مجله علوم پزشکی رازی دوره ۲۶، شماره ۱۰، دی ۱۳۹۸ http://rjms.iums.ac.ir

مقاله پژوهشی

روش جدید بر پایه تحلیل مؤلفه اساسی و پرسپترون چندلایه با بهینه‌سازی الگوریتم ژنتیک برای تشخیص بیماری سرطان ریه

علی شریفی: دانشجوی دکتری شیمی‌تجزیه، گروه شیمی، دانشکده علوم پایه، دانشگاه لرستان، ایران

کمال علیزاده: دانشیار، گروه شیمی، دانشکده علوم پایه، دانشگاه لرستان، ایران (*نویسنده مسئول). alizadeh.k@lu.ac.ir

چکیده

زمینه و هدف: سرطان ریه جزء اولین سرطان‌های شناخته شده محسوب می‌گردد. تشخیص زودهنگام این سرطان سبب کاهش طول درمان و صرف هزینه‌های سرسام آور درمانی و عامل اصلی در بقاء و زنده ماندن افراد است. در سال‌های اخیر بهره‌مندی از روش‌های کامپیوتری در استفاده از داده‌کاوی و الگوریتم‌های هوشمند، سبب تسریع در تشخیص زودهنگام این سرطان شده است. هدف این مقاله ارزیابی نقش روش جدید بر پایه تحلیل مؤلفه اساسی و پرسپترون چندلایه با بهینه‌سازی الگوریتم ژنتیک برای تشخیص بیماری سرطان ریه مد نظر است.

روش کار: در این مطالعه، داده‌های مورد استفاده از پایگاه داده یادگیری ماشین UCI شامل ۳۲ پرونده بیمار با ۵۷ ویژگی استخراج شد. پس از انجام مراحل پیش‌پردازش، در مرحله استخراج ویژگی و کاهش ابعاد داده، داده‌های اصلی سرطان ریه با استفاده از تحلیل مؤلفه اساسی به ۱۷ ویژگی کاهش یافت. سپس در مرحله طبقه‌بندی این ویژگی‌های کاهش داده شده به پرسپترون چندلایه با بهینه‌سازی الگوریتم ژنتیک داده شد و حساسیت و ویژگی مدل با توجه به معیارهای دقت، حساسیت و صحت بررسی گردید. کلیه تحلیل‌ها با استفاده از نرم‌افزار MATLAB و SPSS صورت گرفت.

یافته‌ها: برای مدل پیشنهادی، نتایج حاصل از شبیه‌سازی‌ها به ترتیب میانگین پارامترهای دقت طبقه‌بندی، حساسیت و صحت، ۹۸/۶۵، ۹۸، ۹۹/۱۶ درصد بدست آمد.

نتیجه‌گیری: نتایج حاصل از داده‌های واقعی نشان می‌دهد که سیستم پیشنهادی در تشخیص سرطان ریه بسیار مؤثرتر و سریع‌تر از سایر روش‌ها بوده است و می‌تواند برای برنامه‌های کاربردی بالینی به عنوان دستیار پزشک مورد استفاده قرارگیرد.

تعارض منافع: گزارش نشده است.

منبع حمایت کننده: دانشگاه لرستان

کلیدواژه‌ها

سرطان ریه،
تحلیل مؤلفه اساسی،
شبکه عصبی مصنوعی،
ساختار پرسپترون چند لایه،
الگوریتم ژنتیک

تاریخ دریافت: ۹۸/۰۴/۲۲

تاریخ پذیرش: ۹۸/۰۹/۰۹

شیوه استناد به این مقاله:

Sharifi A, Alizadeh K. A novel technique based on principal component analysis and multi-layer perceptron with genetic algorithm optimization for diagnosis of lung cancer. Razi J Med Sci. 2019;26(10):48-56.

Razi Journal of Medical Sciences.2019;26(10):48-56. http://rjms.iums.ac.ir

Original Article

A novel technique based on principal component analysis and multi-layer perceptron with genetic algorithm optimization for diagnosis of lung cancer

Ali Sharifi, PhD Student in Analytical Chemistry, Department of Chemistry, Lorestan University, Khorramabad, Iran
Kamal Alizadeh, Associate Professor, Department of Chemistry, Faculty of Basic Science, Lorestan University, Khorramabad, Iran (*Corresponding author). alizadeh.k@lu.ac.ir

Abstract

Background: Lung cancer was known as primary cancers. Early detection of lung cancer reduces the length of treatment and spends a great deal of cost on the survival and survival of the individual. In recent years, the use of computer techniques in the use of data mining and intelligent algorithms has accelerated the early diagnosis of this cancer. The purpose of this paper is to evaluate the role of the new method based on Principal Component Analysis and Multi-Layer Perceptron with Genetic Algorithm optimization for Diagnosis of Lung Cancer.
Methods: In this study, the lung cancer dataset used was taken from the UCI machine learning database, including 32 patient records with 57 features. After performing its preprocessing steps, in the process of extraction of features and reduction of data dimensions, the main data of lung cancer were reduced to 17 characteristics using a basic component analysis. Then, in the classification step, these characteristics were reduced to multilayer perceptron by optimizing the genetic algorithm and the sensitivity and specificity of the model were studied according to the accuracy, sensitivity and Specificity. All analysis and synthesis were performed using the software of MATLAB and SPSS.
Results: For the proposed model, the results of the simulations were the mean of classification accuracy, sensitivity and specificity, respectively, 98.86, 98 and 99.16%.
Conclusion: The results on real data indicate that the proposed system is very effective in the diagnosis of lung cancer and can be used for clinical applications.

