renal lesions. Angiomyolipoma (AML) is the most common benign solid renal neoplasm (Jinzaki et al. 2014), and renal cell carcinoma (RCC) accounts for nearly 90% of all renal malignancies (Qayyum et al. 2013). Although RCC is usually managed with partial or radical nephrectomy, AML is typically observed or embolized. Hence, differentiating between AML and RCC is crucial (Siegel et al. 1996). Otherwise, unnecessary surgery may be performed in patients later found to have benign AML. On B-mode ultrasound, AML is classically hyperechoic with a well-circumscribed margin and acoustic shadowing. However, minimal-fat AML may appear isoechoic (Park 2017). Although RCC is most commonly hypoechoic, 32% of RCCs under 3 cm have been found to be hyperechoic, mimicking AML (Forman et al. 1993). One study showed that the diagnostic accuracy of grayscale ultrasound for small solid lesion characterization was 42%, increasing to 78% with power Doppler (Jinzaki et al. 1998). Hence, RCC and AML can be difficult to distinguish on B-mode ultrasound. These lesions are typically either followed clinically or further evaluated *via* computed tomography (CT) or magnetic resonance imaging (MRI). As part of the standard ultrasound examination, elastography could help characterize these common incidental hyperechoic lesions, minimizing unnecessary follow-up exams and reducing costs.

Ultrasound elastography non-invasively assesses tissues' mechanical properties (Sigrist et al. 2017).

Address correspondence to: Daniel L. Rubin, MD, Department of Radiology, School of Medicine, Stanford University, 1265 Welch Road, Room X-335, MC 5464, Stanford, CA 94305-562, USA. E-mail: dlrubin@stanford.edu

Strain elastography, an earlier technology, uses tissue displacement in response to compression from the ultrasound transducer to generate a strain elastogram, demonstrating relative tissue stiffness (Garra 2015). Point shear wave elastography (pSWE), a more recent technology, uses an acoustic radiation force impulse to transmit controllable longitudinal forces, deforming the tissue and generating transverse shear waves (Nowicki and Dobruch-Sobczak 2016). The transducer detects shear wave velocity (SWV) to measure tissue stiffness. Advantages of pSWE over strain elastography include less operator dependence and quantitative SWV measurements.

The results are conflicting and the experience limited using strain (Keskin et al. 2015; Onur et al. 2015; Tan et al. 2013) and pSWE (Goya et al. 2015; Lu et al. 2015) to classify between AML and RCC. Using pSWE, one study showed decent performance with 88% sensitivity and 54% specificity (Goya et al. 2015), another demonstrated only 48% sensitivity and 33% specificity (Lu et al. 2015). Recently, there has been increasing interest in using machine learning in radiology (Erickson et al. 2017). Machine learning algorithms can make autonomous predictions (Kohli et al. 2017) and detect complex patterns imperceptible to humans (Lakhani et al. 2018).

In the field of ultrasound elastography, machine learning has been used for chronic liver disease diagnosis (Gatos et al. 2017), fibrosis assessment in hepatitis B patients (Chen et al. 2017), fibrosis assessment in hepatitis C patients (Fujimoto et al. 2013; Stoean et al. 2011), breast cancer diagnosis (Zhang et al. 2016) and thyroid nodule diagnosis (Ma et al. 2010). For renal lesions, machine learning using texture analysis on CT images has been used to differentiate fat-poor AML from RCC (Feng et al. 2018; Hodgdon et al. 2015), as well as different RCC subtypes (Kocak et al. 2018). Deep learning on CT images was used to differentiate RCC from oncocytoma (Coy et al. 2019), as well as RCC from other benign entities such as AML and cysts (Zhou et al. 2019).

Earlier work has been done on the application of machine learning to B-mode renal ultrasound. For instance, a prior study used a support vector machine to distinguish the following classes on B-mode renal ultrasound with 86% accuracy: normal, medical renal disease and cyst (Subramanya et al. 2015). In fact, multiple studies have sought to differentiate these 3 classes using various techniques, such as the dominant Gabor wavelet features (Raja et al. 2010) and a hybrid fuzzy-neural system (Raja et al. 2008). However, machine learning has not been applied for solid renal lesion characterization with ultrasound elastography. This is of key clinical utility, as these lesions are often incidentally found on initial ultrasound, and their characterization could potentially minimize the time, expense and patient anxiety associated with follow-up examinations.

The purpose of this study is to demonstrate that pSWE is accurate in differentiating between RCC and AML and that machine learning algorithms can better make this distinction than median shear wave velocity.

MATERIALS AND METHODS

Patient population

This prospective, single-center study was approved by the Stanford University Institutional Review Board (Stanford, CA) and compliant with the Health Insurance Portability and Accountability Act. All patients provided signed informed consent. From February 2014 to May 2016, patients scheduled for renal surgery—who were diagnosed with a solid renal mass—were enrolled, as well as additional patients with confirmed AML based on CT and MRI. A total of 58 patients consented. The following patients were excluded (Fig. 1): renal tumor other than RCC and AML (oncocytoma); RCC not confirmed by surgical pathology; AML not confirmed by pathology, CT or MRI; failure to undergo elastography (one for tumor depth, one for tumor excision before elastography could be performed); incomplete pathology (RCC and AML in the same kidney so pathology could not be matched one to one); or incomplete measurements (medulla not measured). After applying the exclusion criteria, 51 patients with 52 renal tumors remained. All RCCs were confirmed by surgical pathology. A total of 3 AMLs were confirmed by CT, 5 were confirmed by MRI and 2 were confirmed by surgery. The study included 42 RCCs and 10 AMLs (Table 1). There were 33 males (28 RCC, 5 AML) and 18 females (13 RCC, 5 AML). The average age was 57.0 ± 13.0 y (range: 16−79) for AML and 56.3 ± 7.6 y (range: 26−84) for RCC. Table 2 summarizes RCC subtypes.

Image acquisition

A certified sonographer or a radiologist (all with at least 18 mo of experience in clinical ultrasound elastography) performed each exam, using an Acuson S2000 (Siemens Medical Solutions, Mountain View, CA, USA) ultrasound system equipped with the Virtual Touch Tissue Quantification mode and a 6C1 convex array transducer (Sigrist et al. 2017). Patients were placed in the neutral or decubitus position, performing breath hold to reduce motion. Tumors were placed in the center of B-mode images, minimizing the distance between the transducer and tumor. The region of interest (ROI) was placed in solid portions of lesions, avoiding cystic areas or calcifications. The average depth of renal lesions was 5.7 cm (range: 2.7−8), of cortical measurements was

