

تشخیص سرطان ریه با استفاده از یادگیری ماشین و تصویر برداری CT Scan

شهره کسایی

علیرضا عیسوند

kasaei@sharif.edu

alireza.isavand@sharif.edu

چکیده - امروزه سرطان ریه به عنوان یکی از بیماری‌های کشنده شناخته می‌شود. با این وجود، در صورتی که این سرطان در مراحل ابتدایی شناسایی شود و مورد درمان قرار بگیرد، احتمال بهبود بیمار افزایش می‌یابد. در حال حاضر یکی از بهترین روش‌های موجود برای تصویربرداری از ریه، تصویربرداری CT Scan است. پزشکان با کمک این تصاویر تصمیم می‌گیرند که آیا فرد مورد نظر دارای سرطان ریه است یا خیر. با این وجود، ممکن است پزشکان نظر متفاوتی در این زمینه داشته باشند و تشخیص برای آن‌ها دشوار باشد. برای برطرف کردن این مشکل می‌توان از روش‌های یادگیری ماشین در تشخیص سرطان ریه استفاده کرد اما دقت این روش‌ها هنوز به اندازه‌ای نیست که بتوان آن‌ها را جایگزین نظر پزشک کرد. در این مقاله به بررسی تعدادی از این روش‌ها و مزایا و معایب هر کدام و مشکلاتی که با آن‌ها روبرو شده‌اند پرداخته شده است. در نهایت نیز روشی ارائه شده که عملکرد بهترین مدل موجود را بهبود داده است. با این روش، صحت (accuracy) مدل از ۸۸.۴٪ به ۹۲٪ افزایش یافته و میزان حساسیت (sensitivity) آن ثابت مانده و specificity از ۴۰٪ به ۵۰٪ افزایش یافته است. مدل پیشنهادی خوش‌خیم یا بدخیم بودن سرطان را نیز پیش‌بینی می‌کند.

کلید واژه - یادگیری ماشین، سرطان ریه، دسته‌بندی، CT Scan

1- مقدمه

مشکوک به سرطان ریه تشخیص داده می‌شدند.

برای برطرف کردن مشکلات ذکر شده می‌توانیم از روش‌های تشخیص کامپیوتری کمک بگیریم [3]. این روش‌ها شامل روش‌های مربوط به پردازش تصویر و یادگیری ماشین هستند که به طور کلی می‌توانند دو فایده داشته باشند: اولاً ثبات و هماهنگی بین تصمیمات پزشک‌ها ایجاد کند و دوماً به پزشکان کم‌تجربه‌تر و کم‌مهارت‌تر کمک کند تا تصمیمات صحیحی بگیرند.

در این مقاله به بررسی برخی از روش‌های تشخیص سرطان ریه به کمک یادگیری ماشین و پردازش تصویر پرداخته شده و مزایا، معایب و مشکلاتی که با آن‌ها رو به رو شده‌اند، ذکر شده است. در نهایت نیز سعی شده با ایجاد تغییراتی در مرحله پیش‌پردازش و استخراج ویژگی (feature extraction) در بهترین روش موجود، عملکرد آن بهبود داده شود. این روش صحت مدل قبلی را از ۸۸.۴٪ به ۹۲٪ و specificity را از ۴۰٪ به ۵۰٪ رسانده است.

سرطان ریه یکی از سرطان‌های کشنده است اما اگر در مراحل اولیه شناسایی شود، می‌تواند در بسیاری از موارد مورد درمان قرار گیرد و بهبود یابد. با این حال، تشخیص آن در مراحل اولیه سخت است و معمولاً زمانی قادر به شناسایی آن هستند که سرطان به مراحل انتهایی رسیده باشد. در حال حاضر بهترین راه تصویربرداری از ریه، استفاده از CT Scan است که می‌تواند انواع ندول‌های (nodule) مشکوک و غیرمشکوک را مشخص کند [1]. با این وجود، ممکن است بین پزشکان و رادیولوژیست‌های مختلف، در مورد تشخیص سلول‌های سرطانی اختلاف نظر وجود داشته باشد.