Conflicts of interest: None
Funding: Lorestan University

Keywords
Lung cancer,
Principal component analysis,
Artificial neural network,
Multilayer perceptron,
Genetic algorithm

Received: 13/07/2019
Accepted: 30/11/2019

Cite this article as:
Sharifi A, Alizadeh K. A novel technique based on principal component analysis and multi-layer perceptron with genetic algorithm optimization for diagnosis of lung cancer. Razi J Med Sci. 2019;26(10):48-56.

مقدمه

سرطان ریه، سرطانی است که در بافت ریه و معمولاً سلول‌های پوشاننده مجاری هوایی ایجاد می‌شود. دو نوع عمده سرطان ریه شامل سرطان ریه سلول کوچک و سرطان ریه غیرسلول کوچک است. این انواع سرطان ریه براساس نمای ظاهری سلول‌ها زیر میکروسکوپ طبقه‌بندی می‌شوند (۱). این نوع سرطان در سراسر دنیا از اهمیت بالایی در نظام سلامت کشورها برخوردار بوده و می‌توان گفت یکی از مهم‌ترین و شایع‌ترین علت مرگ ناشی از سرطان در مردان و زنان محسوب می‌شود (۲). در کشور ایران نیز متأسفانه این سرطان در حال گسترش در بین مردان و زنان می‌باشد (۳). سرطان ریه در ایران بعد از سرطان معده دومین علت مرگومیر افراد به علت سرطان است (۴).

با توجه به عوامل متعدد در تشخیص این بیماری، یکی از اساسی‌ترین مشکلات، تشخیص دیرهنگام آن در مراحل پیشرفته و خطرناک بوده و امکان جراحی و درمان موفقیت‌آمیز در آن بسیار پایین است (۲). بنابراین تشخیص سریع این سرطان در مراحل اولیه، شانس درمان و طول عمر بیمار را بالا می‌برد بعلاوه در هزینه‌های بالای درمان برای بیماران مبتلا به آن صرفه‌جویی می‌شود. از آنجا که در تحقیقات مربوط به حوزه پیش‌بینی و تشخیص بیماری، مسئله سلامت انسان مطرح است، پیش‌بینی درست نتایج اهمیت بیشتری می‌یابد بنابراین باید از روش‌هایی استفاده شود که تشخیص بر اساس آنها دارای کمترین خطا و بیشترین دقت، صحت و اطمینان باشد. بهره‌گیری از روش‌های هوشمند تشخیص بیماری سرطان ریه می‌تواند گام مؤثری در جهت تشخیص زود هنگام این بیماری باشد. یکی از نوین‌ترین روش‌ها برای بهره‌وری موثر و سریع از داده‌های حجیم استفاده از داده‌کاوی و شبکه‌های عصبی مصنوعی است که این روش‌ها با هدف تولید مدل‌ها و الگوهای پیشبینی کننده در زمینه‌های مختلف پزشکی رواج یافته و مدل‌های کامپیوتری و تکنیک‌های داده‌کاوی می‌توانند راهنمای خوبی برای پزشکان باشد. در زمینه تشخیص بیماری سرطان ریه با استفاده از روش‌های داده‌کاوی طی سال‌های اخیر کارهایی انجام شده که خود تاییدی بر این است که روش‌های نوین داده‌کاوی و بهره‌گیری از شبکه هوش مصنوعی، ابزار موثری را برای شناسایی روابط مهم سلامت از درون پرونده‌های پزشکی فراهم می‌کند که در این قسمت به چند نمونه آن اشاره می‌کنیم.

