



## پردازش تصاویر و استخراج ندول‌های ریوی با ترکیب ویژگی‌های بافتی و هندسی

هانیه قهوجی خلیق: دانشکده مهندسی برق، دانشگاه صنعتی ارومیه، ارومیه، ایران

یعقوب پوراسد: دانشکده مهندسی برق، دانشگاه صنعتی ارومیه، ارومیه، ایران (\* نویسنده مسئول) y.pourasad@uut.ac.ir

سرلی مقدس قولیان: دانشکده مهندسی برق، دانشگاه صنعتی سهند، شهر جدید سهند، تبریز، ایران

### چکیده

#### کلیدواژه‌ها

پردازش تصاویر،

سرطان ریه،

ندول،

غده،

تصاویر سی تی اسکن،

طبقه‌بندی

**زمینه و هدف:** از آنجایی که تشخیص غدد سرطانی و بدخیم ریه با استفاده از روش‌های عکس‌برداری نظیر CT-Scan و بدون نیاز به نمونه‌برداری باعث کاهش ریسک پخش شدن ندول سرطانی می‌شود، بنابراین توسعه یک سیستم تشخیصی کامپیوتری جهت پردازش تصاویر و غدد ریوی و سپس طبقه‌بندی آن‌ها به دو دسته خوش خیم و بدخیم، در تشخیص زودهنگام سرطان ریه و نجات جان بیماران نقش بسزایی ایفا می‌کند. هدف از این پژوهش، دستیابی به دقت طبقه‌بندی بالاتر و در نتیجه دقت تشخیص بالاتر غده‌های بدخیم و خوش خیم می‌باشد.

**روش کار:** در این پژوهش الگوریتم‌هایی که پیش از این برای طبقه‌بندی غدد ریوی استفاده شده، معرفی می‌شود و در نهایت الگوریتم پیشنهادی ارائه می‌شود. در الگوریتم پیشنهادی ابتدا تصاویر سی تی اسکن ریه پیش پردازش شده و سپس به وسیله کاتور فعال چن-وسه، ناحیه ندول استخراج می‌شود. از ناحیه قطعه‌بندی شده، ویژگی‌های هیستوگرام، بافت و هندسی استخراج می‌شود. سپس این ویژگی‌ها با استفاده از دو طبقه بند SVM و KNN، ندول‌های ریوی را به دو دسته خوش خیم و بدخیم طبقه بندی می‌کنند.

**یافته‌ها:** نتایج حاصل از پیش‌پردازش‌های اعمالی بررسی می‌گردد. سپس تصاویر پیش‌پردازش شده توسط الگوریتم چن-وسه قطعه‌بندی شده و ناحیه استخراج شده و تحت الگوریتم‌های استخراج ویژگی قرار گرفته و ۲۵ ویژگی مختلف بافتی و هندسی برای هر غده از این نواحی استخراج می‌گردد. در مرحله آخر، توسط داده‌های استخراج شده، طبقه‌بندهای SVM و KNN اقدام به طبقه‌بندی غدد می‌کنند. معیارهای دقت، حساسیت و میزان اختصاصی بودن در طبقه بند برتر  $90.8\%$ ،  $100\%$  و  $89\%$  بدست می‌آید.

**نتیجه‌گیری:** این روش علاوه بر دقت بالا در تشخیص، روشی کم هزینه و کم خطر نیز می‌باشد. روش پیشنهادی بخاطر دارا بودن حساسیت بسیار بالا و همچنین دارا بودن مقادیر مطلوب دو معیار دقت و میزان اختصاصی بودن و تعداد پایین ویژگی‌های مورد استفاده جهت طبقه‌بندی، بعنوان یک روش کارآمد و مناسب جهت طبقه‌بندی غدد ریوی پیشنهاد می‌گردد.

**تعارض منافع:** گزارش نشده است.

**منبع حمایت کننده:** حامی مالی نداشته است.

**شیوه استناد به این مقاله:**

Ghahvechi Khaligh H, Pourasad Y, Moghadas Gholian S. Image processing and extraction of pulmonary nodules by combining tissue and geometric features. Razi J Med Sci. 2020;27(4):106-121.

\* انتشار این مقاله به صورت دسترسی آزاد مطابق با [CC BY-NC-SA 3.0](https://creativecommons.org/licenses/by-nc-sa/3.0/) صورت گرفته است.

## Image processing and extraction of pulmonary nodules by combining tissue and geometric features

Haniyeh Ghahvechi Khaligh, Department of Electrical Engineering, Urmia University of Technology, Urmia, Iran

Yaghoob Pourasad, Faculty of Electrical Engineering, Urmia University of Technology, Urmia, Iran (\* Corresponding author) y.pourasad@uut.ac.ir

Serly Moghadas Gholian, Department of Electrical Engineering, Sahand University of Technology, Tabriz, Iran

### Abstract

**Background:** Background: Since the diagnosis of cancerous and malignant lung glands using imaging techniques such as CT-Scan without the need for sampling reduces the risk of spreading cancerous nodules, the development of a computerized diagnostic system for processing images and pulmonary glands and then class Their classification into two categories, benign and malignant, plays an important role in the early diagnosis of lung cancer and the survival of patients. Access to a database with a uniform statistical population of malignant and benign glands is one of the most fundamental steps in the implementation and evaluation of computerized diagnostic systems for cancer patients. In the present study, the image database consortium image collection of lung images has been used. This database includes images of CT scans of lung cancer, along with the diagnostic opinion of a specialist doctor and the identified areas of the glands. This database has been compiled by the National Cancer Organization of America, the National Health Organization and the Food and Drug Administration made available to the public. In this database, the CT images of each patient are stored in DICOM format, which is the standard for storing and suitable for processing medical images. The average incision of each scan for each patient is 254 incisions, the distance between each incision is 9 to 9 mm. The aim of this study is to achieve higher classification accuracy and therefore higher diagnostic accuracy of malignant and benign glands.

**Methods:** In this study, the algorithms that have been used to classify the pulmonary glands are introduced and finally the proposed algorithm is presented. In the proposed algorithm, the CT scan images of the lungs are pre-processed and then extracted from the nodule area by the active Chen-Wess contour. From the fragmented area, the histogram, texture and geometric features are extracted. These features then classify the pulmonary nodules into two categories, benign and malignant, using two classes, SVM and KNN. After extracting the features from the fragmented areas and normalizing them, with a large amount of data (feature) we are faced with using this data to make the final decision and classification about whether the glands are benign or malignant due to the large number. It is necessary to choose the best and most valuable features for the correct classification. There are different ways to select the feature, but due to the time consuming nature of this process, in the present study, this step is eliminated and first all the features are classified. Then, by trying and making a mistake, the best features are selected for each class. Therefore, by doing this, the feature and classification are selected at the same time and the computational load and processing time are reduced. In this research, the extracted features is given to the two well-known classifiers of the support vector machine and the nearest neighbor parameter. This database has three Excel files, the first / adjacent file contains 6 information such as the number of nodules in each patient along with the size of each nodule, the main source of the nodule and the

### Keywords

Image processing,  
Lung cancer,  
Nodules,  
Glands,  
CT scans,  
Classification

Received: 18/04/2020

Published: 05/07/2020

final diagnosis of the radiologist and the treatment for each patient. The second file contains general information such as the date of each scan, the name of the company that made the CT device, the device model, the software version, and the image ID. In the third Excel file, the number of nodules larger than 9 mm and smaller than 9 mm for each patient is given.

**Results:** Each scan was examined separately by four radiologists, and scans identified by all four radiologists were added to the database. Experts have divided each nodule into one of four unknown categories: benign, benign, primary malignant, and metastatic malignancy. In this study, for each patient, the incision in which the nodule appears is selected. Also, since the nodules are classified into 4 categories, in this study, we have classified the unknown and benign category as benign and the primary malignant and malignant metastatic categories as malignant. In this study, lung images of 65 cancer patients from the mentioned database were classified, of which 49 patients had malignant nodules and 25 patients had benign nodules. In the proposed algorithm, a semi-automatic method is used to segment the pulmonary nodule area. Using an automatic classification algorithm that does not require the selection of two border points of the gland, it can increase the rate of fragmentation. Of course, the advantage of the semi-automatic segmentation method is its high classification accuracy, which is much lower in automatic segmentation. Therefore, proposing and implementing an automated segmentation algorithm that is both highly accurate is at the top of the project's future plans. In this study, the lung glands were classified into two categories: benign and malignant. The results of the proposed actions are examined. The pre-processed images are then aligned by the Chen-and-Three algorithm and the area is extracted and subjected to feature extraction algorithms, and 25 different tissue and geometric features are extracted for each gland from these areas. Finally, by extracted data, the SVM and KNN classifications classify the glands. Criteria for accuracy, sensitivity and specificity in the top class are obtained by 90.8%, 100% and 89%.