در کارآزمایی ملی غربالگری ریه در ایالات متحده آمریکا (NLST)، مرگ و میر در اثر سرطان ریه ۲۰ درصد کاهش یافت [2]. این موضوع ایالات متحده را قانع کرد تا افراد مشکوک به سرطان ریه را تحت نظارت منظم قرار دهند. یکی از مشکلاتی که در این راه وجود داشت، تعداد زیاد مواردی بود که به اشتباه

2- کارهای مرتبط

تلاش‌های مختلفی در زمینه‌ی تشخیص سرطان ریه به کمک پردازش تصویر و یادگیری ماشین انجام شده است. یکی از این راه‌ها، وارد کردن اطلاعات بیمار (از قبیل سن، جنسیت و مصرف سیگار) و همچنین اطلاعات ندول‌های یافت‌شده (مانند طول ندول) به صورت دستی است. سپس به کمک روش‌های یادگیری ماشین و با استفاده از ویژگی‌های داده شده، این ندول‌ها به دسته‌های خوش‌خیم و بدخیم دسته‌بندی می‌شود. یکی از این روش‌ها که توسط انجمن قفسه سینه بریتانیا (BTS) [4] توصیه شد، استفاده از مدل Brock بود [5].

در این دسته از روش‌ها، وارد کردن اطلاعات به صورت دستی است و می‌تواند منجر به بروز خطا شود. برای مثال در یکی از این مدل‌ها، اگر طول ندول ۱ میلی‌متر بزرگ‌تر گزارش شود، شانس ابتلا تا دو برابر افزایش می‌یابد. این درحالی است که دقت افرادی که اطلاعات را وارد می‌کنند کمتر از این مقدار است [6].

کارهایی برای برطرف کردن این مشکلات انجام شده است. از جمله اینکه ciompi [7] به صورت خودکار ندول‌ها را به یکی از دسته‌های جامد (solid)، غیرجامد (non-solid)، نیمه‌جامد (part-solid)، calcified، پیرامونی (perifissural) و speculated دسته‌بندی می‌کند. مشکل این کار در این است که دسته‌های مختلف ندول‌ها تعریف دقیقی ندارند و خود پزشکان به طور شهودی این مفاهیم را ایجاد کرده‌اند.

یکی از مفاهیم مرتبط با موضوع مورد بحث، radiomic است. radiomic به استخراج ویژگی به صورت خودکار از عکس‌های پزشکی گفته می‌شود [8][9]. ریشه اکثر این دسته از روش‌ها به کارهایی که در دهه ۱۹۷۰ میلادی در زمینه دسته‌بندی تصاویر بافت‌دار (textured) [10] انجام شد برمی‌گردد. در روش radiomic ابتدا مدل تعداد زیادی ویژگی را از طریق داده‌های آموزش استخراج می‌کند. برای این کار بخش مورد نظر (مثل تومور) به شکلی مشخص شده است

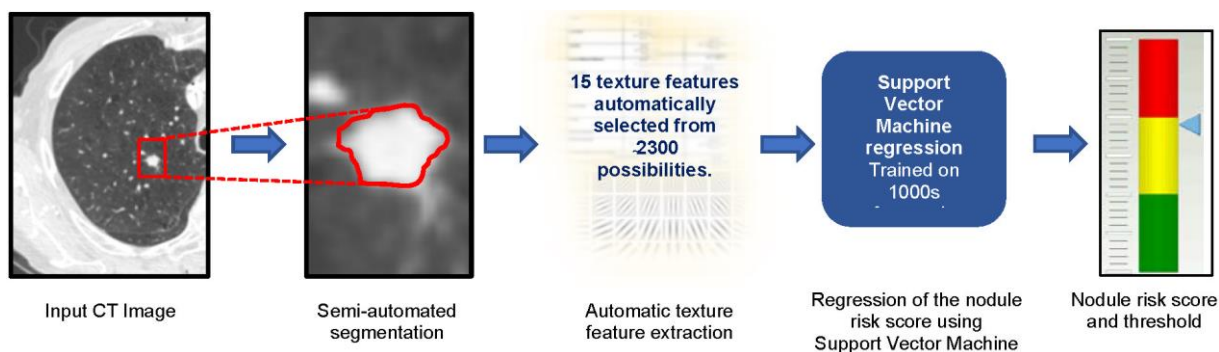
که بتوان ویژگی‌های را به طور خودکار به دست آورد. سپس از بین این ویژگی‌ها، تعدادی به عنوان ویژگی‌های اصلی انتخاب می‌شود به طوری که مدل بهترین عملکرد را داشته باشد.

اگر تعداد این ویژگی‌ها خیلی کم باشد، می‌توان با در نظر گرفتن یک آستانه (threshold) دسته‌بندی را انجام داد و اگر تعداد آن‌ها به اندازه کافی باشد، از روش‌های پیچیده‌تری مثل Support Vector Machines (SVM) یا Random Forests می‌توان استفاده کرد. برخی از رویکردهایی که از این روش در دسته‌بندی ندول‌های ریه استفاده کرده‌اند در مقاله Wilson et al [11] بررسی شده‌اند.