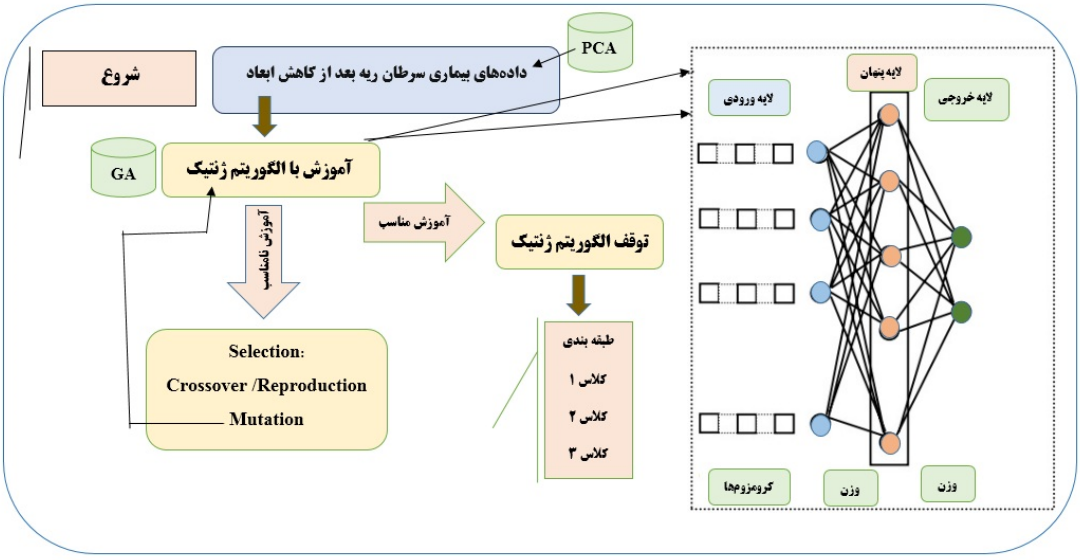
اباذری و همکارانش در سال ۲۰۱۵ به ارزیابی متغیرهای پیش‌آگهی در رده‌بندی نرخ بقای بیماران مبتلا به سرطان ریه با استفاده از درخت تصمیم‌گیری در استان آذربایجان‌غربی پرداختند و با استفاده از مدل درخت تصمیم‌گیری متغیرهای وضعیت درمان، وضعیت سیگاری بودن، سن تشخیص بیماری و مرحله‌ی تومور به عنوان متغیرهای مهم شناسایی کردند (۵). دلیری در سال ۲۰۱۲ یک سیستم اتوماتیک ترکیبی برای تشخیص سرطان ریه بر اساس الگوریتم ژنتیک و ماشین یادگیری حداکثر فازی (ELM) ارائه داده و میزان دقت به دست آمده با روش فوق را ۹۸/۸۵٪ گزارش کرده است (۶).

تعداد مطالعات صورت گرفته در خصوص تشخیص سرطان ریه با روش‌های داده‌کاوی و شبکه عصبی مصنوعی در ایران بسیار محدود است لذا در این مقاله، روشی جدید مبتنی بر تحلیل مؤلفه اساسی به عنوان روشی جهت کاهش ابعاد داده و شبکه عصبی مصنوعی با ساختار پرسپترون چند لایه با بهینه‌سازی الگوریتم ژنتیک به عنوان الگوریتمی جهت پردازش اطلاعات، برای تشخیص بیماری سرطان ریه پیشنهاد شده است و در پایان میزان دقت تشخیص بدست آمده با سایر روش‌ها مقایسه شده است (شکل ۱).

ساختار کلی شبکه عصبی مصنوعی مبتنی بر ساختار پرسپترون چند لایه مبتنی بر الگوریتم ژنتیک برای طبقه بندی داده‌های بیماری سرطان ریه در این مقاله نشان داده شده است.

روش کار

در این پژوهش به ارائه مدلی برای تشخیص بیماری سرطان ریه بااستفاده از پایگاه داده UCI پرداخته شد.



شکل ۱- ساختار کلی شبکه عصبی مصنوعی مبتنی بر ساختار پرسپترون چند لایه مبتنی بر الگوریتم ژنتیک برای طبقه‌بندی سرطان ریه

مدل ارائه شده دارای چهار مرحله جمع‌آوری داده، پیش‌پردازش و نرمال‌سازی داده‌ها، کاهش ابعاد داده بااستفاده از تحلیل مؤلفه اساسی و پردازش اطلاعات و طبقه‌بندی با استفاده از پرسپترون چندلایه با بهینه‌سازی الگوریتم ژنتیک است که در ادامه به تشریح هریک از مراحل پرداخته می‌شود.

جمع‌آوری داده: پایگاه داده سرطان ریه مورد استفاده در این مقاله شامل ۳۲ مورد پرونده بیمار که هریک دارای ۵۷ ویژگی بودند که ویژگی نهایی مربوط به کلاس‌بندی آنهاست استفاده شد. ویژگی‌های استخراج شده بر اساس داده‌های بالینی و داده‌های اشعه X این بیماران به دست آمده است و تمام مقادیر این ویژگی‌ها از ۰ تا ۳ مشخص شده بود. این پایگاه داده مربوط به سه نوع سرطان ریه بوده که در سه کلاس دسته‌بندی شده است. کلاس اول شامل ۹ نمونه، کلاس دوم شامل ۱۳ نمونه و کلاس سوم شامل ۱۰ نمونه مورد استفاده قرار گرفت. مجموعه داده‌های مربوط به بیماران مبتلا به سرطان ریه موجود در انبار داده یادگیری ماشین دانشگاه ایروین، کالیفرنیا (UCI: University of California, Irvine)استفاده شد (۷). کلیه تحلیل‌ها با استفاده از نرم‌افزار MATLAB و SPSS صورت گرفت.

پیش‌پردازش و نرمال‌سازی داده‌ها: پیش‌پردازش و پاکسازی داده با هدف ارتقاءکیفیت داده‌ها به کار بسته می‌شوند. قالب مناسب داده‌ها به عنوان ورودی داده‌کاوی در نتایج و خروجی تأثیرگذار است. اگر مقادیر ویژگی‌های مجموعه داده در دامنه متفاوتی قرار داشته باشند، احتمال بروز خطا در یافته‌ها افزایش می‌یابد. به قرار دادن داده‌های یک جامعه آماری در دامنه مشابه، نرمال‌سازی گفته می‌شود (۸). مقادیر اندازه‌گیری شده به محدوده‌ی جدیدی تبدیل شدند. این محدوده در این مقاله ۱±تعریف شد.