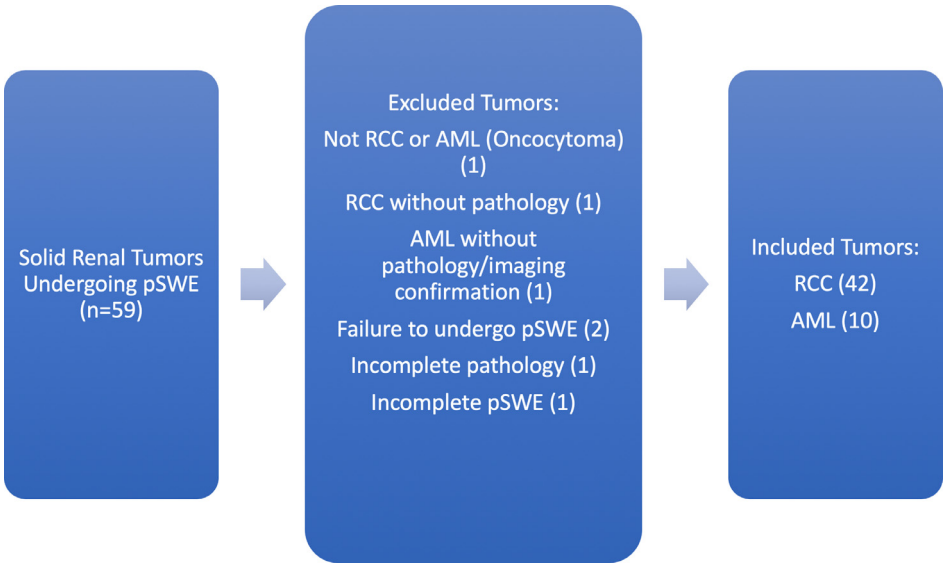

Fig. 1. Tumors excluded from analysis. Out of the original 59 solid renal tumors, 52 were included and 7 were excluded. pSWE = point shear wave elastography; RCC = renal cell carcinoma; AML = angiomyolipoma.

5.2 cm (range: 2.9−8) and of medullary measurements was 5.7 cm (range: 2.9−8). The average difference in depth between renal lesions and both cortical and medullary measurements was 1.0 cm. Image quality was optimized by individually adjusting imaging parameters (depth, focus, gain). The pSWE technique was subsequently performed as described in Goya et al. (2015) and in Lu et al. (2015). Keeping the transducer still, a fixed box-shaped ROI (10 mm axial by 6 mm lateral) was placed in the tumor, cortex and medulla, and 10 consecutive valid measurements of SWV were obtained at each location (Fig. 2). Examples of AML and RCC are presented in Fig. 3.

Table 1. Summary of the patient population: Comparing the AML and RCC groups

Characteristics	Angiomyolipoma	Renal Cell Carcinoma
Patients ($n = 51$)	10	41
Age (y)	57.0 ± 13.0 [16−79]	56.3 ± 7.6 [26−84]
Gender: Male ($n = 33$)	5	28
Gender: Female ($n = 18$)	5	13
Diameter (mm)	22.3 ± 22.5 [10−85]	34.8 ± 14.4 [14−87]
Location: Right ($n = 32$)	7	25
Location: Left ($n = 20$)	3	17

Note: Values in brackets represent ranges.
AML = angiomyolipoma; RCC = renal cell carcinoma.

Table 2. Types of tumors included in the study population

Group	*n*	Percent (%)
Total	52	100
Renal cell carcinoma	42	80.8
Clear cell type	32	61.5
Papillary type	6	11.5
Chromophobe type	3	5.8
Unclassified	1	1.9
Angiomyolipoma	10	19.2

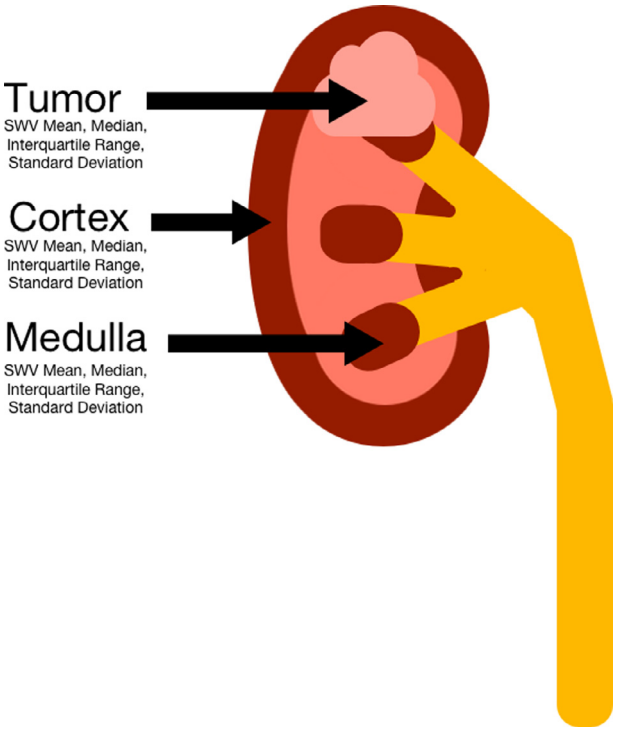

Fig. 2. Parameters under study. For each patient, 10 measurements of shear wave velocity were performed in the tumor, renal cortex and renal medulla. For each region (tumor, cortex and medulla), the mean, median, interquartile range and standard deviation of the 10 measurements of shear wave velocity was obtained.