یکی از مشکلات روش‌های ابتدایی radiomics، بیش‌برازش (overfitting) است که به دلیل مستقل نبودن داده‌های آزمون (test data) و داده‌های اعتبارسنجی (validation data) رخ می‌دهد [12].

مسابقه‌ای با موضوع دسته‌بندی ندول‌های ریه به دسته‌های خوش‌خیم و بدخیم برگزار شد که چالش دسته‌بندی ندول‌های ریه SPIE_AAPM_NCI LungX نام داشت [13]. شکل ۱ به ترتیب مراحل روش برنده این مسابقه را نشان می‌دهد که از چهار بخش تشکیل شده است: ۱) قطعه‌بندی ندول (nodule segmentation)، ۲) استخراج ویژگی بافت (texture feature extraction)، ۳) رگرسیون ریسک (risk score regression) و ۴) آستانه‌بندی ریسک برای انتخاب دسته (risk score thresholding).

روش ما در این مسابقه، ۱۵ ویژگی را از بین حدود ۱۳۰۰ ویژگی معرفی‌شده مانند Gabor [14] و Haralick [10] و ویژگی‌های رایجی مثل میانگین و مرکز انتخاب می‌کرد. یکی از مهم‌ترین ایده‌ها تقسیم ویژگی‌ها به دو دسته بود که دسته اول مربوط به خود ندول و دسته دوم مربوط به ناحیه اطراف آن بود. این ۱۵ ویژگی را به کمک جستجوی جامع (exhaustive search) و به صورت حریصانه به شکلی استخراج کردیم که مدل بهترین عملکرد را داشته باشد. در نهایت با یک دسته‌بند SVM نمونه‌ها را دسته‌بندی



شکل 1 نمودار بلوکی مدل برنده در LungX

۸۴.۶٪، میزان حساسیت مدل ۸۲.۵٪ و specificity آن ۸۶.۷٪ است. مزیت این روش، استفاده از فیلتر دایره‌ای در مرحله تشخیص ناحیه مطلوب (region of interest) است که هزینه‌ی مرحله‌ی آموزش را کم می‌کند.

Sangamitraa و Govindaraju [19] از روش k-مرکزی (k-means) برای خوشه‌بندی پیکسل‌ها و قطعه‌بندی استفاده می‌کنند. همچنین دسته‌بندی با شبکه انتشار پس‌رو (back propagation network) انجام می‌شود و ویژگی‌های مختلف از قبیل آنتروپی و همبستگی به کمک ماتریس همزمانی سطح خاکستری (gray-level co-occurrence matrix) استخراج می‌شوند. این مدل از فیلتر میانه (median filter) برای کاهش نویز استفاده می‌کند. صحت این مدل ۹۰.۷٪ است.

Roy، Sirohi و Patle [20] روشی ارائه کردند که ابتدا با کمک تبدیل خاکستری (gray transformation) تباين (contrast) عکس را افزایش می‌دهد. سپس عکس را دودویی می‌کند و به کمک مدل هم‌دور فعال (active contour model) قطعه‌بندی می‌کند. حال ویژگی‌هایی مانند مساحت، میانگین، طول و عرض برای دسته‌بندی استخراج شده و دسته‌بندی به کمک روش تداخل فازی (fuzzy interference) انجام می‌شود. ضعف این مدل مشخص نکردن خوش‌خیم یا بدخیم بودن ندول است. صحت این روش ۹۴.۱۲٪ است.

Gonzalez و Ponomaryvo [22] روشی ارائه کردند که

کردیم. برای آموزش مدل از مجموعه دادگان LIDC_IDRI استفاده کردیم و از مجموعه دادگان PLAN نیز در مرحله‌ی استخراج ویژگی کمک گرفتیم.

امروزه با رشد مجموعه دادگان در دسترس، روش‌های شبکه عصبی پیچشی (convolutional neural network) عملکرد بسیار بهتری نسبت به روش‌های پیشین داشته‌اند. از مزایای این رویکرد می‌توان به استخراج خودکار ویژگی‌ها از پایه به جای انتخاب دستی آن‌ها، چندلایه بودن و توانایی یادگیری توابع پیچیده، بهینه‌سازی یکپارچه و همزمان به جای بهینه‌سازی جداگانه و نیاز نداشتن به قطعه‌بندی اشاره کرد. در مسابقات تشخیص سرطان ریه علوم داده Kaggle نیز برندگان از شبکه عصبی پیچشی استفاده کردند [15][16].