کاهش ابعاد داده: تحلیل مؤلفه اساسی (PCA: Principal Component Analysis) روشی برای کاهش ابعاد داده‌ها می‌باشد (۹). تحلیل مؤلفه‌های اساسی در تعریف ریاضی (۱۰) یک تبدیل خطی متعامد است که داده را به دستگاه مختصات جدید می‌برد به‌طوری‌که بزرگترین واریانس داده بر روی اولین محور مختصات، دومین بزرگترین واریانس بر روی دومین محور مختصات قرار می‌گیرد و همین طور برای بقیه.

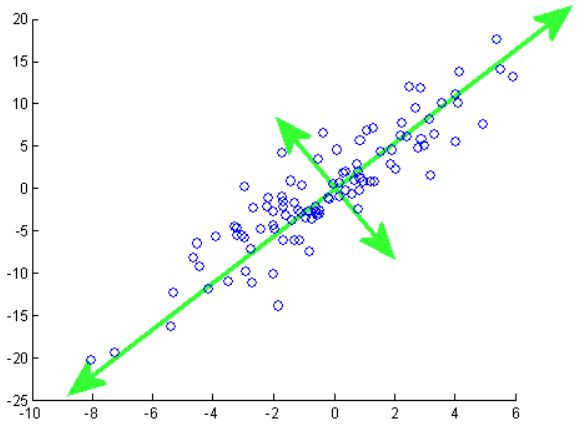
برای مجموعه داده‌های $\{x_1 \dots x_m\}$، محاسبه ماتریس کوواریانس نمونه‌ها از طریق رابطه زیر بدست می‌آید.

$$\sum = \frac{1}{m}\sum_{i=1}^{m}(x_i - \bar{x})(x_i - \bar{x})^T \qquad (1)$$

که $\bar{x}$ میانگین نمونه ها و m تعداد نمونه ها است. در حالی که:

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i \qquad (2)$$

در محاسبه مقدار ویژه و بردار ویژه ماتریس



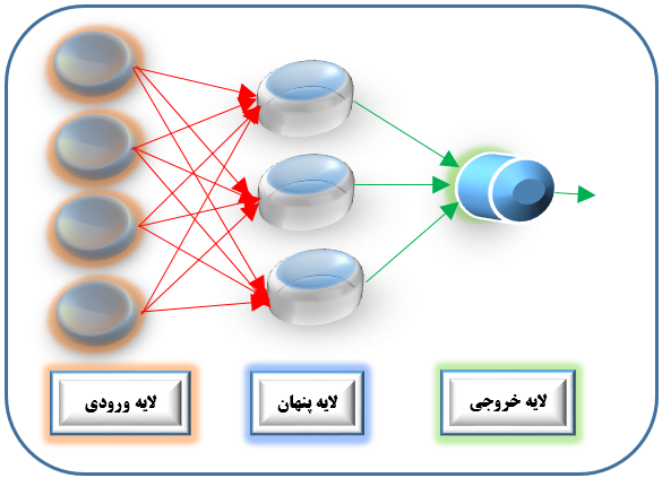
شکل ۲- تحلیل مؤلفه اساسی (محورهای جدید در جهت پرتراکم‌ترین نقاط قرار دارند) (۹)

کوواریانس داده‌ها، بیشترین مقدار ویژه بدست آمده دارای بالاترین اهمیت بردار ویژه در تحلیل مؤلفه‌های اساسی است.

با استفاده از تحلیل مؤلفه اساسی در این مقاله به استخراج متغیرهای مهم (به شکل مؤلفه) از مجموعه متغیرهای موجود در مجموعه داده می‌پردازیم تا به ثبت اطلاعات بیشتر با تعداد کمتری از متغیرها کمک نماییم. بدین شکل، بصری‌سازی داده‌ها نیز معنادارتر می‌شود. تحلیل مؤلفه اساسی هنگامی که با داده‌های دارای سه یا تعداد بیشتری بُعد سروکار داشته باشید، کاربردپذیرتر است (۱۱) (شکل۲). این مفهوم برای داده‌های دو بعدی برای این پژوهش نشان داده شده است.