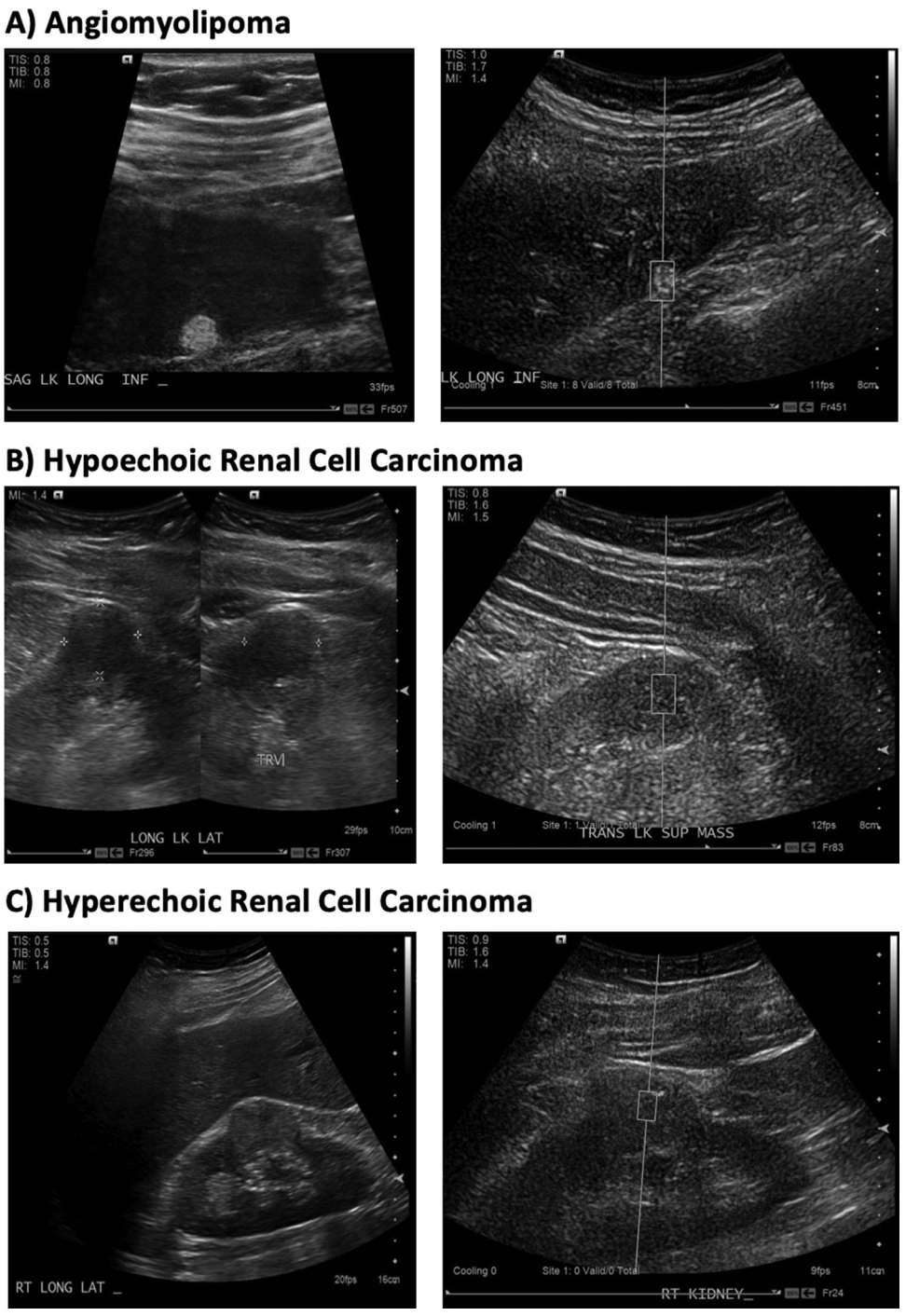

Fig. 3. (a) B-mode and corresponding ultrasound elastography image of an angiomyolipoma. (b) B-mode and elastography image of a hypoechoic renal cell carcinoma. (c) B-mode and elastography image of a hyperechoic renal cell carcinoma.

Tumor diameters and SWV measurements were retrieved from the picture archiving and communication system (Centricity, GE Healthcare, Chicago, IL, USA). All 10 measurements of SWV were not always available from each location. A total of 91.2% of cases involved exactly 10 measurements, 5.4% of cases involved fewer and 3.4% of cases involved more. The statistical features that ultimately served as inputs in machine learning models were the mean, median, interquartile range and standard deviation of either the 10 measurements of SWV or the number of measurements available when fewer. Additional information (age, gender, pathology

diagnosis) was obtained from patient records (Epic Systems Corporation, Verona, WI, USA).

Statistical analysis

All data analysis was performed in Matlab R2015b (MathWorks, Natick, MA, USA). We first used median tumor SWV to classify between RCC and AML. Median tumor SWV and the true class labels were input into the Matlab *perfcurve* function, which recursively examined performance at various thresholds to generate a receiver operating characteristic (ROC) curve, used to calculate the area under the curve (AUC). Next, tumor-to-cortex shear wave ratio (SWR) and tumor-to-medulla SWR were evaluated in their ability to differentiate RCC and AML, again using *perfcurve* to generate ROC curves and AUC.

Machine learning

The classification accuracy of 4 supervised machine learning algorithms for distinguishing between RCC versus AML was compared, using four statistical measures of SWV (mean, median, interquartile range and standard deviation) as inputs: quadratic discriminant analysis (Guo et al. 2007), logistic regression (Dobson 1990), naïve Bayes (Hastie et al. 2009) and a non-linear support vector machine (Schölkopf and Smola 2002). We selected these algorithms because they are commonly used in the literature. Quadratic discriminant analysis used a pseudo-quadratic transformation. Support vector machines used the Gaussian radial basis function kernel. The Appendix presents additional detail about each algorithm.

The machine learning models were separately run, using 4 sets of inputs. The first set included the 4 statistical measurements (mean, median, interquartile range and standard deviation) in the tumor alone; the second set included the 4 statistical measures in the cortex alone; the third set included the 4 statistical measures in the medulla alone; and the fourth set included all 4 statistical measures within the tumor, cortex and medulla, for a total of 12 features.

Validation was performed using leave-one-out cross validation (Hastie et al. 2009). During each run, the training data were used to train the model, and the Matlab *predict* function applied this model to the validation data point to output a score representing the likelihood (posterior probability) that the label came from each class, RCC or AML. Next, the Matlab *perfcurve* function used the scores and true class labels to generate ROC curves, subsequently used to calculate AUC, sensitivity, specificity, positive and negative predictive value and accuracy. Hence, the reported ROC AUC values are from cross validation. Next, the distribution of scores was compared between RCC and AML *via* a Wilcoxon rank-sum test, with p values demonstrating the strength of class separation. Finally, the statistical significance of the difference in AUC between ROC curves generated by different models was calculated using the DeLong method (DeLong et al. 1988). Supplementary Figure 1 details the machine learning workflow.

RESULTS

Using the median value of 10 SWV measurements did not show significant differentiation of RCC from AML with a ROC AUC of 0.62 ($p = 0.23$; Table 3). The tumor-to-cortex SWR also did not perform well, with ROC AUC of 0.64. The tumor-to-medulla ratio had moderate performance with ROC AUC of 0.72. Supplemental Figure 2 shows that the distribution of SWV measurements for AML and RCC was different, as SWV measurements are more clustered toward the median for AML than for RCC. This information would not be incorporated by using median SWV alone.

Support vector machines demonstrated the highest level of performance and represented the only machine learning algorithm that showed statistically significant separation of scores *via* the Wilcoxon rank-sum test, using measurements in the lesion alone (AUC = 0.94; $p = 4.6 \times 10^{-3}$), the cortex alone (AUC = 0.79; $p = 2.3 \times 10^{-5}$), or the medulla alone (AUC = 0.84; $p = 1.1 \times 10^{-3}$; Table 4). Moreover, using the combination of features within the tumor, cortex and medulla resulted in the highest level of performance for each machine-learning algorithm. In fact, using all 12 statistical features, all machine learning algorithms demonstrated statistically significant performance (AUC = 0.78−0.98; $p < 7.1 \times 10^{-3}$) in separating RCC from AML, except logistic regression (AUC = 0.71; $p = 0.099$; Table 4).

Analyzing score separation between RCC and AML in particular, the median value of SWV within the tumor, cortex and lesion, respectively, did not demonstrate adequate score separation between the 2 classes (Fig. 4a). With support vector machines, there was improved score separation between RCC and AML either using the 4

Table 3. Performance of traditional measures and machine learning in classifying between RCC and AML: Traditional measures (individual values and ratios)

Region	Value	ROC AUC
Tumor SWV	Median values	0.62
Cortex SWV		0.40
Medulla SWV		0.28
Tumor to cortex	Ratio	0.64
Tumor to medulla		0.72

Note: Performance of median SWV and shear wave ratios in predicting RCC versus AML. ROC AUC was relatively low.