Aggarwal، Furquan و Kalra [17] روشی ارائه کردند که ویژگی‌های هندسی و آماری ندول را استخراج می‌کند و سپس دسته‌بندی را با روش تحلیل جداکننده خطی (Linear Discriminant Analysis) و قطعه‌بندی را با روش آستانه‌بندی بهینه انجام می‌دهد. صحت این مدل ۸۴٪، میزان حساسیت آن ۹۷.۱۴٪ و specificity آن ۵۳.۳۳٪ است که با میزان مورد انتظار فاصله دارد. این مدل از روش‌های یادگیری ماشین استفاده نمی‌کند و روش‌های پردازش تصویر آن بسیار ساده هستند.

Jin و Zhang [18] از یک شبکه عصبی پیچشی برای دسته‌بندی استفاده کردند که صحت مدل

در آن برای یافتن ناحیه مطلوب از اطلاعات پیشینی (priori) و HouseFieldUnit استفاده می‌شود. سپس ویژگی‌های مربوط به شکل ندول و ویژگی‌های بافتی برای استفاده در دسته‌بند SVM استخراج می‌شوند. مزیت این روش پیش‌بینی خوش‌خیم و بدخیم بودن ندول است و مشکل آن نیاز به اطلاعات پیشین است.

Joseph و Ignatious [21] از روشی استفاده کردند که ابتدا به کمک فیلتر Godel کیفیت تصویر را بهبود می‌دهد. سپس از قطعه‌بندی آب‌گیر (watershed segmentation) برای قطعه‌بندی استفاده می‌شود. مزیت این روش، جلوگیری از قطعه‌بندی بیش از حد (over-segmentation) است. صحت این مدل ۹۰.۱٪ است که از سایر مدل‌های موجود عملکرد بهتری دارد و از آن در روش پیشنهادی استفاده شده است. یکی از مشکلات این روش، حذف نکردن نویزهای عکس ورودی در مرحله پیش‌پردازش است که می‌تواند در تصمیم‌گیری مدل مشکل ایجاد کند. مشکل بعدی، استفاده از ویژگی‌های بسیار کم در مدل است و دسته‌بندی به خوش‌خیم و بدخیم نیز انجام نمی‌شود.

3- روش پیشنهادی

بهترین روش حال حاضر، روش Joseph و Ignatious [21] است. در روش پیشنهادی از این مدل استفاده شده است و سعی شده مشکلات این مدل برطرف شود و عملکرد آن بهبود یابد.

3-1- پردازش تصویر

در مرحله پردازش تصویر، به جای استفاده از فیلتر Godel، از فیلترهای میانه و گاوسی استفاده شده است. ابتدا نویزهای نمک و فلفل در تصویر را با استفاده از فیلتر میانه برطرف می‌کنیم [23]. سپس برای هموارسازی تصویر و از بین بردن نویزهای لکه‌ای (speckle) از فیلتر گاوسی استفاده می‌کنیم.

3-2- قطعه‌بندی

این مرحله به منظور جداکردن بخش‌های مختلف

تصویر است که ناحیه مطلوب در تصویر از سایر اجزا جدا شود. در تشخیص سرطان ریه، ناحیه مطلوب همان ندول‌ها هستند. در روش پیشنهادی، از روش قطعه‌بندی آب‌گیر استفاده شده است. مزیت این روش در تشخیص اشیایی است که با هم در تماس هستند.

3-3- استخراج ویژگی

در این مرحله ویژگی‌های لازم استخراج می‌شوند تا در مرحله دسته‌بندی از آن‌ها استفاده شود. محیط، مساحت، مرکز و قطر از جمله این ویژگی‌ها هستند.

3-4- دسته‌بندی

برای دسته‌بندی از SVM که یک روش دسته‌بندی نظارت‌شده (Supervised) است، استفاده شده است. در این روش، تابع دسته‌بند ما به صورت [22]

$$D(x) = W^T x_i + b \quad (1)$$

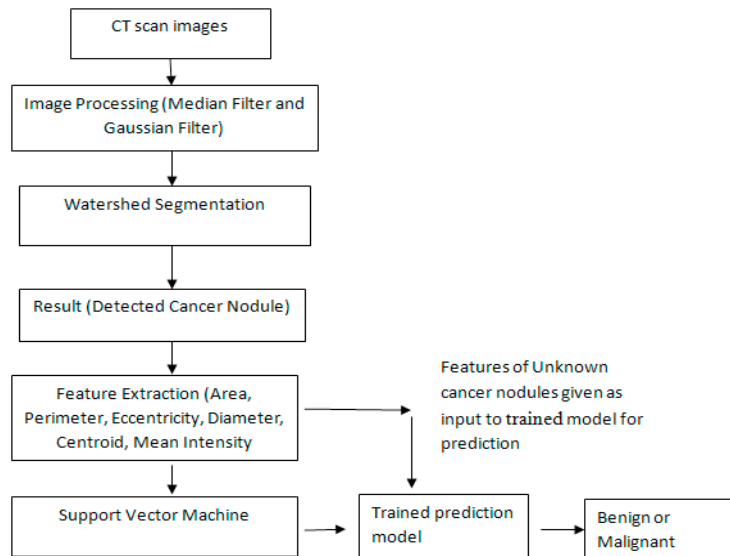
است که x همان ویژگی‌های انتخاب‌شده ورودی و W^T بردار وزن‌های به دست آمده برای دسته‌بند خطی است و b جانب‌داری (bias) مدل است و داریم:

$$D(x) = W^T x_i + b \geq 1 \text{ (for } y_i = 1) \quad (2)$$

$$D(x) = W^T x_i + b \leq -1 \text{ (for } y_i = -1) \quad (3)$$

مراحل مختلف مدل در شکل ۲ نشان داده شده است. مزایای روش پیشنهادی عبارتند از: (۱) صحت مدل نسبت به بهترین مدل موجود بهبود پیدا کرده است. (۲) سرطان ریه را به دسته‌ی خوش‌خیم و بدخیم دسته‌بندی می‌کند. (۳) نویزهای نمک و فلفل و نویزهای لکه‌ای را حذف می‌کند تا بر روی عملکرد مدل اثر منفی نگذارند.

مشکلات روش پیشنهادی عبارتند از: (۱) با وجود افزایش صحت مدل، همچنان فاصله‌ی معناداری تا صحت ۱۰۰٪ داریم. (۲) با وجود دسته‌بندی سرطان به خوش‌خیم و بدخیم، مرحله‌ی آن را مشخص نمی‌کند.



شکل 2 مدل پیشنهادی

4- نتایج تجربی

مدل ارائه شده در موارد مختلف نشان داده شده است. صحت مدل از ۸۸.۴٪ به ۹۲٪ افزایش پیدا کرده، میزان حساسیت آن ثابت مانده و specificity آن از ۴۰٪ به ۵۰٪ رسیده است. آموزش دسته‌بند ۵.۹۳ ثانیه زمان می‌برد. نمودار پراکندگی مدل آموزش دیده در شکل‌های ۹، ۱۰ و ۱۱ نشان داده شده است. در این نمودارها، نقاط آبی برای نمایش سرطان خوش‌خیم درست پیش‌بینی شده و نقاط قرمز برای نمایش سرطان بدخیم درست پیش‌بینی شده هستند و ضربدرها نیز برای نمایش پیش‌بینی غلط هستند.

نتیجه‌ی دسته‌بندی ۵ تصویر اعتبارسنجی و ندول‌های مختلف آن‌ها در جدول ۲ نمایش داده شده است.

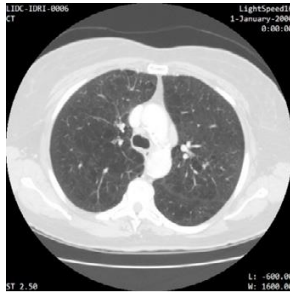
یکی دیگر از مواردی که بررسی کردیم، تاثیر اندازه ندول بر عملکرد مدل بود. برای این منظور دو مجموعه داده با ۶۴۰ تصویر با استفاده از دادگان NLST ساختیم که در یکی اندازه ندول عنصر تمایزدهنده بود اما در دیگری خیر. در مجموعه اول به ازای هر تصویر بدخیم، یک تصویر خوش‌خیم با همان اندازه در مجموعه دادگان قرار دادیم اما در دسته دوم تصاویر خوش‌خیم را به صورت تصادفی انتخاب کردیم. هر دو به کمک یک دسته‌بند SVM دسته‌بندی

تصاویر مورد استفاده از کنسرسیوم پایگاه داده تصاویر ریه (LIDC) گرفته شده‌اند. ابعاد عکس‌ها ۵۱۲*۵۱۲ است و فرمت آن‌ها DICOM است. استفاده از این فرمت سخت است و با برنامه‌ی MicroDicom آن‌ها را به فرمت JPEG تبدیل می‌کنیم. مدل ما در MATLAB R2016a پیاده‌سازی شده و برای پیاده‌سازی بخش یادگیری ماشین، از جعبه ابزار یادگیری ماشین متلب استفاده شده است. از اعتبارسنجی تقاطعی پنج‌تایی (5 folds cross validation) استفاده شده تا بیش‌برازش اتفاق نیفتد. ۱۶ تصویر برای آموزش و ۵ تصویر دارای ۱۵ ندول برای اعتبارسنجی مورد استفاده قرار گرفته‌اند.