**Conclusion:** In addition to high accuracy in diagnosis, this method is also a low cost and low risk method. By comparing the results of the proposed method with the previous methods, the proposed method received the most sensitivity, and in many studies, the highest classification accuracy. Given that the criterion of sensitivity means the ability of classification in the correct diagnosis of the disease in a person and the criterion of specificity means the ability of classification in the correct diagnosis of the absence of the disease in the person, so it can be concluded that the criterion Sensitivity is very important in research related to glandular diagnosis in medical imaging. This is because correctly diagnosing the presence of a disease or cancer is much more important and vital than diagnosing its absence in a candidate. Therefore, the proposed method is recommended as an efficient and suitable method for classification of the pulmonary glands due to its very high sensitivity and also having the desired values of two criteria of accuracy and specificity and low number of features used for classification.

**Conflicts of interest:** None

**Funding:** None

#### Cite this article as:

Ghahvechi Khaligh H, Pourasad Y, Moghadas Gholian S. Image processing and extraction of pulmonary nodules by combining tissue and geometric features. Razi J Med Sci. 2020;27(4):106-121.

**\*This work is published under CC BY-NC-SA 3.0 licence.**

## مقدمه

اخیراً، استفاده از توموگرافی کامپیوتری با دوز کم (CT) برای معاینه سرطان ریه بسیار مورد استفاده قرار گرفته است (۱-۳). بطور کلی در سیستم‌های CAD، جهت طبقه‌بندی ندول‌های ریوی سه مرحله پیش‌پردازش، قطعه‌بندی، استخراج ویژگی و طبقه‌بندی وجود دارد. هر یک از این مراحل را می‌توان با الگوریتم‌های مختلف پیاده‌سازی کرد که در این بخش به مرور مقالات مربوطه و روش‌های بکار گرفته شده و میزان دقت تشخیص بررسی و مقایسه می‌گردد. شکل ۱ بلوک دیاگرام مراحل طبقه‌بندی ندول‌های ریوی را نشان می‌دهد.

گو و همکارانش (۴)، روش جدیدی به نام قطعه‌بندی با یک کلیک (Single click ensemble segmentation) که براساس الگوریتم کلیک و رشد (Click-SCES) پایه‌ریزی شده، ارائه کردند. آنها در این روش اقدام به قطعه‌بندی ناحیه ندول ریه از سایر بافت‌ها تنها با یک کلیک و ایجاد بذر اولیه کردند. این روش در مقایسه با روش‌های قطعه‌بندی پیشین که به چندین بذر نیاز داشت، بسیار کارآمدتر، دقیق‌تر و خودکار است. در مطالعه‌ای (۵)، اقدام به طبقه‌بندی مراحل سرطان ریه بر روی تصاویر CT اسکن کردند. در این روش، نواحی‌ای از تصویر که به هم تماس دارند با یک رنگ نمایش داده می‌شود. پیکسل‌هایی که بیشترین اندازه گرادیان را دارند بعنوان مرزهای این نواحی انتخاب می‌شوند.

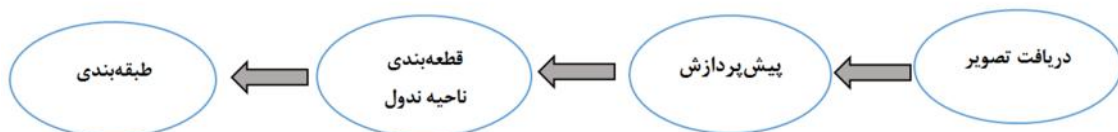
در پژوهشی دیگر (۶) روش نوینی را ارائه دادند که بجای انجام مراحل قطعه‌بندی و استخراج ویژگی، تصاویر خام CT اسکن را مدل‌سازی کردند و یک معماری یادگیری ماشین انتها-به-انتهای نام شبکه عصبی کانوال شده چند برشی (Multi-crop convolutional neural network-MC-CNN) تعبیه کردند. در مطالعه‌ای دیگر (۷)، پس از پیش‌پردازش

تصویر با استفاده از فیلتر گوسی و میانه و حذف نویز و سایر اغتشاشات، از روش جغرافیای رنگ جهت ناحیه بندی جهت ناحیه بندی و تشخیص ندول استفاده کردند. در (۸) روشی جدید بر مبنای شبکه عصبی کانولوشنی ارائه دادند که علاوه بر طبقه‌بندی ۴ نوع ندول ریه، قادر به طبقه‌بندی ندول‌های از نوع کدورت شیشه‌ای زمینی (GGO-Ground glass opacity) و حتی بافت‌های غیرندولی می‌باشد. نتایج این روش، عملکرد خوب در طبقه‌بندی ندول‌های ریوی، بخصوص در ندول‌های پیچیده GGO و نواحی غیرندولی را نشان می‌دهد. در تحقیقی دیگر (۹) با استفاده از روش‌های بر مبنای توپولوژی ندول، اقدام به طبقه‌بندی ندول‌های خوش خیم و بدخیم نمودند. آنها برای استخراج ویژگی از بافت ندول‌ها از روش درخت تکاملی (Phylogenetic tree) استفاده کردند و با اینکار توانستند اطلاعات رفتاری از بافت هر نوع ندول را استخراج کنند.

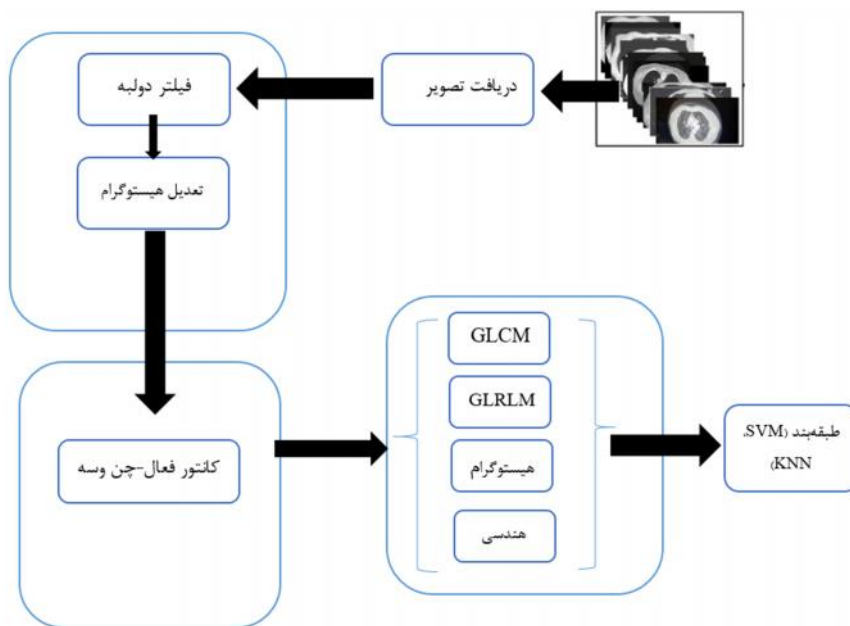
## روش کار

در این بخش، ابتدا روش‌های پیش‌پردازش تصاویر CT اسکن استفاده شده، در مرحله دوم روش‌های قطعه‌بندی ناحیه ندول از تصاویر CT، سپس روش‌های استخراج ویژگی از نواحی قطعه‌بندی شده، و در مرحله ۴ طبقه‌بندی‌های استفاده شده، ارائه می‌گردد. در نهایت اطلاعاتی درباره دیتاست استفاده شده ارائه می‌شود. شکل ۲ بلوک دیاگرام روش پیشنهادی را نشان می‌دهد.