RCC = renal cell carcinoma; AML = angiomyolipoma; ROC = receiver operating characteristic; AUC = area under the curve; SWV = shear wave velocity.

Table 4. Performance of traditional measures and machine learning in classifying between RCC and AML: Machine learning

		Machine learning algorithm ROC AUC with associated p value from rank-sum test							
Region	Value	Logistic	p	Bayesian	p	QDA	p	SVM	p
Tumor only	Using all four statistical measures.	0.57	0.26	0.67	0.097	0.49	0.29	0.94	**4.6 E-03**
Cortex only		0.65	0.62	0.67	0.11	0.61	0.95	0.79	**2.3 E-05**
Medulla only		0.66	0.13	0.72	0.032	0.51	0.90	0.84	**1.1 E-03**
Cortex, tumor and medulla		0.71	0.099	0.78	**7.1 E-03**	0.9	**1.2 E-05**	0.98	**3.1 E-06**

Note: Performance of various machine learning techniques, including logistic regression, Bayesian classification, QDA and SVM, in predicting RCC versus AML. When all 4 statistical measures (mean, median, IQR, standard deviation) from all 3 regions (lesion, cortex and medulla) were used to predict RCC versus AML (12 features total), performance was substantially improved, with SVMs performing best. The p values were calculated using the Wilcoxon rank-sum test comparing the distribution of scores between RCC and AML. Values of $p < 0.01$ are in bold.

RCC = renal cell carcinoma; AML = angiomyolipoma; ROC = receiver operating characteristic; AUC = area under the curve; QDA = quadratic discriminant analysis; SVM = support vector machines.

statistical measures in the tumor alone, the cortex alone or the medulla alone (Fig. 4b, *first three columns*). However, the best score separation for support vector machines occurred when statistical measures were used from the tumor, cortex and medulla (AUC = 0.98; $p = 3.1 \times 10^{-6}$; Fig. 4b, *fourth column*).

Associated ROC curves are presented in Figure 5. Detailed values of model sensitivity, specificity, positive and negative predictive value, accuracy and AUC are presented in Supplementary Table 1. The significance of the difference in ROC AUC between machine learning models is presented in Tables 5–8. Support vector machines demonstrated a significantly different ROC AUC compared with median SWV when statistical features were analyzed in the tumor alone, cortex alone, medulla alone or all regions.

DISCUSSION

Support vector machines significantly outperformed median tumor SWV in distinguishing between RCC and AML. After combining features from the tumor, cortex and medulla, both support vector machines and quadratic discriminant analysis demonstrated significantly improved performance over median SWV, suggesting that SWV values outside the tumor may contain meaningful diagnostic information. Support vector machines (SVMs) performed best, and the literature confirms that non-linear SVMs perform well with high dimensional data (Ben-Hur and Weston 2010). Our methodology differs from other studies using machine learning in elastography by analyzing composite statistical features from various regions, but other studies either directly analyze color maps quantifying stiffness or combine clinical data with single elastography measurements. Our study is the first to assess machine learning for characterizing solid renal lesions using pSWE. In addition, other studies investigating renal tumors using ultrasound did not leverage the heterogeneity of various tissue regions as we did by using measurements from these regions as inputs to machine learning.

The literature shows conflicting results using strain elastography and pSWE. Strain elastography has demonstrated high sensitivity (89%–94%) and high specificity (83%–100%) for differentiating RCC from AML (Keskin et al. 2015; Onur et al. 2015; Tan et al. 2013). However, pSWE actually performed worse for this distinction than strain elastography. One study demonstrated 88% sensitivity and 54% specificity (Goya et al. 2015), but another demonstrated 48% sensitivity and 33% specificity (Lu et al. 2015). Median tumor SVW also performed poorly in our study, with ROC AUC of 0.62. The SWR of the tumor to the peripheral parenchyma was previously proposed. Although SWR was useful in the diagnosis of liver fibrosis and breast cancer (Grgurevic et al. 2015; Jia 2014), it performed poorly for solid renal lesions (Lu et al. 2015). It similarly did not perform well in our study. Other research has shown that tissue stiffness is different in the renal cortex and medulla, both with pSWE and MR elastography (Bensamoun et al. 2011; Zheng et al. 2015). By incorporating statistical features from the renal lesion, cortex and medulla with machine learning, we improved performance.

Machine learning has been used in ultrasound elastography. One study investigated chronic liver disease diagnosis by inputting information from color maps quantifying stiffness values into an SVM (Gatos et al. 2017), with ROC AUC of 0.87. Another study analyzed fibrosis in hepatitis B patients, using color maps obtained during real time elastography, and a random forest classifier performed best (Chen et al. 2017). Another study analyzing the use of SWE in breast cancer diagnosis employed deep learning on color maps quantifying stiffness, with ROC AUC of 0.95 (Zhang et al. 2016). Additional studies include diagnosing thyroid nodules *via* strain elastography using SVMs (Ma et al. 2010), using Fibroscan data and clinical/laboratory values to evaluate fibrosis in hepatitis C using SVMs (Stoean et al. 2011),

and using color maps from real-time elastography to evaluate fibrosis in hepatitis C *via* multivariate linear regression (Fujimoto et al. 2013).

Of note, SWV outside the tumor demonstrated predictive ability in this study. A malignant tumor may alter its surrounding microarchitecture, changing SWV. For instance, areas of high stiffness can be observed outside the visualized tumor margin in breast elastography (Zhang et al. 2015, 2016). Moreover, tumors may alter perfusion in surrounding tissues, changing their elasticity. A study analyzing animals both *in vivo* and *ex vivo* showed that experimental changes in renal perfusion induced by clamping the renal artery or vein altered SWV (Liu et al. 2017). Another study showed that the reduction of elasticity after diminished blood flow was the major factor influencing SWV in patients with chronic kidney disease (Asano et al. 2014). Thus, it is reasonable that incorporating SWV measurements outside the lesion can improve performance.