پردازش داده و طبقه‌بندی: شبکه عصبی مصنوعی (ANN: Artificial Neural Network) (۱۲) الگویی برای پردازش اطلاعات می‌باشند که با تقلید از شبکه‌های عصبی بیولوژیکی مثل مغز انسان ساخته شده‌اند. عنصر کلیدی این الگو ساختار جدید سیستم پردازش اطلاعات آن می‌باشد و از تعداد زیادی عناصر (نرون) با ارتباطات قوی داخلی که هماهنگ با هم برای حل مسائل مخصوص کار می‌کنند تشکیل شده‌اند. یکی از شبکه‌های عصبی پرکاربرد شبکه عصبی پرسپترون چند لایه (MLP: Multi-Layer Perceptron) با روش یادگیری پس‌انتشار است که در صورت انتخاب صحیح ساختار داخلی، قادر است هر نوع سیستم غیر خطی را مدل‌بندی نماید. یک شبکهٔ پرسپترون چند لایه می‌تواند به سادگی با تعریف اوزان و توابع مناسب مورد استفاده قرارگیرد . توابع فعالیت مختلفی به فراخور اسلوب مسئله در نرون ها مورد استفاده قرار می‌گیرد (۱۳). شبکه عصبی مصنوعی مورد استفاده در این مقاله، از سه لایه‌ی ورودی، خروجی و پردازش تشکیل می‌شود. هر لایه شامل گروهی از سلول‌های عصبی (نورون) است که عموماً با کلیه‌ی نورون‌های لایه‌های دیگر در ارتباط هستند، مگر این که کاربر ارتباط بین نورون‌ها را محدود کند؛ ولی نورون‌های هر لایه با سایر نورون‌های همان لایه، ارتباطی ندارند (۱۴). قبل از به‌کار بردن مدل شبکه عصبی، وزن‌ها و اریب‌های ارتباط دهنده نرون‌های شبکه تعیین می‌شوند. به همین منظور تمام داده‌ها برای تدوین ساختار مدل به سه گروه تقسیم می‌شوند (۱۵). اولین گروه داده‌ها، به عنوان داده‌های آموزش برای تعیین وزن‌ها و اریب‌های شبکه به‌کار می‌روند. دومین گروه از داده‌ها، که داده‌های اعتباری نامیده می‌شوند، برای ارزیابی نتایج مرحله آموزش و تصمیم‌گیری درخصوص توقف آموزش شبکه استفاده می‌شوند. تعیین دقت مدل، و یا به عبارتی آزمون مدل، با استفاده از سومین گروه داده‌ها، یعنی داده‌های آزمون که در تدوین مدل استفاده نشده‌اند، انجام می‌شود. در شکل ۳ ساختار کلی شبکه عصبی مصنوعی مبتنی بر ساختار پرسپترون چند لایه نشان داده شده است.

الگوریتم ژنتیک (GA: Genetic Algorithm) یک روش برای حل مسائل بهینه‌سازی می‌باشد که اساس آن بر انتخاب، بقاء و تکامل در محیط‌های طبیعی استوار است. روش الگوریتم ژنتیک را می‌توان برای انواع



شکل ۳- ساختار کلی شبکه عصبی مصنوعی مبتنی بر ساختار پرسپترون چند لایه (۱۲)

مسائل بهینه‌یابی که مناسب برای روش‌های استاندارد بهینه‌سازی نیستند، استفاده نمود. همچنین این روش در مسائلی که در آنها تابع هدف ناپیوسته، غیرقابل مشتق‌گیری، آماری، بسیار غیرخطی و یا دارای مقادیر بهینه موضعی باشد، به‌کار برده می‌شود (۱۶). در این روش سعی بر مدل‌سازی ریاضی نظام انتخاب طبیعی است. در طبیعت عملگرهای ژنتیکی بسیاری دخیل می‌باشند اما در الگوریتم ژنتیک این عملگرها عبارتند از: به‌گزینی، تولید مثل (Reproduction)، تلاقی (Cross over) و پرش (Mutation). ابتدا داده‌های مورد استفاده پس از پیش‌پردازش، نرمال‌سازی و کاهش ابعاد داده، قبل از معرفی به شبکه عصبی مصنوعی توسط الگوریتم ژنتیک بهینه‌سازی گردید.

شاخص‌های ارزیابی: به طور کلی در سیستم‌های تشخیص بیماری‌ها برای بررسی میزان موفقیت و کارایی این سیستم‌ها از ماتریس آشفتگی (Confusion Matrix) استفاده می‌شود. در این مقاله، برای ارزیابی مدل‌ها از شاخص‌های دقت، حساسیت و صحت استفاده می‌شود که میزان دقت یک روش دسته‌بندی بر روی مجموعه داده‌های آزمون، درصد مشاهداتی از مجموعه آزمون است که به درستی توسط مدل مورد استفاده دسته‌بندی شده است. حساسیت عبارت است از میزانی برای مشخص‌کردن توانایی سیستم در تشخیص و دسته‌بندی موارد واقعاً بیمار که سیستم آنها را صحیح و سرطانی تشخیص می‌دهد. صحت عبارت است از میزانی برای مشخص‌کردن توانایی سیستم در تشخیص و دسته‌بندی موارد واقعاً سالم که سیستم آنها را به صورت صحیح تشخیص می‌دهد و توسط فرمول‌های ۳. ۴. و ۵. بیان می‌شوند.