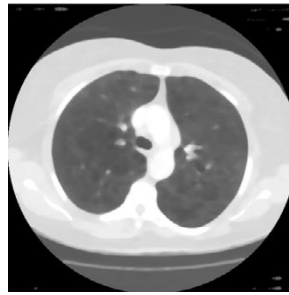
یکی از مشکلاتی که با آن رو به رو شدیم حجم بسیار زیاد پایگاه داده LIDC (۱۲۴ گیگابایت) بود. مشکل دیگر نیز فرمت علامت‌گذاری سرطان در این عکس‌ها بود که با فرمت xml انجام شده و کار با آن سخت بود.

در شکل‌های ۳، ۴، ۵، ۶، ۷ و ۸ تصویر به دست آمده در هر مرحله از مدل نمایش داده شده است.

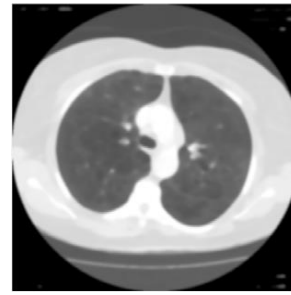
در جدول ۱ نتیجه‌ی مقایسه‌ی بهترین مدل موجود و



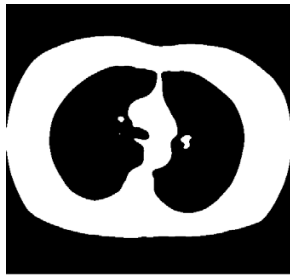
شکل 3 تصویر خاکستری اولیه



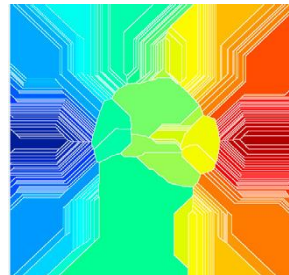
شکل 4 تصویر فیلتر میانه



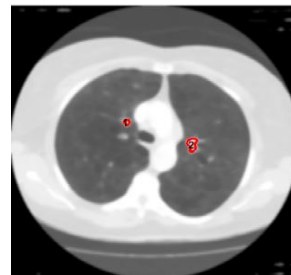
شکل 5 تصویر فیلتر گاوسی



شکل 6 تصویر دودویی شد








شکل 7 تصویر قطعه‌بندی آب‌گیر



شکل 8 علامت‌گذاری سرطان

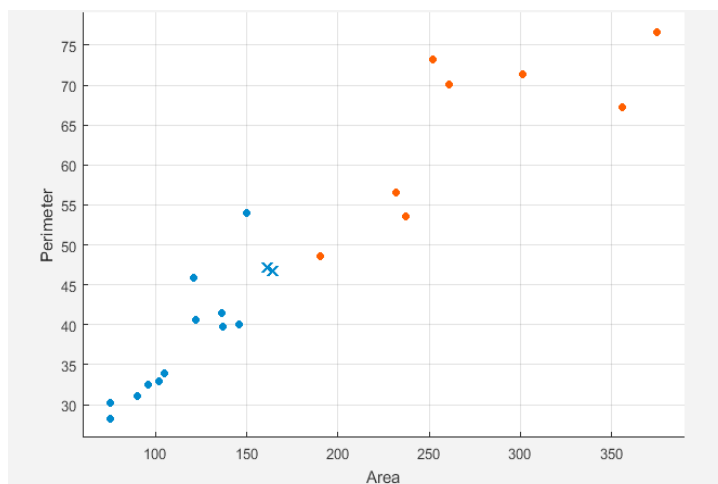
جدول 2 دسته‌بندی گره‌ها با مدل پیشنهادی

تصویر	ندول	دسته‌بندی	علامت
	Nodule1	Malignant	True
	Nodule2	Malignant	True
	Nodule3	Malignant	True
	Nodule1	Benign	True
	Nodule2	Benign	True
	Nodule3	Benign	True
	Nodule1	Benign	True
	Nodule2	Malignant	True
	Nodule3	Malignant	True
	Nodule1	Malignant	False
	Nodule2	Malignant	False
	Nodule3	Benign	True
	Nodule4	Benign	True
	Nodule1	Malignant	True
	Nodule2	Malignant	True