We acknowledge several study limitations, including the sample size of 51 patients. A future study could enroll more patients. However, we did incorporate 30 measurements of SWV for each patient from various tissue regions, which is substantially higher than earlier

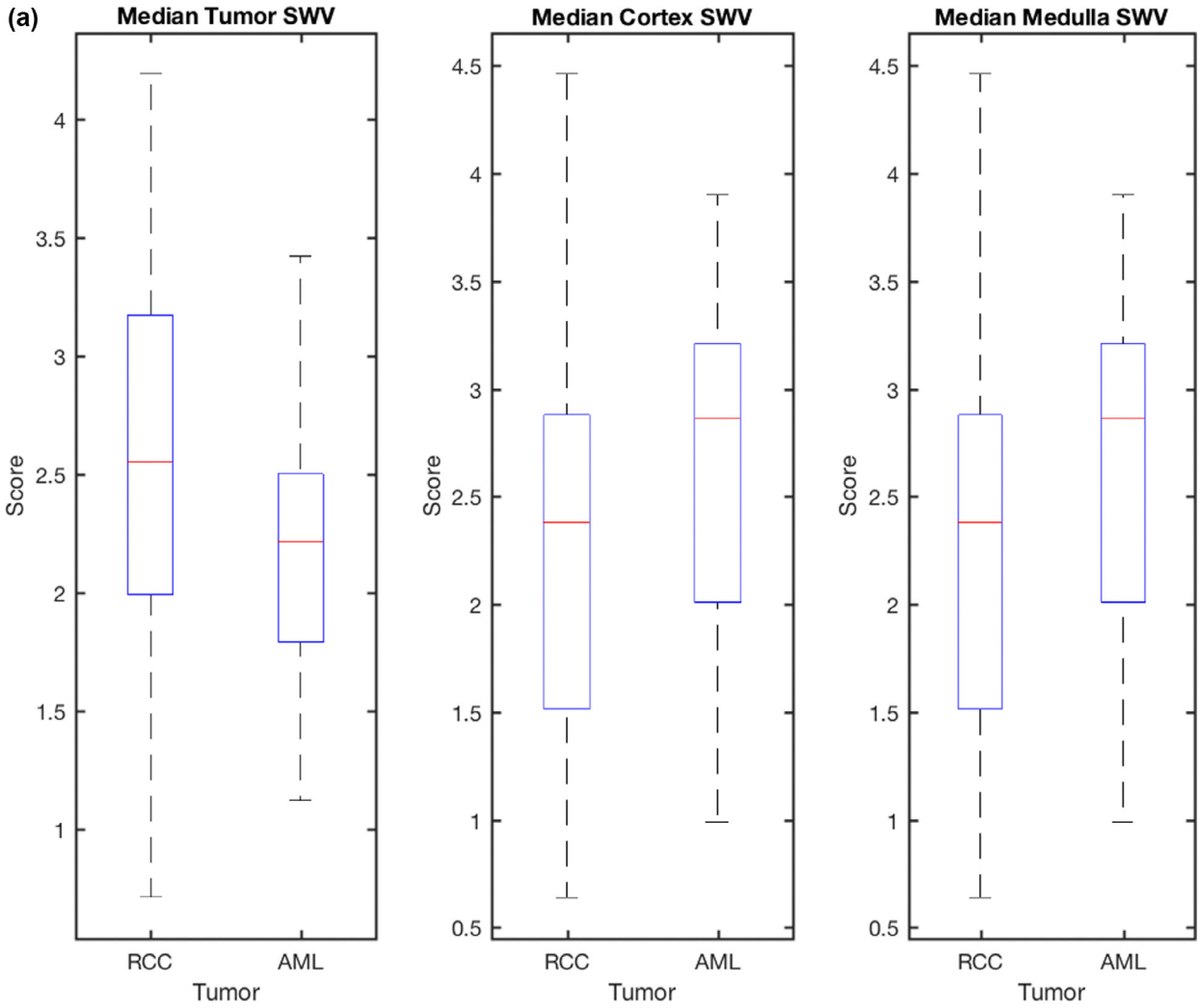

Fig. 4. Scores for AML and RCC separation using median SWV and machine learning. Scores reflect the likelihood (posterior probability) that the label came from each class. The *red line* in the boxplot indicates the median. The box edges indicate the 25th and 75th percentile, the whiskers extend to the farthest points not representing outliers, and outliers are marked with a "+" sign. (a) Boxplots show poor separation between RCC and AML when median SWV is analyzed in the tumor (*left plot*), cortex (*middle plot*) or medulla (*right plot*). (b) Improved score separation when the four statistical features in the lesion (*first plot*), cortex (*second plot*) or medulla (*third plot*) are used with a support vector machine. The best separation occurs when all four statistical features are used from the tumor, cortex and medulla (*fourth plot*). AML = angiomyolipoma; RCC = renal cell carcinoma; SWV = shear wave velocity.