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$\text{Sensitivity}=\frac{TP}{TP+FN} \quad (4)$$

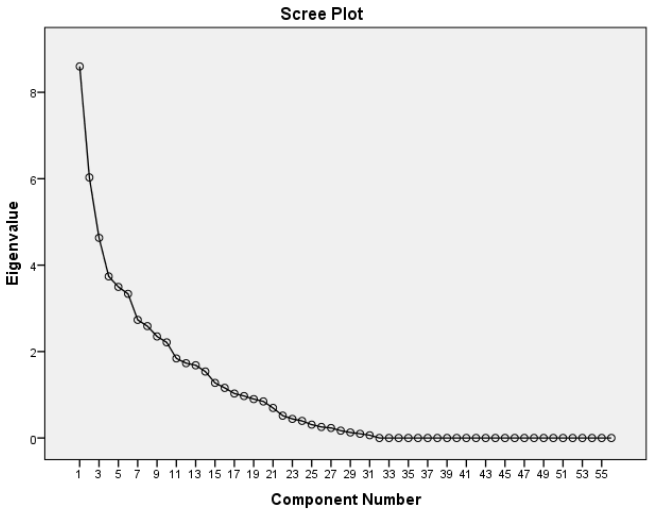
$$\text{Specificity}=\frac{TN}{TN+FP} \quad (5)$$

TP (True Positive)، TN (True Negative)، FP (False Positive)، FN (False Negative) به‌ترتیب تعداد مثبت‌های درست، منفی‌های درست، مثبت‌های نادرست و منفی‌های نادرست می‌باشد.

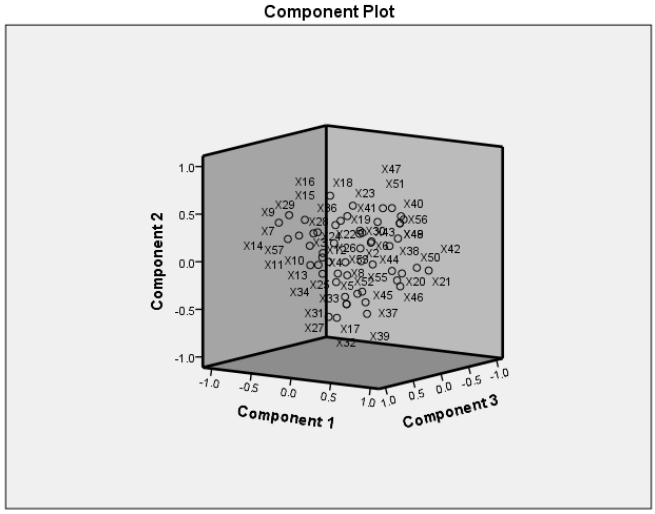
یافته‌ها

در این پژوهش ابتدا داده‌های مورد مطالعه پیش پردازش شد و سپس از الگوریتم PCA جهت کاهش ابعاد داده استفاده گردید. یافته‌های حاصل از تحلیل مولفه اصلی به صورت زیر است. شکل ۴ تغییرات مقادیر ویژه را در ارتباط با عامل‌ها نشان می‌دهد. این نمودار برای تعیین تعداد بهینه مؤلفه‌ها به‌کار می‌رود. با توجه به این نمودار مشاهده می‌شود که از عامل هفدهم به بعد تغییرات مقدار ویژه کم می‌شود، پس می‌توان هفده عامل را به‌عنوان عوامل مهم که بیشترین نقش را در تبیین واریانس داده‌ها دارند، استخراج کرد.

بعد از مشخص نمودن داده‌های ورودی و مطلوب توسط PCA که شامل ۱۷ مؤلفه اصلی بود، عمل چیدن



شکل ۴- نمودار اسکری‌گراف برای تعیین تعداد عامل



شکل ۵- نمودار سه بعدی پراکنش متغیرها نسبت به سه عامل اول استخراج شده

تصادفی (Randomization) ردیف‌ها روی صفحه گسترده انجام گردید. پارامترهای شبکه عصبی، شامل متغیرهای ورودی و تعداد نرون‌ها در لایه پنهان، با الگوریتم ژنتیک بهینه‌یابی شدند که در مرحله آزمون شبکه عصبی، نتایج برآورد شده توسط مدل، یا داده‌های خروجی بهترین برازش را با مؤلفه‌های اصلی داشته باشند. در این مرحله از ۷۰ درصد داده‌های پیش‌پردازش شده توسط PCA که شامل هفده مؤلفه اول بود، جهت آموزش شبکه عصبی مصنوعی با استفاده از الگوریتم ژنتیک استفاده شده است. در مرحله بعد ۳۰ درصد داده‌های پیش‌پردازش شده توسط الگوریتم PCA به‌صورت بردار به شبکه عصبی مصنوعی پیاده‌سازی شده در نرم‌افزار اعمال گردید. همچنین معماری‌های مختلف شبکه عصبی برای حصول بهترین کلاسبندی مورد بررسی قرار گرفت و دقت، حساسیت و صحت به دست آمده با استفاده از بهینه‌سازی الگوریتم ژنتیک برای بهترین معماری به‌ترتیب برابر ۹۸/۶۵، ۹۸، ۹۹/۱۶ درصد بدست آمد. همان‌طور که در جدول ۱. مشاهده می‌کنید انتخاب دقیق بهترین معماری نقش عمده‌ای در افزایش دقت، حساسیت و صحت تشخیص شبکه دارد.