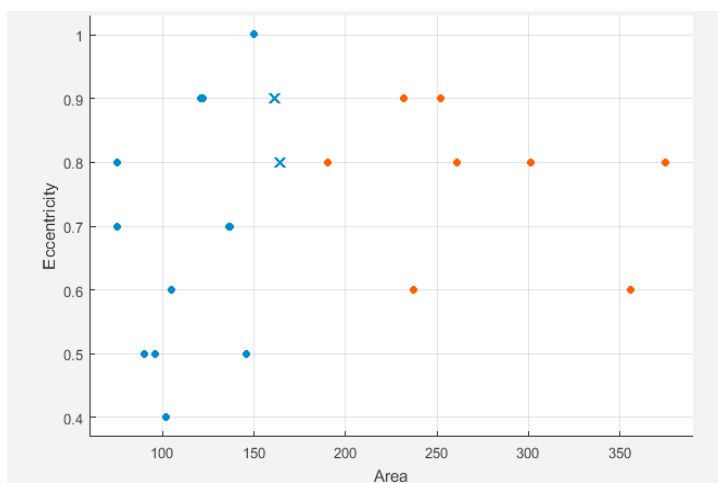
شدند. معیار AUC (Area Under ROC Curve) برای دسته اول ۰.۷۰ بود در حالی که اگر تنها از اندازه ندول استفاده می‌کردیم AUC برابر ۰.۵۰ بود. این معیار برای مجموعه دوم ۰.۹۱ بود و در صورتی که تنها از اندازه ندول استفاده می‌کردیم تقریباً برابر ۰.۷ بود.

جدول 1 مقایسه مدل پیشنهادی و بهترین مدل کنونی

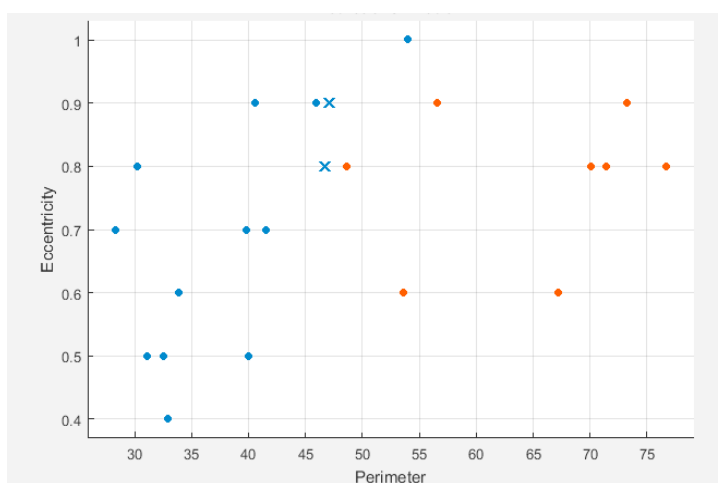
	مدل پیشنهادی	بهترین مدل کنونی
تعداد گره پیداشده	23	24
تعداد مثبت صحیح	21	21
تعداد مثبت اشتباه	2	2
تعداد منفی صحیح	2	3
تعداد منفی اشتباه	0	0
صحت	92%	88.4%
حساسیت	100%	100%
Specificity	50%	40%



شکل 9 نمودار پراکندگی برای مساحت در برابر محیط در مدل آموزش دیده



شکل 10 نمودار پراکندگی برای مساحت در برابر دوری از مرکز



شکل 11 نمودار پراکندگی برای محیط در برابر دوری از مرکز

5- نتیجه‌گیری
بهترین مدل موجود نیز خوش‌خیم و بدخیم بودن را
مدل‌های موجود فاصله‌ی زیادی با صحت ۱۰۰٪ دارند و مشخص نمی‌کند. در روش پیشنهادی، برای

قطعه‌بندی از روش قطعه‌بندی آب‌گیر و برای دسته‌بندی از SVM استفاده کردیم. صحت مدل در تشخیص سرطان به ۹۲٪ و در تشخیص نوع آن به ۸۶.۶٪ رسید. با این حال هنوز فاصله‌ی زیادی تا صحت مورد انتظار داریم. مدل ما همچنان مرحله‌ی سرطان را مشخص نمی‌کند که می‌تواند در آینده مورد بررسی قرار بگیرد. با این حال همچنان در استفاده از این روش‌ها، سوالات بسیاری وجود دارد از قبیل اینکه دخالت پزشک و اعتماد بر مدل باید چه اندازه باشد.