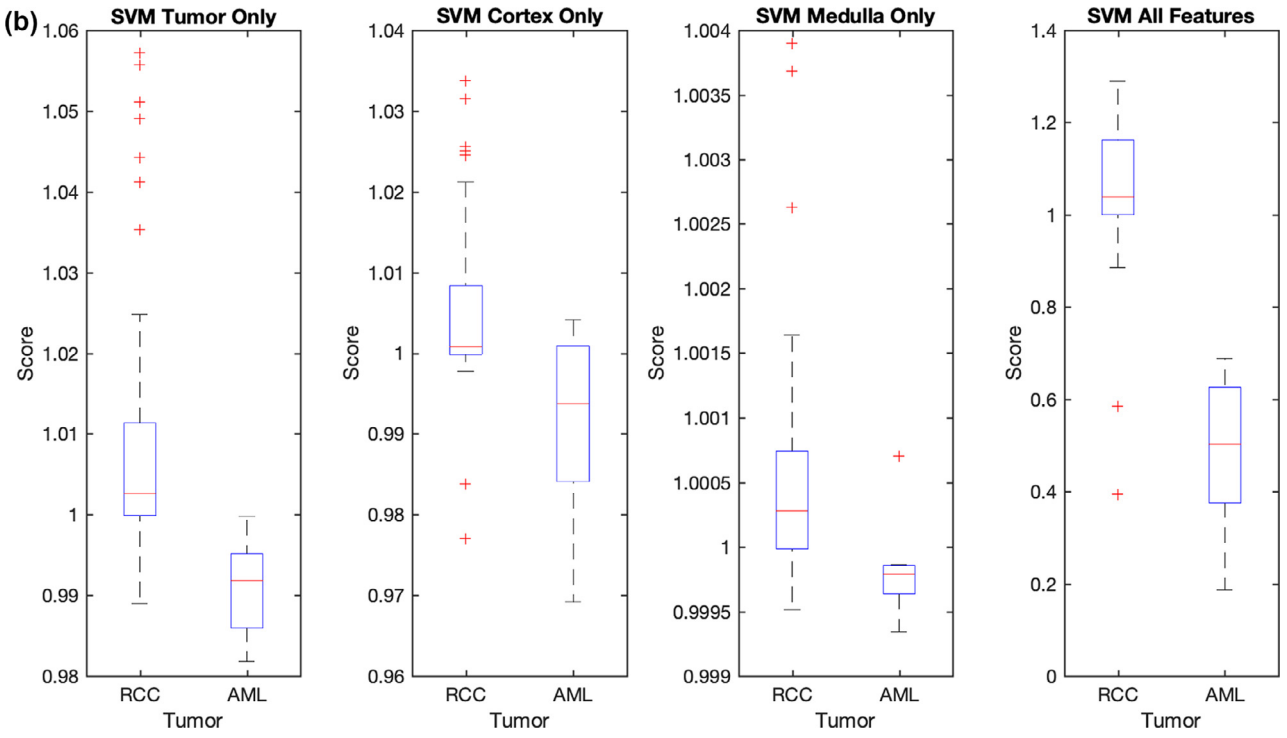

Fig. 4. Continued

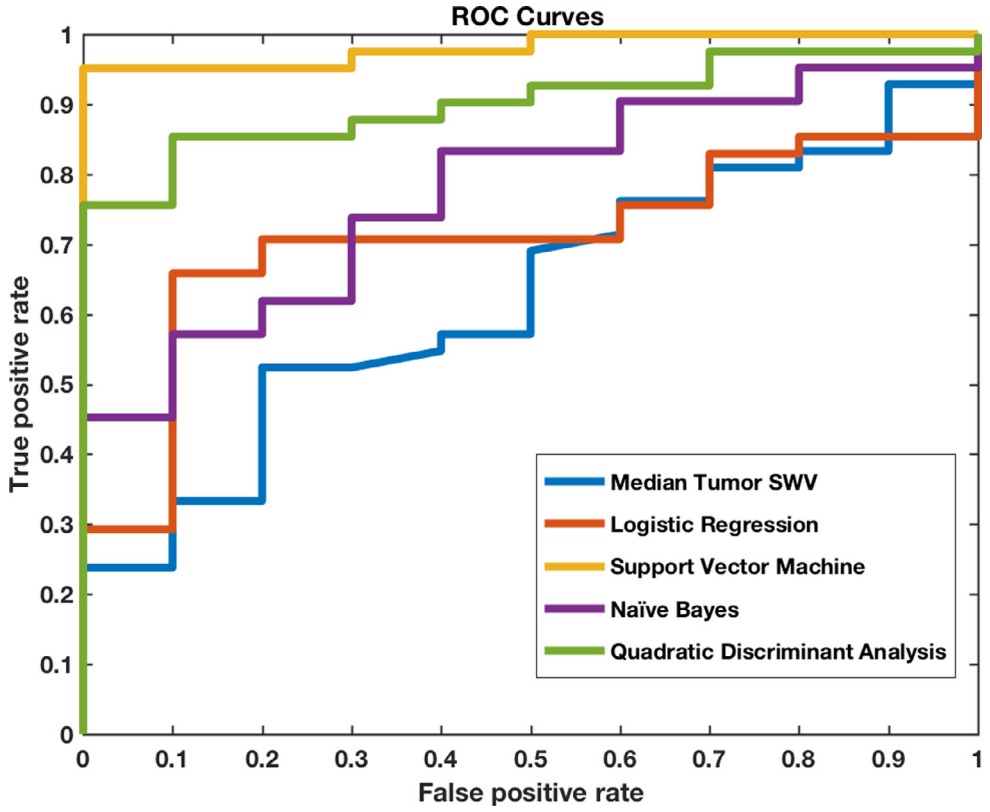

Fig. 5. ROC curves compare the performance of each machine learning algorithm and median tumor SWV to predict renal cell carcinoma versus angiomyolipoma. Support vector machines had the highest performance of all machine-learning algorithms. ROC = receiver operating characteristic; SWV = shear wave velocity.

Table 5. Comparison of performance between different models: All features in the lesion, cortex and medulla

Combination	*p* value
Median SWV versus logistic	0.490
Median SWV versus Bayesian	0.0870
Median SWV versus QDA	**8.70 e-3**
Median SWV versus SVM	**2.80 e-4**
Logistic versus Bayesian	0.279
Logistic versus QDA	0.0142
Logistic versus SVM	**1.68 e-4**
Bayesian versus QDA	0.105
Bayesian versus SVM	**8.74 e-3**
QDA versus SVM	0.0335

Note: The difference in ROC AUC between each pair of models was compared using the DeLong method. This comparison was performed using features from the lesion, cortex and medulla. Values of $p < 0.01$ are in bold.

SWV = shear wave velocity; QDA = quadratic discriminant analysis; SVM = support vector machines.

Table 6. Comparison of performance between different models: Features in the lesion only

Combination	*p* value
Median SWV versus logistic	0.537
Median SWV versus Bayesian	0.570
Median SWV versus QDA	0.275
Median SWV versus SVM	**2.50 e-4**
Logistic versus Bayesian	0.120
Logistic versus QDA	0.444
Logistic versus SVM	**4.71 e-7**
Bayesian versus QDA	0.0165
Bayesian versus SVM	**1.17 e-5**
QDA versus SVM	**8.82 e-6**

Note: The difference in ROC AUC between each pair of models was compared using the DeLong method. This comparison was performed using features from the lesion only. Values of $p < 0.01$ are in bold.

SWV = shear wave velocity; QDA = quadratic discriminant analysis; SVM = support vector machines.

Table 7. Comparison of performance between different models: Features in the cortex only

Combination	*p* value
Median SWV versus logistic	0.147
Median SWV versus Bayesian	0.0919
Median SWV versus QDA	0.125
Median SWV versus SVM	**6.37 e-3**
Logistic versus Bayesian	0.495
Logistic versus QDA	0.592
Logistic versus SVM	0.0511
Bayesian versus QDA	0.276
Bayesian versus SVM	0.133
QDA versus SVM	0.0138

Note: The difference in ROC AUC between each pair of models was compared using the DeLong method. This comparison was performed using features from the cortex only. Values of $p < 0.01$ are in bold.

SWV = shear wave velocity; QDA = quadratic discriminant analysis; SVM = support vector machines.

Table 8. Comparison of performance between different models: Features in the medulla only

Combination	*p* value
Median SWV versus logistic	0.0123
Median SWV versus Bayesian	**0.00168**
Median SWV versus QDA	0.108
Median SWV versus SVM	**2.66 e-5**
Logistic versus Bayesian	0.179
Logistic versus QDA	0.0508
Logistic versus SVM	0.0421
Bayesian versus QDA	**0.00233**
Bayesian versus SVM	0.177
QDA versus SVM	**0.00124**

Note: The difference in ROC AUC between each pair of models was compared using the DeLong method. This comparison was performed using features from the medulla only. Values of $p < 0.01$ are in bold.

SWV = shear wave velocity; QDA = quadratic discriminant analysis; SVM = support vector machines; ROC = receiver operating characteristic; AUC = area under the curve.

studies. Moreover, although this study analyzed RCC and AML, the most common benign renal neoplasm and the most common renal malignancy, respectively, a future study could analyze other benign tumors (oncocytoma), different RCC subtypes (clear cell, papillary, chromophobe), different tumors (transitional cell carcinoma) and pediatric renal tumors (Wilm's tumor). Most patients in this study had clear cell RCC, and all were 16 years of age or older. Another limitation is that examinations were performed by 1 operator, which did not allow evaluation of interobserver variability. However, an earlier report did suggest that SWE measurements demonstrate decent reproducibility within the kidney (Bob et al. 2014). Moreover, the limitations of this study are not markedly different from those of other published results. Another study using SWE to study renal allograft dysfunction used 6 measurements per kidney (vs. 30 in our study), was retrospective (vs. prospective in our study), demonstrated weaker results (AUC = 0.70 vs. AUC = 0.98) and similarly did not assess interobserver variability (Ghonge et al. 2018). Additional limitations include the small sample size of AMLs ($n = 10$) and a lack of a pathology gold standard for some AML cases. In the future, we could evaluate this technique in a multi-institutional study. We could also apply our technique to other organs, to further validate its generalizability. Finally, we could perform texture analysis on the original B-mode image to ascertain whether it provides additional diagnostic utility.