طبقه‌بندی K-نزدیکترین همسایه: طبقه‌بندی KNN یک روش یادگیری تحت نظارت غیرپارامتری است که جهت طبقه‌بندی و رگرسیون (Regression) استفاده می‌شود (۱۰). در حالت طبقه‌بندی، از معیار شباهت (معیار فاصله یا ...) جهت طبقه‌بندی داده‌ها استفاده می‌شود. در این روش، برای مشخص شدن برچسب (کلاس) هر داده کاندید، برچسب اکثریت K داده



شکل ۱- بلوک دیاگرام مراحل طبقه‌بندی ندول‌های ریوی



شکل ۲- بلوک دیاگرام روش پیشنهادی

کنسرسیونم پایگاه داده تصاویر ریه (۹) (LIDC-IDRI Lung Image Database Consortium image) استفاده شده است. این پایگاه داده شامل تصاویر CT اسکن سرطان ریه به همراه نظر تشخیصی پزشک متخصص و ناحیه مشخص شده غدد می‌باشد. این پایگاه داده توسط سازمان ملی سرطان آمریکا (NCI)، سازمان ملی سلامت (FNIH) و سازمان غذا و دارو (FDA) گردآوری و در دسترس عموم قرار گرفته است. در این پایگاه داده، تصاویر CT هر بیمار با فرمت DICOM که استاندارد ذخیره سازی و مناسب جهت پردازش تصاویر پزشکی است، ذخیره شده است. میانگین برش‌های هر اسکن برای هر بیمار ۲۵۴ برش می‌باشد که فاصله هر بین برش‌های مجاور بین ۰/۶ الی ۳ میلیمتر می‌باشد (۱۲-۱۵). این پایگاه داده دارای سه فایل اکسل می‌باشد که در فایل اول اطلاعاتی همچون تعداد ندول‌های هر بیمار به همراه اندازه هر ندول، منشا اصلی پخش شده ندول و تشخیص نهایی رادیولوژیست و اقدام درمانی انجام شده برای هر بیمار آورده شده است. در فایل دوم اطلاعات عمومی همچون تاریخ انجام هر اسکن، نام شرکت سازنده دستگاه CT به همراه مدل دستگاه، ورژن نرم‌افزار و شناسه تصاویر می‌باشد. در فایل اکسل سوم، تعداد ندول‌های بزرگتر از ۳ میلیمتر و کوچکتر از ۳ میلیمتر برای هر بیمار آورده

نزدیک‌ترین همسایه‌اش به آن داده کاندید تعلق می‌گیرد. برای تعیین نزدیک‌ترین داده‌ها به داده کاندید، معیارهای فاصله زیادی مانند فاصله اقلیدسی، منهن، مینکووسکی، همینگ، مربعی و ... تعریف شده است. فاصله اقلیدسی مرسوم‌ترین متریک فاصله مورد استفاده در طبقه‌بند KNN می‌باشد. فاصله اقلیدسی دو بردار  $x$  و  $y$  مطابق رابطه (۱) تعریف می‌شود (۱۱):

$$\begin{aligned} d(x, y) &= d(y, x) \\ &= \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \dots + (y_n - x_n)^2} \\ &= \sqrt{\sum_{i=1}^n (y_i - x_i)^2} \end{aligned} \quad (1)$$

در رابطه بالا،  $n$  تعداد ابعاد بردار می‌باشد. همانطور که مشاهده می‌شود، با انتخاب  $K=3$  نمونه تست به کلاس دوم (مثلث قرمز) تعلق می‌گیرد. همچنین با انتخاب  $K=5$  نمونه تست به کلاس اول (مربع آبی) تعلق می‌گیرد.

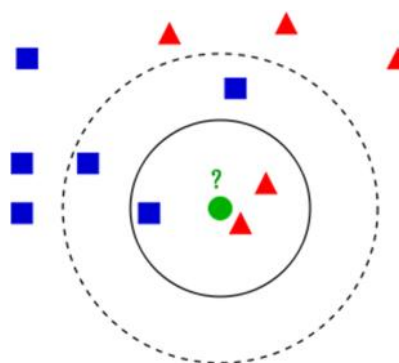
معرفی دیتاست سرطان ریه: دسترسی به پایگاه داده با جامعه آماری یکنواخت از غدد بدخیم و خوش‌خیم یکی از اساسی‌ترین مراحل پیاده‌سازی و ارزیابی سیستم‌های تشخیصی کامپیوتری برای بیماری‌های سرطانی می‌باشد. در پژوهش حاضر، از مجموعه تصاویر



### یافته‌ها

در این بخش، نتایج حاصل از پیاده‌سازی روش پیشنهادی بر پایگاه داده LIDC-IDRI بررسی و تشریح می‌گردد. در مرحله اول، نتایج حاصل از پیش‌پردازش‌های اعمالی بررسی می‌گردد. در ادامه و در بخش قطعه‌بندی، تصاویر پیش‌پردازش شده توسط الگوریتم چن-وسه قطعه‌بندی شده و ناحیه استخراج شده در بخش سوم تحت الگوریتم‌های استخراج ویژگی قرار گرفته و ۲۵ ویژگی مختلف بافتی و هندسی برای هر غده از این نواحی استخراج می‌گردد. در مرحله آخر، توسط داده‌های استخراج شده، طبقه‌بندی‌های SVM و KNN اقدام به طبقه‌بندی غدد می‌کنند. نتایج حاصل از طبقه‌بندی به صورت دقت و حساسیت برای هر طبقه‌بندی و هر کرنلی ارائه می‌شوند.

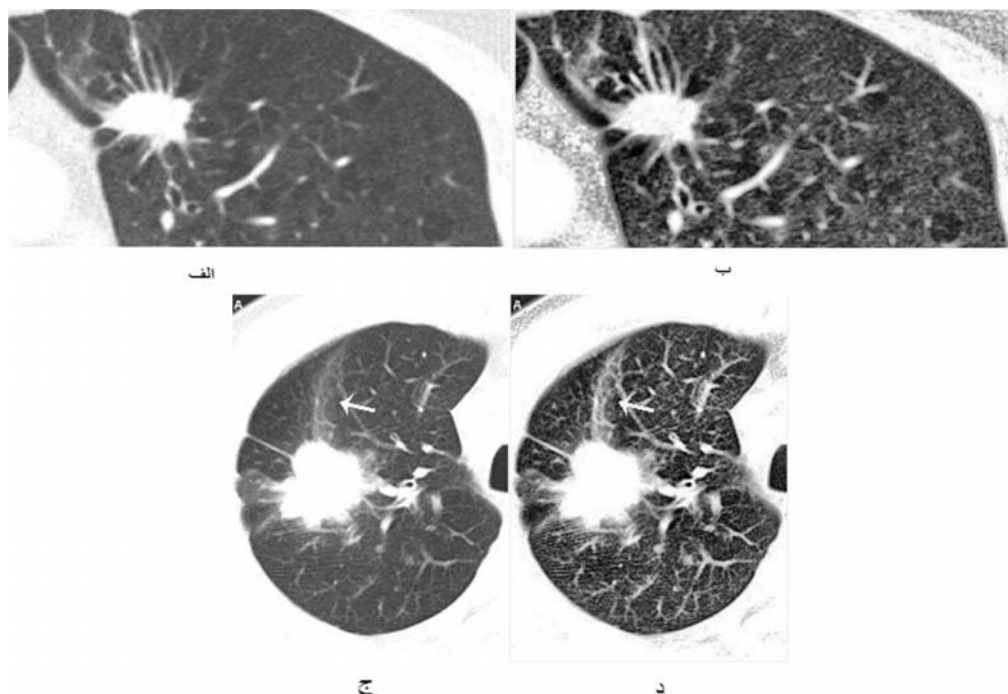
**پیش‌پردازش:** جهت کاهش نویز و افزایش گیرایی تصویر، باید تصاویر در مرحله اول پیش‌پردازش شوند. در این مرحله، ابتدا فیلتر دولبه جهت کاهش نویزها و علامت‌ها و فلش‌هایی که توسط رادیولوژیست ممکن است ثبت شده باشند، بر روی تصاویر اعمال می‌شود. در مرحله بعد، الگوریتم تعدیل هیستوگرام تطبیقی بر روی



شکل ۳- نمونه اعمال طبقه‌بندی KNN جهت طبقه‌بندی داده تست (دایره سبز)

شده است.

در پژوهش حاضر، روش پیشنهادی توسط نرم‌افزار Matlab 2019a پیاده‌سازی شده است. ویژگی‌های استخراجی از نواحی ندول‌ها به طور خودکار در یک فایل اکسل ذخیره شده و سپس در مرحله بعد، فایلی به فرمت .mat. از این داده‌ها ساخته شده که توسط نرم‌افزار متلب قابلیت خواندن دارد و سپس عملیات طبقه‌بندی انجام گرفته است.