مراجع

- [11] Wilson, R., Devaraj, A. (2017) "Radiomics of pulmonary nodules and lung cancer". *Transl Lung Cancer Res* 2017;6:86-91.
- [12] Chalkidou, A., O'Doherty, M.J., Marsden, P.K. (2015) "False Discovery Rates in PET and CT Studies with Texture Features: A Systematic Review". *PLoS One* 2015;10:e0124165.
- [13] Armato, S.G. 3rd, Drukker, K., Li, F., et al. (2016) "LUNGx Challenge for computerized lung nodule classification". *J Med Imaging (Bellingham)* 2016;3:044506.
- [14] Lee, T.S. (1996) "Image Representation Using 2D Gabor Wavelets". *IEEE Trans Pattern Anal Mach Intell* 1996;18:1-13.
- [15] Li, Y., Chen, K.Z., Sui, X.Z., et al. (2011) "Establishment of a mathematical prediction model to evaluate the probability of malignancy or benign in patients with solitary pulmonary nodules". *Beijing Da Xue Xue Bao Yi Xue Ban* 2011;43:450-4.
- [16] Hammack, D. (2017) "Forecasting Lung Cancer Diagnoses with Deep Learning". Available online: https://raw.githubusercontent.com/dhammack/DSB2017/master/dsb_2017_daniel_hammack.pdf
- [17] Aggarwal, T., Furqan, A., & Kalra, K. (2015) "Feature extraction and LDA based classification of lung nodules in chest CT scan images." 2015 International Conference On Advances In Computing, Communications And Informatics (ICACCI), DOI: 10.1109/ICACCI.2015.7275773.
- [18] Jin, X., Zhang, Y., & Jin, Q. (2016) "Pulmonary Nodule Detection Based on CT Images Using Convolution Neural Network." 2016 9Th International Symposium On Computational Intelligence And Design (ISCID). DOI: 10.1109/ISCID.2016.1053.
- [19] Sangamithraa, P., & Govindaraju, S. (2016) "Lung tumour detection and classification using EK-Mean clustering." 2016 International Conference On Wireless Communications, Signal Processing And Networking (Wispnet). DOI: 10.1109/WISPNET.2016.7566533.
- [20] Roy, T., Sirohi, N., & Patle, A. (2015) "Classification of lung image and nodule detection using fuzzy inference system." International Conference On Computing, Communication & Automation. DOI: 10.1109/CCAA.2015.7148560.
- [21] Ignatious, S., & Joseph, R. (2015) "Computer aided lung cancer detection system." 2015 Global Conference On Communication Technologies (GCCT), DOI: 10.1109/GCCT.2015.7342723.
- [22] Rendon-Gonzalez, E., & Ponomaryov, V. (2016) "Automatic Lung nodule segmentation and classification in CT images based on SVM." 2016 9Th International Kharkiv Symposium On Physics And Engineering Of Microwaves, Millimeter And Submillimeter Waves (MSMW). DOI: 10.1109/MSMW.2016.7537995.
- [23] Miah, M.B.A., & Yousuf, M.A. (2015) "Detection of lung cancer from CT image using image processing and neural network." 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT): 1-6.
- [1] Gindi, A.M., Al Attiatalla, T.A., & Sami, M.M. (2014) "A Comparative Study for Comparing Two Feature Extraction Methods and Two Classifiers in Classification of Earlystage Lung Cancer Diagnosis of chest x-ray images". *Journal of American Science*, **10**(6): 13-22.
- [2] National Lung Screening Trial Research Team, Aberle, D. R., Adams, A. M., et al. (2011) "Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening". *N Engl J Med* 2011. 365: 395-409.
- [3] Xiuhua, G., Tao, S., & Zhigang, L. (2011) "Prediction Models for Malignant Pulmonary Nodules Based-on Texture Features of CT Image." In *Theory and Applications of CT Imaging and Analysis*. DOI: 10.5772/14766.
- [4] Callister, M.E., Baldwin, D.R., Akram, A.R., et al. "British Thoracic Society guidelines for the investigation and management of pulmonary nodules". *Thorax* 2015;70 Suppl 2:ii1-54.
- [5] McWilliams, A., Tammemagi, M.C., Mayo, J.R., et al. (2013) "Probability of Cancer in Pulmonary Nodules Detected on First Screening CT". *N Engl J Med* 2013;369:910-9.
- [6] Revel, M.P., Bissery, A., Bienvenu, M., et al. (2004) "Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable?". *Radiology* 2004;231:453-8.
- [7] Ciompi, F., Chung, K., van Riel, S.J., et al. (2017) "Towards automatic pulmonary nodule management in lung cancer screening with deep learning". *Sci Rep* 2017;7:46479.
- [8] Aerts, H.J., Velazquez, E.R., Leijenaar, R.T., et al. (2014) "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach". *Nat Commun* 2014;5:4006.
- [9] Lambin, P., Rios-Velazquez, E., Leijenaar, R., et al. (2012) "Extracting more information from medical images using advanced feature analysis". *Eur J Cancer* 2012;48:441-6.
- [10] Haralick, R.M., Shanmugam, K., Dinstein, I. (1973) "Textural Features for Image Classification". *IEEE Trans Syst Man Cybern Syst* 1973;3:610-21.