CONCLUSION

Analyzing all the statistical features from the lesion, cortex and medulla with machine learning, particularly with SVMs, is significantly better able to distinguish between RCC and AML than median SWV using pSWE. Statistical measurements outside the lesion may reflect changes in the surrounding renal parenchyma. The superior performance of SVMs likely reflects the

non-linear nature of the Gaussian radial basis function kernel. Overall, pSWE can differentiate RCC from AML with high classification accuracy when the most robust machine learning algorithm takes the maximum available information from various regions into account.

Acknowledgments—The authors would like to acknowledge the instrumental contribution to this work from the late Dr. Jürgen K. Willmann, who spearheaded this project from start to finish. He was supported by departmental funding and in part by a Siemens research grant. Hersh Sagreiya was awarded an RSNA Research Fellow grant by the Radiological Society of North America (#RF1727) and received funding from the Stanford Cancer Imaging Training Program (NIH 2 T32 CA009695-26). Dandan Li was awarded a State Scholarship from the Chinese Scholarship Council (#201606260092) to pursue her studies at Stanford University.

SUPPLEMENTARY DATA

Supplementary data related to this article can be found online at doi:10.1016/j.ultrasmedbio.2019.04.009.

APPENDIX

Summary of Machine Learning Techniques

In quadratic discriminant analysis, a transformation function is optimized to maximize the ratio of between-class variance to within-class variance and to minimize the overlap of the transformed distributions. A pseudo-quadratic transformation was used in which an inverse covariance matrix was used as a cost function (how well the machine learning algorithm maps training data to outcomes) to measure the variability of covariance matrices among the classes.

Generalized linear models consist of linear models based on the following three components: a random component, a systematic component and a link function. The random component identifies the dependent variable (Y) and its probability distribution. The systematic component identifies the set of explanatory variables (X_1,..., X_k). The link function identifies the function of the mean that is a linear function of the explanatory variables. If the outcome is binary (*i.e.,* benign versus malignant) and assuming that the random component has a binomial distribution, then the model is simply multivariate logistic regression, which is what was used here. Merging these three components leads to the following relationship between the prediction and input data (statistical features derived from 10 measurements of shear wave velocity): $g(\mu) = \alpha + \beta_1 X_1 + \ldots + \beta_k X_k$, where $g(\mu)$ is the prediction. The function could be linear or non-linear.

The naïve Bayes classifier is based on Bayes theorem, and it tends to perform well when the inputs have high dimensionality. It is based on the assumption that input data (statistical features from SWV in our case) have some multivariate distribution, but the outcomes are independent. Despite its simplicity, it is capable of performance comparable with more sophisticated classification methods. It is based on prior probabilities, derived from previous experience, which can be used to predict outcomes.

With the support vector machine, the original input (feature) space is mapped into a higher dimensional feature space in which an optimal separating hyperplane is constructed such that the distance from the hyperplane to the nearest data point is maximized. In this case, features represent the statistical quantification of intermeasurement variability and differences in SWV across 10 measurements. This aids in the generalizability of the SVM classifier. We used the Gaussian radial basis function kernel in Matlab (MathWorks, Natick, MA, USA), a non-linear support vector machine.

REFERENCES

Asano K, Ogata A, Tanaka K, Ide Y, Sankoda A, Kawakita C, Nishikawa M, Ohmori K, Kinomura M, Shimada N, Fukushima M. Acoustic radiation force impulse elastography of the kidneys: Is shear wave velocity affected by tissue fibrosis or renal blood flow?. J Ultrasound Med 2014;33:793–801.

Ben-Hur A, Weston J. A user's guide to support vector machines. Methods Mol Biol 2010;609:223–239.

Bensamoun SF, Robert L, Leclerc GE, Debernard L, Charleux F. Stiffness imaging of the kidney and adjacent abdominal tissues measured simultaneously using magnetic resonance elastography. Clin Imaging 2011;35:284–287.

Bob F, Bota S, Sporea I, Sirli R, Petrica L, Schiller A. Kidney shear wave speed values in subjects with and without renal pathology and inter-operator reproducibility of acoustic radiation force impulse elastography (ARFI)—Preliminary results. PLoS One 2014;9 e113761.

Chen Y, Luo Y, Huang W, Hu D, Zheng RQ, Cong SZ, Meng FK, Yang H, Lin HJ, Sun Y, Wang XY, Wu T, Ren J, Pei SF, Zheng Y, He Y, Hu Y, Yang N, Yan H. Machine-learning-based classification of real-time tissue elastography for hepatic fibrosis in patients with chronic hepatitis B. Comput Biol Med 2017;89:18–23.

Coy H, Hsieh K, Wu W, Nagarajan MB, Young JR, Douek ML, Brown MS, Scalzo F, Raman SS. Deep learning and radiomics: The utility of Google TensorFlow Inception in classifying clear cell renal cell carcinoma and oncocytoma on multiphasic CT. Abdom Radiol (NY) 2019;.

DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 1988;44:837–845.

Dobson AJ. An introduction to generalized linear models 1990.

Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. Radiographics 2017;37:505–515.

Feng Z, Rong P, Cao P, Zhou Q, Zhu W, Yan Z, Liu Q, Wang W. Machine learning-based quantitative texture analysis of CT images of small renal masses: Differentiation of angiomyolipoma without visible fat from renal cell carcinoma. Eur Radiol 2018;28:1625–1633.

Forman HP, Middleton WD, Melson GL, McClennan BL. Hyperechoic renal cell carcinomas: Increase in detection at US. Radiology 1993;188:431–434.

Fujimoto K, Kato M, Kudo M, Yada N, Shiina T, Ueshima K, Yamada Y, Ishida T, Azuma M, Yamasaki M, Yamamoto K, Hayashi N, Takehara T. Novel image analysis method using ultrasound elastography for noninvasive evaluation of hepatic fibrosis in patients with chronic hepatitis C. Oncology 2013;84(suppl 1):3–12.

Garra BS. Elastography: History, principles, and technique comparison. Abdom Imaging 2015;40:680–697.

Gatos I, Tsantis S, Spiliopoulos S, Karnabatidis D, Theotokas I, Zoumpoulis P, Loupas T, Hazle JD, Kagadis GC. A machine-learning

algorithm toward color analysis for chronic liver disease classification, employing ultrasound shear wave elastography. Ultrasound Med Biol 2017;43:1797–1810.

Ghonge NP, Mohan M, Kashyap V, Jasuja S. Renal allograft dysfunction: Evaluation with shear-wave sonoelastography. Radiology 2018;288:146–152.