شکل ۴- نمونه تصاویر CT scan قبل و بعد از اعمال الگوریتم پیش‌پردازش پیشنهادی ( تصاویر الف و ج قبل از پیش‌پردازش و تصاویر ب و د پس از اعمال پیش‌پردازش)

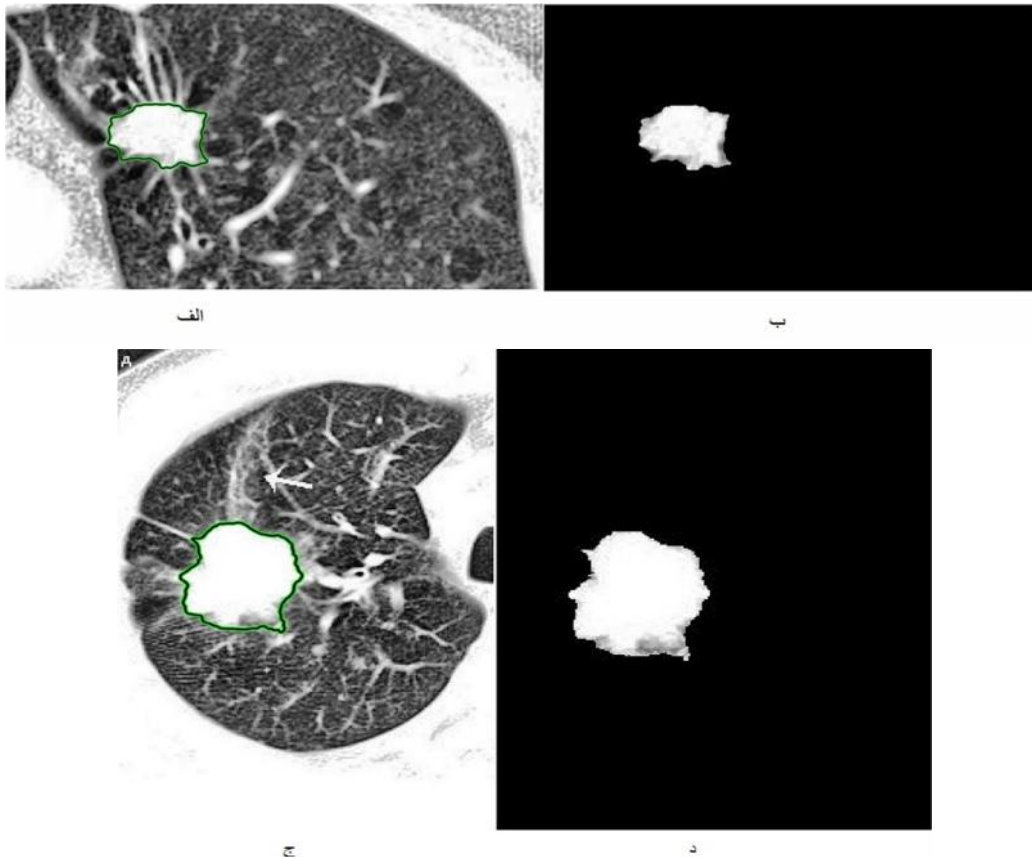
در نهایت، ناحیه قطعه‌بندی شده بریده شده و از تصویر خارج می‌شود و جهت اعمال الگوریتم‌های استخراج ویژگی، آماده می‌شود. در شکل ۵ نتایج حاضر از عملکرد این الگوریتم نمایش داده شده است.

**استخراج ویژگی:** ویژگی‌های بافتی شامل شاخص‌های آماری مرتبه اول، دوم، سوم و بالاتر و همچنین ویژگی‌های هندسی از نواحی قطعه‌بندی شده ندول‌ها استخراج می‌شوند. در مجموع از هر تصویر CT ریه، ۲۶ ویژگی استخراج شده‌است. تمام این ویژگی‌های استخراج شده، در یک فایل اکسل ذخیره می‌شوند. با توجه به اختلاف بازه مقادیر این ویژگی‌ها، در مرحله بعد، این ویژگی‌ها با الگوریتم نرمالیزاسیون مینیمم-ماکزیمم، نرمالیزه شده و مقداری بین ۰ تا ۱ پیدا می‌کنند.

**طبقه‌بندی:** در این بخش، غدد ریوی با استفاده از ویژگی‌های استخراج شده و طبقه‌بندی‌های SVM و KNN به دو دسته خوش‌خیم و بدخیم طبقه‌بندی

این تصاویر اعمال می‌شود که گیرایی و کنتراست تصویر را افزایش می‌دهد. در شکل ۴ نتایج حاصل از مرحله پیش‌پردازش آورده شده است.

**قطعه‌بندی:** پس از مرحله پیش‌پردازش و افزایش گیرایی تصویر، در این مرحله باید تصاویر قطعه‌بندی شده و ناحیه غده استخراج شود. الگوریتم چن-وسه، روش استفاده شده در این مرحله از پژوهش می‌باشد. در این روش، برای هر تصویر، دو نقطه بعنوان ناحیه مرزی غده توسط کاربر انتخاب می‌شود، سپس، الگوریتم چن-وسه ۱۰۰ بار بطور تکرار شونده بر روی این مرزها اعمال شده که در هر مرحله ناحیه مرز پیش‌روی می‌کند و مرز غده با دقت بسیار خوبی استخراج می‌شود. این روش که یک روش نیمه اتوماتیک می‌باشد و با توجه به اینکه در بسیاری از موارد مرز غدد کاملاً مشخص نمی‌باشد، با کمک گرفتن از کاربر در انتخاب دو نقطه مرزی، تاثیر بسیار شگرفی در افزایش دقت قطعه‌بندی ایجاد می‌کند.



**شکل ۵-** نمونه تصاویر CT-scan قبل و بعد از اعمال الگوریتم ناحیه بندی پیشنهادی ( تصاویر الف و ج ناحیه مرز ندول ها روی تصویر پیش پردازش شده و تصاویر ب و د ناحیه استخراج شده مرز ندولها)

استفاده نشده است. به همین دلیل، در هر طبقه‌بندی، جهت انتخاب بهترین ویژگی‌ها، ابتدا تمام ویژگی‌ها انتخاب و الگوریتم طبقه‌بندی اعمال می‌شود و سپس در هر مرحله یکی از ویژگی‌ها حذف می‌شود و عملکرد طبقه‌بندی بررسی می‌شود. اگر در هر مرحله حذف ویژگی، عملکرد طبقه‌بندی بهبود یابد، آن ویژگی حذف باقی می‌ماند، در غیر اینصورت ویژگی بازیابی شده و در طبقه‌بندی دخیل می‌شود.

**نتایج طبقه‌بندی بر حسب سه معیار اصلی:** در پژوهش‌های مرتبط با طبقه‌بندی و تشخیص بیماری‌ها با پردازش تصاویر پزشکی، معمولاً نتایج را با سه معیار، دقت (Accuracy)، حساسیت (Sensitivity) و میزان اختصاصی (Specificity) بودن بیان می‌کنند:

TP: تعداد غدد بدخیمی که بعنوان غدد بدخیم طبقه‌بندی شده‌اند.

FP: تعداد غدد خوش‌خیمی که بعنوان غدد بدخیم طبقه‌بندی شده‌اند.

FN: تعداد غدد بدخیمی که بعنوان غدد خوش‌خیم طبقه‌بندی شده‌اند.

TN: تعداد غدد خوش‌خیمی که بعنوان غدد خوش‌خیم طبقه‌بندی شده‌اند.

روابط بالا به راحتی از ماتریس درهم‌ریختگی (Confusion matrix)، که ماتریسی جهت بیان عملکرد مدل طبقه‌بندی می‌باشد و پراکندگی توزیع دسته‌ها را از لحاظ درستی یا نادرستی نمایش می‌دهد، قابل استخراج می‌باشند. شکل ۶ ماتریس درهم‌ریختگی باینری (دو کلاس) را نمایش می‌دهد.

با استفاده از تعاریف فوق، سه معیار دقت، حساسیت و میزان اختصاصی بودن، مطابق روابط زیر محاسبه می‌شوند:

می‌شوند. در ادامه، روند کلی این فرآیند تشریح می‌گردد.

آماده‌سازی داده‌ها: مرحله طبقه‌بندی خود شامل دو مرحله آموزش و آزمایش می‌باشد. معمولاً بخشی از داده‌های موجود جهت آموزش طبقه‌بند و بخش دیگر جهت آزمایش و مشاهده نتایج و عملکرد طبقه‌بند استفاده می‌گردد. در این پژوهش، نسبت داده‌های آموزش به آزمایش، ۷۰ به ۳۰ می‌باشد. داده‌های آموزش به طور تصادفی از کل داده‌ها انتخاب می‌شوند. این انتخاب تصادفی با الگوریتم برگزاری (HoldOut method) انجام می‌پذیرد. در این روش، داده‌ها با نسبت از قبل تعیین شده به طور تصادفی به دو دسته آموزش و آزمایش تقسیم‌بندی می‌شوند. الگوریتم‌های دیگری نیز جهت انجام اینکار وجود دارند. الگوریتم‌هایی مانند: روش اعتبارسنجی متقابل k تایی (K-fold cross validation)، اعتبارسنجی تصادفی زیرنمونه تکراری، اعتبارسنجی Leave-M-Out.