بحث و نتیجه‌گیری

در این مطالعه سعی شده است برای بهبود تشخیص سرطان ریه از روش جدید بر پایه تحلیل مؤلفه اساسی جهت استخراج ویژگی و کاهش ابعاد داده با حفظ

جدول ۱- نتایج معماری‌های مختلف شبکه عصبی پرسپترون چندلایه

صحت	حساسیت	دقت	نوع معماری
۹۰/۶۴	۸۷/۱۹	۸۸/۲۷	۱۷-۸-۳-۳
۹۳/۵۷	۹۰/۲۵	۹۲/۱۴	۱۷-۱۰-۳-۳
۹۹/۱۶	۹۸	۹۸/۶۵	۱۷-۱۲-۳-۳
۹۴/۸۴	۹۱/۶۵	۹۳/۳۵	۱۷-۱۵-۳-۳
۸۶/۶۶	۸۴/۷۸	۸۵/۶۸	۱۷-۱۸-۳-۳

بیشترین اطلاعات و پرسپترون چندلایه با بهینه‌سازی الگوریتم ژنتیک برای تشخیص بیماری سرطان ریه استفاده شود. در این بخش کارایی روش پیشنهادی جدید با تعدادی از پژوهش‌های صورت گرفته تا امروز مقایسه می‌شود. اوسی در سال ۲۰۱۲ با استفاده سیستم تشخیص تخصصی GDA—LS_SVM ابتدا به کاهش ابعاد و استخراج ویژگی‌های سرطان ریه و سپس به طبقه‌بندی آنها پرداخته و میزان دقت به دست آمده با روش فوق را ۹۶/۸۷۵٪ گزارش کرده است که نسبت به روش پیشنهادی با توجه به معیار دقت، عملکرد پایین‌تری دارد (۲۲). لو و همکارانش در سال ۲۰۱۴ یک روش هوشمند برای تشخیص سرطان ریه با استفاده از الگوریتم ژنتیک بر پایه روش انتخاب ویژگی ارائه داده میزان دقت تشخیص به دست آمده با روش فوق را ۹۹٪ گزارش کرده‌اند که نسبت به روش پیشنهادی علی‌رغم دقت خوب بدست آمده به علت زمان‌بر بودن و حجم بالای دیتای ورودی از قابلیت اجرایی شدن کمتری برخوردار است (۲۳). همچنین این نویسنده و همکارانش در سال ۲۰۰۱ با بهره‌گیری از روش محلول فیلتر FCBF به دقت ۸۷/۵٪ رسیده بودند که نسبت به روش پیشنهادی با توجه به معیار دقت، عملکرد ضعیف‌تری به دست آورده بودند (۱۷). تان و همکارانش در سال ۲۰۰۴ با استنباط MML از روش درخت تصمیم‌گیری تنها توانسته‌اند به دقت ۴۶/۸٪ دست یابند که نسبت به روش پیشنهادی عملکرد ضعیف‌تری دارد (۱۸). در مقایسه با نتایج به‌دست‌آمده در مطالعات پیشین، در این پژوهش که روشی متفاوت با مقالات ذکرشده بود طراحی و ارزیابی صورت گرفت که نتیجه آن رسیدن به دقت، حساسیت و صحت به‌ترتیب برابر ۹۸/۶۵، ۹۸، ۹۹/۱۶ درصد بوده است. آن‌چه که می‌تواند پژوهش حاضر را از مطالعات پیشین متمایز کند علاوه بر دقت بالا، بهره‌گیری از الگوریتم فراابتکاری و نیز سرعت بالای آن در تشخیص سرطان ریه است که با استخراج ویژگی و کاهش ابعاد داده، پیچیدگی‌های مسئله کاهش یافته است. با این‌حال برخی محدودیت‌ها از جمله عدم استفاده از دیتای بومی به علت دردسترس نبودن اطلاعات بیماران و تعداد محدود مراکز تخصصی درمان بر روی نتایج این پژوهش موثر بوده است. در جدول۲. نتایج به دست آمده با روش‌های دیگر مقایسه شده‌اند که روش پیشنهادی ما جزء بالاترین دقت‌ها می‌باشد.

با توجه به نتایج این مقاله، استفاده از الگوریتم تحلیل

جدول ۲- مقایسه دقت داده‌ها در پیاده‌سازی‌های مختلف در تشخیص سرطان ریه

ردیف	روش	دقت	سال	نویسندگان و منبع
۱	FCBF	۸۷/۵	۲۰۰۱	لو و همکاران (۱۷)
۲	MML oblique tree	۴۶/۸	۲۰۰۴	تان و همکاران (۱۸)
۳	RS/RSP	۷۰/۰۱	۲۰۰۵	بوستروم و همکاران (۱۹)
۴	MAXENT-H	۴۳/۳	۲۰۰۵	هندریک (۲۰)
۵	k-NN + Reduction	۷۵	۲۰۰۵	بادیو و همکاران (۲۱)
۶	GDA-LS_SVM	۹۶/۸۷	۲۰۱۱	اوسی (۲۲)
۷	GA-FELM	۹۸/۸۵	۲۰۱۲	دلیری (۶)
۸	GA based feature selection method	۹۹	۲۰۱۴	لو و همکاران (۲۳)
۹	PCA-GA_MLP	۹۸/۶۵	۲۰۱۹	پژوهش حاضر