Goya C, Daggulli M, Hamidi C, Yavuz A, Hattapoglu S, Cetincakmak MG, Teke M. The role of quantitative measurement by acoustic radiation force impulse imaging in differentiating benign renal lesions from malignant renal tumours. Radiol Med 2015;120:296–303.

Grgurevic I, Puljiz Z, Brnic D, Bokun T, Heinzl R, Lukic A, Luksic B, Kujundzic M, Brkljacic B. Liver and spleen stiffness and their ratio assessed by real-time two dimensional-shear wave elastography in patients with liver fibrosis and cirrhosis due to chronic viral hepatitis. Eur Radiol 2015;25:3214–3221.

Guo Y, Hastie T, Tibshirani R. Regularized linear discriminant analysis and its application in microarrays. Biostatistics 2007;8:86–100.

Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: Data mining, inference, and prediction. New York, NY: Springer; 2009.

Hodgdon T, McInnes MD, Schieda N, Flood TA, Lamb L, Thornhill RE. Can Quantitative CT Texture analysis be used to differentiate fat-poor renal angiomyolipoma from renal cell carcinoma on unenhanced CT images? Radiology 2015;276:787–796.

Jia C, Alam S, Azar R, Garra B. Estimation of shear modulus ratio between inclusion and background using strain ratios in 2-D ultrasound elastography. IEEE Trans Ultrason Ferroelect Freq Control 2014;61:611–619.

Jinzaki M, Ohkuma K, Tanimoto A, Mukai M, Hiramatsu K, Murai M, Hata J. Small solid renal lesions: Usefulness of power Doppler US. Radiology 1998;209:543–550.

Jinzaki M, Silverman SG, Akita H, Nagashima Y, Mikami S, Oya M. Renal angiomyolipoma: A radiological classification and update on recent developments in diagnosis and management. Abdom Imaging 2014;39:588–604.

Keskin S, Guven S, Keskin Z, Ozbiner H, Kerimoglu U, Yesildag A. Strain elastography in the characterization of renal cell carcinoma and angiomyolipoma. Can Urol Assoc J 2015;9:e67–e71.

Kocak B, Yardimci AH, Bektas CT, Turkcanoglu MH, Erdim C, Yucetas U, Koca SB, Kilickesmez O. Textural differences between renal cell carcinoma subtypes: Machine learning-based quantitative computed tomography texture analysis with independent external validation. Eur J Radiol 2018;107:149–157.

Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. AJR Am J Roentgenol 2017;208 754−560.

Lakhani P, Prater AB, Hutson RK, Andriole KP, Dreyer KJ, Morey J, Prevedello LM, Clark TJ, Geis JR, Itri JN, Hawkins CM. Machine learning in radiology: Applications beyond image interpretation. J Am Coll Radiol 2018;15:350–359.

Liu X, Li N, Xu T, Sun F, Li R, Gao Q, Chen L, Wen C. Effect of renal perfusion and structural heterogeneity on shear wave elastography of the kidney: An in vivo and ex vivo study. BMC Nephrol 2017;18:265.

Lu Q, Wen JX, Huang BJ, Xue LY, Wang WP. Virtual Touch quantification using acoustic radiation force impulse (ARFI) technology for the evaluation of focal solid renal lesions: Preliminary findings. Clin Radiol 2015;70:1376–1381.

Ma JLS, Dighe M, Lim D, Kim Y. Differential diagnosis of thyroid nodules with ultrasound elastography based on support vector machines. IEEE Int Ultrasonics Symp Proc 2010;1372–1375.

Nowicki A, Dobruch-Sobczak K. Introduction to ultrasound elastography. J Ultrason 2016;16:113–124.

Onur MR, Poyraz AK, Bozgeyik Z, Onur AR, Orhan I. Utility of semiquantitative strain elastography for differentiation between benign and malignant solid renal masses. J Ultrasound Med 2015;34:639–647.

Park BK. Renal angiomyolipoma: Radiologic classification and imaging features according to the amount of fat. AJR Am J Roentgenol 2017;209:826–835.

Qayyum T, Oades G, Horgan P, Aitchison M, Edwards J. The epidemiology and risk factors for renal cancer. Curr Urol 2013;6:169–174.

Raja BK, Madheswaran M, Thyagarajah K. A hybrid fuzzy-neural system for computer-aided diagnosis of ultrasound kidney images using prominent features. J Med Syst 2008;32:65–83.

Raja KB, Madheswaran M, Thyagarajah K. Texture pattern analysis of kidney tissues for disorder identification and classification using dominant Gabor wavelet. Machine Vision and Applications 2010;21:287–300.

Schölkopf B, Smola AJ. Learning with kernels: Support vector machines, regularization, optimization, and beyond. Cambridge, MA: MIT Press; 2002.

Siegel CL, Middleton WD, Teefey SA, McClennan BL. Angiomyolipoma and renal cell carcinoma: US differentiation. Radiology 1996;198:789–793.

Sigrist RMS, El Kaffas A, Jeffrey RB, Rosenberg J, Willmann JK. Intra-individual comparison between 2-D shear wave elastography (GE System) and Virtual Touch Tissue Quantification (Siemens System) in grading liver fibrosis. Ultrasound Med Biol 2017;43:2774–2782.

Sigrist RMS, Liau J, Kaffas AE, Chammas MC, Willmann JK. Ultrasound elastography: Review of techniques and clinical applications. Theranostics 2017;7:1303–1329.

Stoean R, Stoean C, Lupsor M, Stefanescu H, Badea R. Evolutionary-driven support vector machines for determining the degree of liver fibrosis in chronic hepatitis C. Artif Intell Med 2011;51:53–65.

Subramanya MB, Kumar V, Mukherjee S, Saini M. SVM-based CAC system for B-mode kidney ultrasound images. J Digit Imaging 2015;28:448–458.

Tan S, Ozcan MF, Tezcan F, Balci S, Karaoglanoglu M, Huddam B, Arslan H. Real-time elastography for distinguishing angiomyolipoma from renal cell carcinoma: Preliminary observations. AJR Am J Roentgenol 2013;200:W369–W375.

Zhang Q, Xiao Y, Chen S, Wang C, Zheng H. Quantification of elastic heterogeneity using contourlet-based texture analysis in shear-wave elastography for breast tumor classification. Ultrasound Med Biol 2015;41:588–600.

Zhang Q, Xiao Y, Dai W, Suo J, Wang C, Shi J, Zheng H. Deep learning based classification of breast tumors with shear-wave elastography. Ultrasonics 2016;72:150–157.

Zheng XZ, Yang B, Fu NH. Preliminary study on the kidney elasticity quantification in patients with chronic kidney disease using virtual touch tissue quantification. Iran J Radiol 2015;12:e12026.

Zhou L, Zhang Z, Chen YC, Zhao ZY, Yin XD, Jiang HB. A deep learning-based radiomics model for differentiating benign and malignant renal tumors. Transl Oncol 2019;12:292–300.