همانطور که گفته شد، با استفاده از الگوریتم برگزاری و انتخاب نسبت آموزش به آزمایش ۷۰ به ۳۰، الگوریتم‌های طبقه‌بندی بر روی داده‌های آموزش انجام می‌گیرد. بدین صورت، طبقه‌بند با دقت بالا جهت انجام طبقه‌بندی و تشخیص برچسب داده‌های تست آماده می‌شود. همچنین، در هر بار انتساب داده‌ها به دسته‌های آموزش و آزمایش، نتایج طبقه‌بندی ممکن است کمی متغیر باشد، به همین دلیل، تمام این مراحل ۱۰ بار تکرار شده و نتیجه هر مرحله ثبت و میانگین نتایج به عنوان کارایی نهایی آن طبقه‌بندی ثبت می‌گردد. بحث دیگری که در اینجا مطرح می‌شود، مساله انتخاب ویژگی است. همانطور که پیشتر گفته شد، از الگوریتم انتخاب ویژگی خاصی در این پژوهش

		مقادیر واقعی	
		مثبت	منفی
مقادیر پیش‌بینی شده	مثبت	درست مثبت TP	نادرست مثبت FP <small>خطای نوع یک</small>
	منفی	نادرست منفی FN <small>خطای نوع دو</small>	درست منفی TN

شکل ۶- ماتریس درهم‌ریختگی



عبارتند از: کرنل خطی، کرنل تابع پایه شعاعی، کرنل گوسی و کرنل چندجمله‌ای.

با استفاده از هر کدام از این کرنل‌ها طبقه‌بندی SVM روی کل دسته ویژگی‌های بافتی و هندسی و سپس تک تک دسته ویژگی‌ها اعمال شده و بهترین عملکرد هر حالت ثبت و بهترین کرنل که بیشترین مقدار معیارها در آن بدست آمده به عنوان طبقه‌بندی برنده انتخاب می‌گردد. در ادامه، در شکل ۷ مقادیر معیارها برای طبقه‌بندی ماشین بردار پشتیبان با تمام ویژگی‌ها آورده شده است. همچنین باید این نکته ذکر شود که در تمامی نمودارهای این فصل مقادیر سه معیار دقت، حساسیت و میزان اختصاصی بودن با سه رنگ آبی، نارنجی و زرد به ترتیب مشخص شده است. در ادامه، در شکل‌های ۸-۱۳، شش کرنل SVM ابتدا با در نظر گرفتن تک تک دسته ویژگی‌ها و سپس ترکیب دوتایی از دسته ویژگی‌ها نشان داده شده است. معیارهای طبقه‌بندی برتر در هر یک از حالت‌ها، در

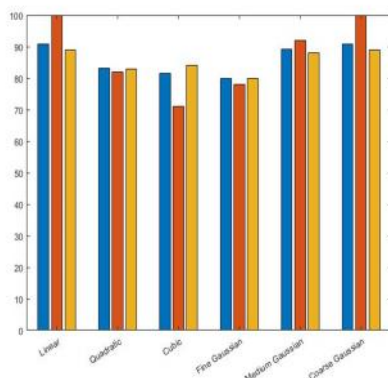
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (3)$$

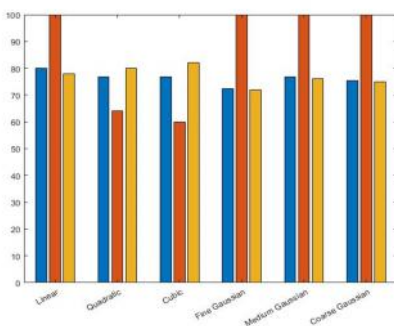
$$Specificity = \frac{TN}{TN+FP} \quad (4)$$

**نتایج طبقه‌بندی:** در این بخش، نتایج طبقه‌بندی داده‌های آزمایش و محاسبه سه معیار دقت، حساسیت و میزان اختصاصی بودن برای دو طبقه‌بندی ماشین بردار پشتیبان و k-نزدیکترین همسایه آورده شده است. قابل ذکر است که داده‌های آموزش و آزمایش ۱۰ بار بطور مختلف به نسبت ۷ به ۳ تقسیم‌بندی شده و مدل طبقه‌بندی هر بار بر داده‌های آزمایش اعمال می‌شود و معیارهای بدست آمده در هر بار ثبت شده و میانگین این معیارها بعنوان معیار آن طبقه‌بندی معرفی می‌گردد.

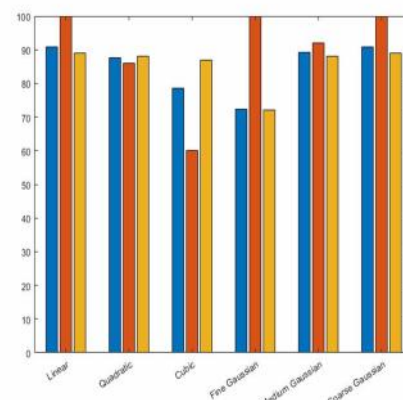
نتایج طبقه‌بندی ماشین بردار پشتیبان: طبقه‌بندی ماشین بردار پشتیبان دارای ۴ کرنل اصلی می‌باشد که



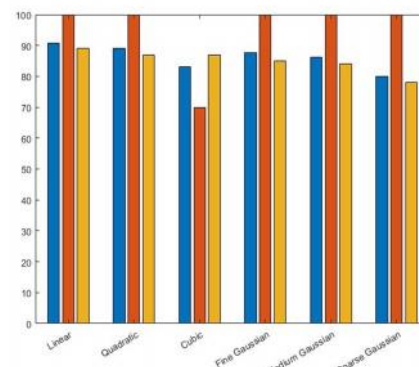
شکل ۹- مقایسه نتایج حاصل از شش طبقه بند SVM با ویژگی‌های بافتی



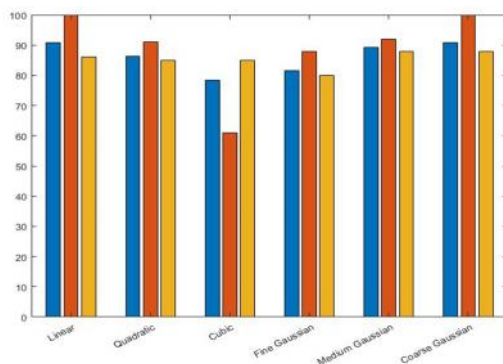
شکل ۱۰- مقایسه نتایج حاصل از شش طبقه بند SVM با ویژگی‌های هندسی



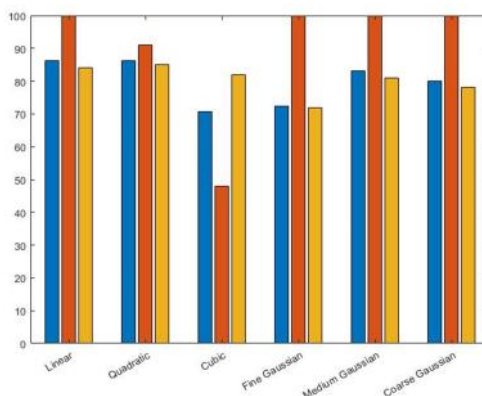
شکل ۷- مقایسه نتایج حاصل از شش طبقه بند SVM با تمام ویژگی‌ها



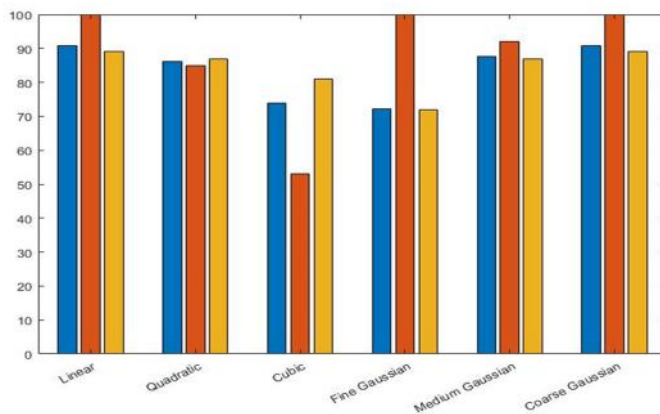
شکل ۸- مقایسه نتایج حاصل از شش طبقه بند SVM با ویژگی‌های هیستوگرام



شکل ۱۱- مقایسه نتایج حاصل از شش طبقه بند SVM با ویژگی‌های هیستوگرام و بافت



شکل ۱۲- مقایسه نتایج حاصل از شش طبقه بند SVM با ویژگی‌های هیستوگرام و هندسی



شکل ۱۳- مقایسه نتایج حاصل از شش طبقه بند SVM با ویژگی‌های هندسی و بافتی

عملکرد را دارند. در نهایت، در انتخاب ترکیب دو تایی از دسته ویژگی‌ها، برای دو حالت، هیستوگرام-بافت و بافت-هندسی، با کرنل‌های خطی و Coarse Gaussian، بیشترین معیارهای دقت، حساسیت و میزان اختصاصی بودن را بدست می‌دهند.