مؤلفه اساسی جهت استخراج ویژگی و کاهش ابعاد داده از ۵۷ ویژگی به ۱۷ مولفه اساسی و استفاده از الگوریتم ژنتیک جهت بهینه‌سازی ساختار پرسپترون چند لایه باعث افزایش دقت تشخیص بیماری سرطان ریه شد. بهره‌گیری از روش‌های کامپیوتری و داده‌کاوی می‌تواند کمک شایانی برای پزشکان در تشخیص بیماری‌ها داشته باشد و پزشکان می‌توانند با این روش‌ها، متغییرهای بیشتری را در مدت زمان کمتری در نظر داشته باشند. این روش جدید غیرتهاجمی می‌تواند به عنوان یک روش موفق و دستیار پزشک جهت شناسایی و کاهش هزینه‌های درمان بیماری سرطان ریه مورد استفاده قرار گیرد.

تقدیر و تشکر

این مطالعه با حمایت مالی دانشگاه لرستان به انجام رسیده است لذا نویسندگان بر خود لازم می‌دانند تا از مدیریت محترم دانشگاه جهت حمایت‌های مالی تشکر و قدردانی نمایند.

References

1. Polat K, Güneş S. Principles component analysis, fuzzy weighting pre-processing and artificial immune recognition system based diagnostic system for diagnosis of lung cancer. Expert Sys Appl. 2008;34(1): 214-221.

2. Liu C, Wen Z, Li Y, Peng L. Application of ThinPrep Bronchial Brushing Cytology in the Early Diagnosis of Lung Cancer: A Retrospective Study. Plos One. 2014;9(4): e90163.

3. Hosseini M, Naghan PA, Karimi S, SeyedAlinaghi S, Bahadori M, Khodadad K, et al. Environmental risk factors for lung cancer in Iran: a case- control study. Int J Epidemiol. 2009;38(4): 989-96.

4. Ghanbari N, Golestan Jahromi F, Bani Hashem K, Moazen M. The Effect of Stress Inoculation Training (SIT) on Depression and Anxiety in Patients with prostate Cancer. J Sabzevar Uni Med Sci. 2014;21(5):735-42. (Persian)

5. Abazari M, Abazari R, Roshanaei G, Gholamnejad M, Mahjub H. Evaluation of Prognostic Variables for Classifying the Survival In lung cancer Patients using The Decision Tree. 2015;22(4):549-66. (Persian)

6. Daliri MR. A hybrid automatic system for the diagnosis of lung cancer based on genetic algorithm and fuzzy extreme learning machines. J Med Syst. 2012;36(2):1001-1005.

7. ftp://ftp.ics.uci.edu/pub/machine-learning-databases (last accessed: 05 March 2006).

8. Dodge Y, the Oxford Dictionary of Statistical Terms, OUP; 2003.

9. Dunteman GH. Principal components analysis. 1989 (No. 69). Sage.

10. Jolliffe IT. Mathematical and statistical properties of population principal components. Principal component analysis. 2002;10-28.

11. Bryant F B, Yarnold PR. Principal components analysis and exploratory and confirmatory factor analysis. 1995.

12. Yegnanarayana B. Artificial neural networks. PHI Learning. 2009 Pvt. Ltd.

13. Anderson A. An introduction to neural network, Cambridge, MA: MIT press. 1995;795-851.

14. Jain AK, Mao J, Mohiuddin KM. Artificial neural networks: A tutorial. Computer. 1996;29(3): 31-44.

15. Coulibaly P, Anctil F Bobee B. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. J Hydrol. 2000;230:214-257.

16. Goldberg DE. Genetic Algorithm in Search Optimization and Machine Learning. Addison Wesley, Reading, Massachusetts, USA, 1989.

17. Yu L, Liu H. Feature selection for high-dimensional data: A fast correlation-based filter solution. In Proceedings of the 20th international conference on machine learning (ICML-03) 2003;856-863.

18. Tan PJ, Dowe DL. MML Inference of Oblique Decision Trees in Lecture Notes in Artificial Intelligence, GI Webb and X. Yu, Eds., 17th Australian Joint Conf. on Advances in AI, Cairns, Australia, 2004:0302-9743.

19. Bostrom H. Maximizing the area under the ROC curve using incremental rediced error pruning. In: Proceedings of the ICML 2005 Workshop on ROC Analysis in Machine Learning, 2005.

20. Hendrickx I, van den Bosch A. Hybrid algorithms with instance based classification. In: Proceedings of the 16th European Conf. on Machine Learning. Lect Notes Comput Sci. 2005;3720: 158–169.

21. Badjio FE, Poulet F. Dimension reduction for visual data mining. In Proceedings of International symposium on applied stochastic models and data analysis, 2005.

22. Avci E. A new expert system for diagnosis of lung cancer: GDA—LS_SVM. J Med Syst. 2012;36(3): 2005-2009.

23. Lu C, Zhu Z, Gu X. An intelligent system for lung cancer diagnosis using a new genetic algorithm based feature selection method. J Med Syst. 2014;38(9): 97.