نتایج طبقه‌بندی K- نزدیکترین همسایه: در این بخش، با استفاده از طبقه‌بندی KNN، ویژگی‌ها طبقه‌بندی و معیارهای دقت، حساسیت و میزان

جداول ۱-۳ آمده است.

با توجه به نمودارها، پیداست که دو طبقه‌بندی خطی و گوسی Coarse بهترین نتیجه را می‌دهند. بنابراین در حالت انتخاب تمام ویژگی‌ها، این دو طبقه‌بندی بعنوان طبقه‌بندهای برتر SVM، انتخاب می‌شوند. در حالت انتخاب تک تک دسته ویژگی‌ها، دسته ویژگی هیستوگرام با کرنل خطی و دسته ویژگی بافت با کرنل‌های خطی و Coarse Gaussian بهترین

**جدول ۱- نتایج معیارهای دو طبقه بند برتر ماشین بردار پشتیبان با تمام ویژگیها**

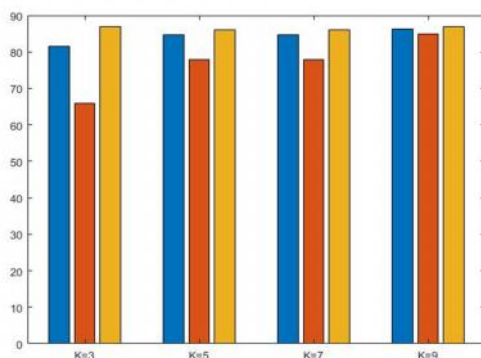
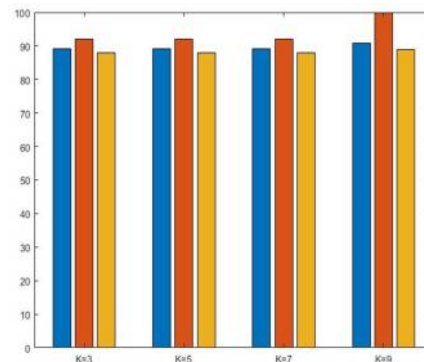
معیارها	Coarse Gaussian	Linear
دقت	۹۰/۸	۹۰/۸
حساسیت	۱۰۰	۱۰۰
میزان اختصاصی بودن	۸۹	۸۹

**جدول ۲- نتایج معیارهای دو طبقه بند برتر ماشین بردار پشتیبان با تک تک ویژگیها**

معیارها	هیستوگرام	بافت	هندسی
طبقه‌بند برتر	Linear	Linear , Coarse Gaussian	Linear
دقت	۹۰/۸	۹۰/۸	۸۰
حساسیت	۱۰۰	۱۰۰	۱۰۰
اختصاصی بودن	۸۹	۸۹	۷۸

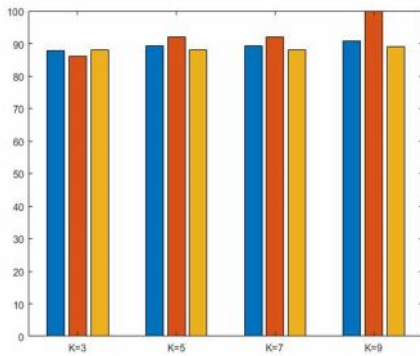
**جدول ۳- نتایج معیارهای دو طبقه بند برتر ماشین بردار پشتیبان با ترکیب دوتایی از ویژگیها**

معیارها	هیستوگرام+بافت	هیستوگرام+هندسی	هندسی+بافت
طبقه‌بند برتر	Linear , Coarse Gaussian	Linear	Linear , Coarse Gaussian
دقت	۹۰/۸	۸۶/۲	۹۰/۸
حساسیت	۱۰۰	۱۰۰	۱۰۰
اختصاصی بودن	۸۹	۸۴	۸۹

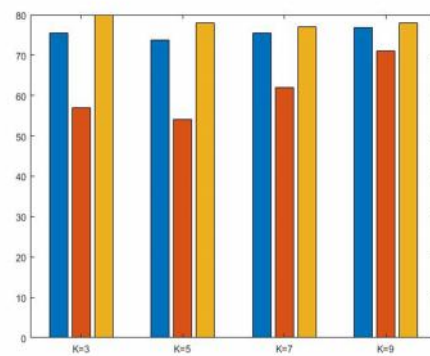
**شکل ۱۵- مقایسه عملکرد سیستم در طبقه بند K نزدیکترین همسایه وزندار در فواصل  $K=3,5,7,9,11$  با ویژگی‌ها هیستوگرام****شکل ۱۴- مقایسه عملکرد سیستم در طبقه بند K نزدیکترین همسایه وزندار در فواصل  $k=3,5,7,9,11$  با تمام ویژگی‌ها**

شده است. با بررسی این نمودارها، می‌توان دریافت که در حالت انتخاب تمام ویژگی‌ها، انتخاب تعداد ۹ نزدیکترین همسایه، یعنی  $K=9$  بهترین معیارها را بدست می‌دهد. همچنین در انتخاب هر دسته ویژگی به تنهایی، ویژگی‌های بافتی با انتخاب  $K=9$ ، بهترین نتیجه را می‌دهند. در نهایت، در انتخاب دو دسته ویژگی، ترکیب دسته ویژگی‌های بافت و هندسی برای  $K=7,9$  بیشترین میزان دقت، حساسیت و میزان

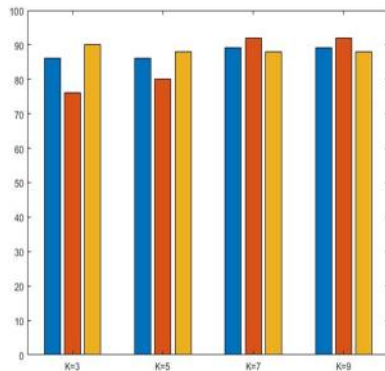
اختصاصی بودن محاسبه می‌شوند. در طبقه‌بندی KNN مقدار چهار مقدار ۳، ۵، ۷ و ۹ بعنوان پارامتر K انتخاب و معیار فاصله از نوع اقلیدسی می‌باشد. در این بخش نیز، همانند طبقه‌بندی SVM، ابتدا تمام ویژگی‌ها، سپس تک تک ویژگی‌ها و در آخر ترکیب دوتایی از ویژگی‌ها جهت طبقه‌بندی انتخاب و طبقه‌بندی برتر (در اینجا K برتر) انتخاب می‌گردد. نتایج طبقه‌بندی در شکل‌های ۱۴-۲۰ نشان داده



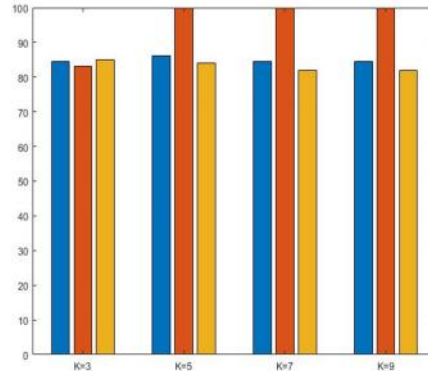
شکل ۱۷- مقایسه عملکرد سیستم در طبقه بند k نزدیکترین همسایه وزندار در فواصل  $k=3,5,7,9,11$  با ویژگی‌ها هندسی



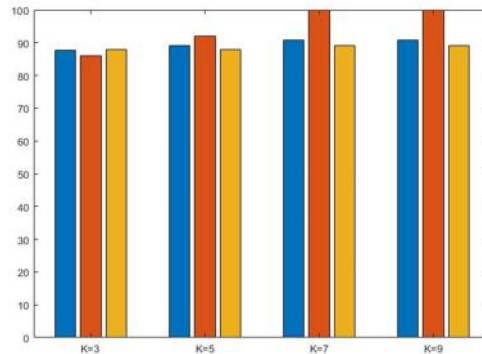
شکل ۱۶- مقایسه عملکرد سیستم در طبقه بند k نزدیکترین همسایه وزندار در فواصل  $k=3,5,7,9,11$  با ویژگی‌ها بافتی



شکل ۱۹- مقایسه عملکرد سیستم در طبقه بند k نزدیکترین همسایه وزندار در فواصل  $k=3,5,7,9,11$  با ویژگی‌ها هیستوگرام و بافت



شکل ۱۸- مقایسه عملکرد سیستم در طبقه بند k نزدیکترین همسایه وزندار در فواصل  $k=3,5,7,9,11$  با ویژگی‌ها هیستوگرام و بافت



شکل ۲۰- مقایسه عملکرد سیستم در طبقه بند k نزدیکترین همسایه وزندار در فواصل  $k=3,5,7,9,11$  با ویژگی‌ها هندسی و بافتی

آمده‌است.

### بحث و نتیجه‌گیری

در این پژوهش، روشی جدید و کارآمد جهت طبقه‌بندی غدد ریوی از تصاویر CT ریه پیشنهاد گردید. در الگوریتم پیشنهادی، ابتدا الگوریتم CLAHE

اختصاصی بودن را نتیجه می‌دهد که این معیارها برابر با حالت انتخاب تمام ویژگی برای  $K=9$  می‌باشد. بنابراین در طبقه‌بندی KNN، حالت ترکیب دو دسته ویژگی بافت و هندسی با  $K=7, 9$  بعنوان طبقه‌بندی برتر انتخاب می‌گردد. مقادیر معیارهای هر دسته ویژگی و K های مختلف بصورت درصدی در جدول‌های ۴-۶

**جدول ۴- نتایج معیارهای طبقه بند برتر KNN با تمام ویژگی‌ها**

معیارها	K=9
دقت	۹۰/۸
حساسیت	۱۰۰
میزان اختصاصی بودن	۸۹

**جدول ۵- نتایج معیارهای طبقه بند برتر KNN با تک تک ویژگی‌ها**

معیارها	هیستوگرام	بافت	هندسی
طبقه‌بند برتر	K=9	K=9	K=9
دقت	۸۶/۲	۹۰/۸	۷۶/۹
حساسیت	۸۵	۱۰۰	۷۱
میزان اختصاصی بودن	۸۷	۸۹	۷۸

**جدول ۶- نتایج معیارهای طبقه بند برتر KNN با ترکیب دوتایی از ویژگی‌ها**

معیارها	هیستوگرام+بافت	هیستوگرام+هندسی	هندسی+بافت
طبقه‌بند برتر	K= 7, 9	K= 5	K= 7, 9
دقت	۸۹/۲	۸۶/۲	۹۰/۸
حساسیت	۹۲	۱۰۰	۱۰۰
میزان اختصاصی بودن	۸۸	۸۴	۸۹

طبقه‌بند ماشین بردار پشتیبان استفاده کردند. آنها با این الگوریتم توانستند ندولهای بدخیم و خوش خیم را با دقت ۸۶٫۶ درصد طبقه بندی کنند (۱۶). پس از حذف نویز تصاویر CT با روش فیلتر میانه وزندار و افزایش گیرایی تصویر با روش هیستوگرام تطبیقی، اقدام به قطعه بندی و جداسازی ناحیه ندول از بخش‌های دیگر تصویر نمودند. آنها برای اینکار از تکنیک خوشه‌یابی فراوان بهبودیافته (Improved profuse clustering) IPCT-technique که بر مبنای شباهت بین پیکسلی عمل می‌کند، استفاده کردند. در مرحله بعد، از ناحیه ندول جدا شده، ویژگی‌های میانگین، چولگی ممان سوم (Third moment skewness)، انحراف معیار، کجی ممان چهارم (Fourth moment kurtosis) استخراج کردند. در مرحله آخر، با استفاده از یادگیری عمیق شبکه عصبی آموزش داده شده بلا درنگ (Deep learning instantaneously trained) اقدام به طبقه‌بندی ندول‌ها به دو دسته سرطانی و غیرسرطانی نمودند.

با مقایسه نتایج حاصل از روش پیشنهادی با روش‌های ارائه شده پیشین، روش پیشنهادی بیشترین حساسیت، و در بسیاری از پژوهش‌ها بیشترین دقت

و فیلتر دولبه بر روی تصاویر CT ریه اعمال شد و تصاویر پیش پردازش گردید. در مرحله بعد، از کانتور فعال چن-وسه جهت جداسازی ناحیه ندول استفاده گردید. سپس، ویژگی‌های مختلف بافتی، هندسی و هیستوگرام از ناحیه قطعه‌بندی شده ندول استخراج شد. در نهایت، با استفاده از ویژگی‌های مذکور، ندول‌ها به دو دسته خوش خیم و بدخیم طبقه‌بندی شدند. طبقه‌بند الگوریتم پیشنهادی، توانست ندول‌ها را با دقت، حساسیت و میزان اختصاصی بودن به ترتیب ۹۰/۸٪، ۱۰۰٪، ۸۹٪ طبقه‌بندی کند. استفاده از این روش در مراکز درمانی و تصویربرداری CT، می‌تواند کمک شایانی به پزشکان و متخصصان رادیولوژی نماید و باعث افزایش دقت تشخیص و صرفه‌جویی در زمان گردد.

پس از پیش پردازش تصویر با استفاده از فیلتر گوسی و میانه و حذف نویز و سایر اغتشاشات، از روش جغرافیای رنگ جهت ناحیه بندی جهت ناحیه بندی و تشخیص ندول استفاده کردند. در مرحله بعد، ویژگی‌های مساحت، محیط، میزان بیضی بودن، قطر ندول، مرکزیت و میانگین شدت روشنایی را استخراج کردند. در مرحله بعد، از ویژگی‌های استخراج شده جهت آموزش و تست



طبقه‌بندی نهایی در خصوص خوشخیم و یا بدخیم بودن غدد انجام میپذیرد. به دلیل وجود تعداد زیاد ویژگی، لازم است تا از این میان بهترین و بارزترین ویژگیها جهت طبقه بندی صحیح انتخاب شود. روشهای مختلفی جهت انتخاب ویژگی وجود دارد ولی به دلیل وقتگیر بودن این فرآیند، در پژوهش حاضر، این مرحله حذف میشود و ابتدا تمام ویژگیها به طبقه بند داده میشود و سپس با انجام سعی و خطا بهترین ویژگیها برای هر طبقه بند انتخاب می‌گردد. بنابراین با انجام این کار، همزمان انتخاب ویژگی و طبقه بندی انجام می‌گیرد و از بار محاسباتی و زمان پردازش کاسته میشود. در این پژوهش، از ویژگیهای استخراج شده جهت طبقه بندی، به دو طبقه بند معروف ماشین بردار پشتیبان و پارامتر نزدیکترین همسایه داده میشود.

دسترسی به پایگاه داده با جامعه آماری یکنواخت از غدد بدخیم و خوشخیم یکی از اساسی ترین مراحل پیاده سازی و ارزیابی سیستم‌های تشخیصی کامپیوتری برای بیماری‌های سرطانی می‌باشد. در پژوهش حاضر، از مجموعه تصاویر کنسرسیون پایگاه داده تصاویر ریه استفاده شده است. این پایگاه داده شامل تصاویر CT اسکن سرطان ریه به همراه نظر تشخیصی پزشک متخصص و ناحیه مشخص شده غدد می‌باشد. این پایگاه داده توسط سازمان ملی سرطان آمریکا، سازمان ملی سلامت و سازمان غذا و دارو گردآوری و در دسترس عموم قرار گرفته است. در این پایگاه داده، تصاویر CT هر بیمار با فرمت DICOM که استاندارد ذخیره سازی و مناسب جهت پردازش تصاویر پزشکی است، ذخیره شده است. میانگین برشهای هر اسکن برای هر بیمار ۲۵۴ برش می‌باشد که فاصله هر بین برشهای ۹ الی ۹ میلی‌متر می‌باشد. این پایگاه داده دارای سه فایل اکسل میباشد که در فایل اول / مجاور بین اطلاعاتی همچون تعداد ندولهای هر بیمار به همراه اندازه هر ندول، منشا اصلی پخش شده ندول و تشخیص نهایی رادیولوژیست و اقدام درمانی انجام شده برای هر بیمار آورده شده است. در فایل دوم اطلاعات عمومی همچون تاریخ انجام هر اسکن، نام شرکت سازنده دستگاه CT به همراه مدل دستگاه، ورژن نرمافزار و شناسه تصاویر میباشد. در فایل اکسل سوم، تعداد ندولهای بزرگتر از ۹ میلی‌متر و کوچکتر از ۹ میلی‌متر برای هر بیمار آورده شده است.

طبقه بندی را نیز بدست آورد. باتوجه به اینکه معیار حساسیت به معنی توانایی طبقه بندی در تشخیص صحیح بیماری در شخص و معیار میزان اختصاصی بودن به معنی توانایی طبقه بندی در تشخیص صحیح عدم وجود آن بیماری در شخص می‌باشد، بنابراین می‌توان نتیجه گرفت که معیار حساسیت در پژوهش‌های مرتبط با تشخیص غدد در تصاویر پزشکی از اهمیت بالایی برخوردار است. این بدین دلیل است که تشخیص درست وجود یک بیماری یا سرطان بسیار مهم تر و حیاتی تر از تشخیص عدم وجود آن در فرد کاندید می‌باشد. بنابراین، روش پیشنهادی بخاطر دارا بودن حساسیت بسیار بالا و همچنین دارا بودن مقادیر مطلوب دو معیار دقت و میزان اختصاصی بودن و تعداد پایین ویژگی‌های مورد استفاده جهت طبقه بندی، بعنوان یک روش کارآمد و مناسب جهت طبقه بندی غدد ریوی پیشنهاد می‌گردد.

پس از قطعه بندی و اسلترجراحی ناحیه مورد نیاز از سایر قسمت‌های تصویر، لازم است تا ویژگی‌هایی از ناحیه مورد نظر استخراج شود. این ویژگیها، اطلاعات مفیدی از تصویر به ما میدهند که در ساده سازی و کاهش حجم داده ها نقش بسزایی دارند. حجم داده بالا نیاز به حافظه و سیستمی با قدرت محاسباتی بالا دارد که ایده آل نیست و با کمک استخراج ویژگی، حجم داده به طور بسزایی کاهش می‌یابد. روشهای مختلفی جهت استخراج ویژگی معرفی شده اسلت که بهترین آنها الگوریتمی است که بتواند داده ها را با دقت بالایی توصیف کند. ویژگیهای استخراج شده در این پژوهش عبارتند از: ویژگیهای هیستوگرام که شامل ۷ ویژگی میباشد، ویژگیهای بافتی که خود شامل ۸۸ ویژگی و ویژگیهای هندسی که خود شامل 1 ویژگی است. همانطور که در بخش قبل ذکر شد، روش پیشنهادی با دقت قابل قبولی توانست ندول‌های خوشخیم و بدخیم ریوی را طبقه بندی کند. اما همچنان با ایجاد تغییراتی می‌توان این سیستم را بهبود داد. پیشنهادات و روش‌های زیاد جهت بهبود این الگوریتم وجود دارد که در این بخش به برخی از این روش‌ها اشاره می‌گردد.

پس از استخراج ویژگیها از نواحی قطعه بندی شده و نرمالیزه کردن آنها، با تعداد زیادی داده (ویژگی) مواجه هستیم که با استفاده از این داده‌ها تصمیمگیری و

Lung Cancer. *J Cancer*. 2020;11(12):3615–3622.

2. Zhao XZ, Liu LY, Qi SL, Teng YY, Li JH, Qian W. Agile convolutional neural network for pulmonary nodule classification using CT images. *Comput Assist Radiol Surg*. 2018;13:585–595.

3. Nobuhiro K, Yoshimasa T, Ryou I, Naoki W. Characteristics of multiple nodules in a patient with pulmonary Langerhans cell histiocytosis: A case report. *World J Clin Cases*. 2019;7(17):2519–2525.

4. Gu Y. Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach. *Pattern Recog*. 2013;46(3):692-702.

5. Kulkarni A, Panditrao A. Classification of lung cancer stages on CT scan images using image processing. in 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies. 2014:56-62.

6. Shen W. Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recog*. 2017;61:663-673.

7. Makaju S. Lung cancer detection using CT scan images. *Procedia Comput Sci*. 2018;125:107-114.

8. Liu X. Multi-view multi-scale CNNs for lung nodule type classification from CT images. *Pattern Recog*. 2018;77:262-275.

9. de Carvalho Filho AO, Corrêa Silva A, de Paiva AC, Nunes RA, Gattass M. Classification of patterns of benignity and malignancy based on CT using topology-based phylogenetic diversity index and convolutional neural network. *Pattern Recognition*, 2018. 81: p. 200-212.

10. Jamshidi-Naeni Y, Bavi AK, Egal A, Oldewage-Theron W. Hemoglobin and ferritin concentrations are positively associated with blood pressure and hypertension risk in older adults: a retrospective cross-sectional study, Sharpeville, South Africa. *Asia Pac J Clin Nutr*. 2019;28(3):533-543.

11. Naeni YJ, Davoodi H, Esmaeili S. Effects of vitamin D on risk of breast cancer. *Iran J Nutr Sci Food Technol*. 2013;7(4, Suppl.).

12. Ali I, Muzammil M, Haq IU, Khaliq AA, Abdullah S. Efficient Lung Nodule Classification Using Transferable Texture Convolutional Neural Network. *IEEE Access*. 2020 Sep 23;8:175859-70..

13. Jalapally P., Suvarna Vani K., Praveen Kumar K., Koduru J. Lung Cancer Detection Using CT Scan Image. In: Jyothi S., Mamatha D., Satapathy S., Raju K., Favorskaya M. (eds) *Advances in Computational and Bio-Engineering*. CBE 2019. Learning and Analytics in Intelligent Systems, 2020; vol 16. Springer, Cham. [https://doi.org/10.1007/978-3-030-46943-6\\_27](https://doi.org/10.1007/978-3-030-46943-6_27).

14. Dehmeshki J, Ye X, Lin X, Valdivieso M, Amin H. Automated detection of lung nodules in CT images using shape-based genetic algorithm. *Comput Med Imag Graphics*. 2017;31:408-17.

هر اسکن توسط ۴ رادیولوژیست به طور مجزا بررسی شده و اسکنهایی که توسط هر ۴ رادیولوژیست تشخیص یکنسانی دارند، به این پایگاه داده اضافه گردیده است. متخصصین هر ندول را به یکی از چهار دسته ناشناخته، خوشخیم، بدخیم ابتدایی و بدخیم متاستازیک تقسیم‌بندی کرده‌اند. در این پژوهش، برای هر بیمار، برشی که ندول در آن نمایان است انتخاب می‌گردد. همچنین، باتوجه به اینکه ندولها به ۴ دسته طبقه بندی شده اند، ما در این پژوهش، دسته ناشناخته و خوشخیم را بعنوان دسته خوشخیم و دسته بدخیم ابتدایی و بدخیم متاستازیک را بعنوان دسته بدخیم طبقه بندی کرده ایم. در این پژوهش، تصاویر ریه ۶۵ بیمار سرطانی از پایگاه داده مذکور مورد طبقه بندی قرار گرفته که از این تعداد، ۴۹ بیمار دارای ندولهای بدخیم و ۲۵ بیمار دارای ندولهای خوشخیم میباشد.

در الگوریتم پیشنهادی از یک روش نیمه خودکار جهت قطعه‌بندی ناحیه ندول ریه استفاده شده است. با استفاده از یک الگوریتم طبقه‌بندی خودکار که نیاز به انتخاب دو نقطه مرز غده نداشته باشد، می‌تواند سرعت قطعه‌بندی را افزایش دهد. البته مزیتی که روش قطعه‌بندی نیمه‌خودکار دارد، دقت طبقه‌بندی بالای آن می‌باشد که این دقت در قطعه‌بندی خودکار به مراتب پایین تر می‌باشد. بنابراین پیشنهاد و پیاده‌سازی یک الگوریتم قطعه‌بندی خودکار که تواما از دقت بالایی نیز برخوردار باشد، در راس کار برنامه‌های آتی این پروژه قرار دارد. در این پژوهش، غدد ریه به دو دسته خوشخیم و بدخیم طبقه‌بندی شدند. با توجه به اینکه هر کدام از این دسته‌ها به دو زیرشاخه تقسیم‌بندی می‌شوند، یکی از کارهای آتی این پژوهش طبقه‌بندی غدد به هر یک از ۴ مرحله پیشرفت سرطان می‌باشد. همچنین استفاده از ویژگی‌های آماری مختلف مانند سن، جنسیت، محل زندگی و رژیم غذایی و تلفیق این ویژگی‌ها با سایر ویژگی‌های استفاده شده در این مقاله، از فعالیت‌های آتی این پژوهش می‌باشد.

## References

1. Lei C, Wanbing F, Zhigang Y, Xiaoming Z, Wei X. Deep Learning Model as a New Trend in Computer-aided Diagnosis of Tumor Pathology for

15. Tan J, Huo Y, Liang Z, Li L. Expert knowledge-infused deep learning for automatic lung nodule detection. *J X-ray Sci Technol*. 01 Jan 2019;27(1):17-35.

16. Shakeel PM, Burhanuddin M, Desa MI. Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks. *Measurement*. 2019;145:702-712.