Archive of SID

ORIGINAL ARTICLE

Iranian Quarterly Journal of Breast Disease 2017; 10(3):55.

Designing a Group Decision-Making System Using a Fuzzy Combination of Regression Methods for Prediction of Benign or Malignant Breast Tumors

Khosravanian A: Department of Electrical and Computer Engineering, Semnan University, Semnan, Iran
Rahmanimanesh M: Department of Electrical and Computer Engineering, Semnan University, Semnan, Iran
Keshavarzi P: Department of Electrical and Computer Engineering, Semnan University, Semnan, Iran

Corresponding Author: Mohammad Rahmanimanesh, rahmanimanesh@semnan.ac.ir

Abstract

Introduction: In cancer research, early detection of breast cancer is effective in prognosis and increasing the survival of patients. Artificial intelligence and data mining methods are predictive methods that can be useful in this case. The purpose of this paper is to design and evaluate an automatic computer-aided diagnosis system to predict the type of breast tumor by using a fuzzy combination of regression methods.

Methods: In this descriptive-analytic study, a group decision-making system was designed by using the combination of PCA (Principal Components Analysis), RSM (Response Surface Methodology) and SVR-Firefly (Support Vector Regression- Firefly) methods to predict the type of breast tumor (benign or malignant). To evaluate the designed system, we used the Wisconsin Breast Cancer Dataset stored in the UCI (University of California, Irvine) Machine Learning Repository. After pre-processing, the data was split into training and testing datasets, which were evaluated by each of mentioned regression methods. Finally the results of the regression methods were combined by fuzzy voting and fuzzy averaging approaches. The proposed group decision-making system was simulated by MATLAB software.

Results: This study was conducted on 683 women with breast cancer, including 444 cases with benign tumors, as well as 239 cases with malignant tumors. For each case, nine clinical variables were used as inputs. The performance of the proposed group decision-making system in test phase is based on accuracy, precision, sensitivity and specificity. The values of these indicators were obtained, on average, 0.9832, 0.9588, 0.9900, 0.9832 in fuzzy voting method and 0.9820, 0.9524, 0.9929, 0.9804 in fuzzy averaging method while the performance indicators were obtained 1 by both methods in the best case.

Conclusion: The results of experiments show the effectiveness of the proposed group decision-making system in predicting the type of breast tumors (benign or malignant). It can help the physicians to choose the best treatment method and prevent the cancer progression. The developed software can also be used for training physicians.

Keywords: Breast Cancer, Fuzzy Voting, Fuzzy Averaging, PCA, RSM, SVR-Firefly, Wisconsin Breast Cancer Dataset.

www.SID.ir

Archive of SID

مقاله پژوهشی

فصلنامهٔ بیماری‌های پستان ایران، سال دهم، شماره سوم، پاییز ۱۳۹۶؛ (۶۶ -۵۶)

تاریخ ارسال: ۹۶/۲/۴ | تاریخ پذیرش: ۹۶/۶/۲۵

طراحی یک سیستم تصمیم‌گیری گروهی با ترکیب فازی روش‌های رگرسیون در پیش‌بینی خوش‌خیم یا بدخیم بودن تومورهای پستان

آسیه خسروانیان: گروه علمی مهندسی کامپیوتر، دانشگاه سمنان، سمنان، ایران

محمد رحمانی‌منش*: گروه علمی مهندسی کامپیوتر، دانشگاه سمنان، سمنان، ایران

پرویز کشاورزی: گروه علمی مهندسی کامپیوتر، دانشگاه سمنان، سمنان، ایران

چکیده

مقدمه: در تحقیقات سرطان، تشخیص زودهنگام سرطان پستان بر پیش‌آگهی و طول عمر بیمار تأثیرگذار است. روش‌های هوش مصنوعی و داده‌کاوی ازجمله روش‌های پیش‌بینی هستند که در این مورد می‌توانند کمک‌کننده باشد. هدف از این مقاله طراحی و ارزیابی یک سیستم جدید تشخیص خودکار کامپیوتری جهت پیش‌بینی نوع توده پستان با استفاده از ترکیب فازی روش‌های رگرسیون است.

روش بررسی: در این مطالعه توصیفی- تحلیلی یک سیستم تصمیم‌گیری گروهی با ترکیب روش‌های رگرسیون PCA، RSM و SVR-Firefly به‌منظور پیش‌بینی نوع توده پستان اعم از تومور خوش‌خیم یا بدخیم طراحی‌شده است. به‌منظور ارزیابی سیستم طراحی‌شده از مجموعه داده مربوط به بیماران مبتلا به سرطان پستان بیمارستان ویسکانسین موجود در انبار داده یادگیری ماشین استفاده شده است. داده‌های ورودی پس از پیش‌پردازش به‌صورت تصادفی به دو دسته داده آموزش و آزمون تقسیم شد. سپس با هر یک از روش‌های رگرسیون ذکرشده مورد آموزش و آزمون قرار گرفت. درنهایت خروجی این روش‌ها با رویکردهای رأی‌گیری فازی و میانگین‌گیری فازی ترکیب شد. سیستم تصمیم‌گیری گروهی پیشنهادی با نرم‌افزار MATLAB شبیه‌سازی شده است.

یافته‌ها: این مطالعه بر ۶۸۳ مورد زن مبتلا به سرطان پستان انجام شد که از این تعداد ۴۴۴ نمونه دارای تومور خوش‌خیم و ۲۳۹ نمونه دارای تومور بدخیم می‌باشد. برای هر مورد از ۹ متغیر بالینی به‌عنوان ورودی استفاده شد. عملکرد سیستم تصمیم‌گیری گروهی طراحی‌شده بر اساس شاخص‌های صحت، دقت، حساسیت و اختصاصیت در مرحله آزمون در حالت متوسط با روش رأی‌گیری فازی معادل اعداد ۰/۹۸۳۲، ۰/۹۵۸۸، ۰/۹۹۰۰ و ۰/۹۸۳۲ و در روش میانگین‌گیری فازی معادل اعداد ۰/۹۸۲۰، ۰/۹۵۲۴، ۰/۹۹۲۹ و ۰/۹۸۰۴ به دست آمد. درحالی‌که این شاخص‌ها در بهترین حالت با هر دو روش معادل عدد یک به دست آمد.

نتیجه‌گیری: بالا بودن شاخص‌های عملکردی سیستم تصمیم‌گیری گروهی در این مقاله نشان داد که سیستم پیشنهادی عملکرد مناسبی در پیش‌بینی خوش‌خیم یا بدخیم بودن توده‌های پستان دارد. تعیین دقیق نوع توده می‌تواند در انتخاب روش درمانی مناسب به پزشک کمک کند و از پیشروی این سرطان جلوگیری کند. همچنین نرم‌افزار طراحی‌شده بر اساس شبیه‌سازی این مقاله می‌تواند در آموزش پزشکان به کار رود.

واژه‌های کلیدی: سرطان پستان، رأی‌گیری فازی، میانگین‌گیری فازی، PCA، RSM، SVR-Firefly، مجموعه داده ویسکانسین.

* نشانی نویسنده مسئول: : سمنان، دانشگاه سمنان، دانشکده مهندسی برق و کامپیوتر، گروه علمی مهندسی کامپیوتر، محمد رحمانی‌منش.

نشانی الکترونیک: rahmanimanesh@semnan.ac.ir

www.SID.ir

Archive of SID

سرطان نوعی بیماری ژنتیکی است که در اثر رشد و تکثیر غیرقابل‌کنترل سلول در بدن به وجود می‌آید (۱). سلول‌های تکثیرشده در نهایت تشکیل یک توده به نام تومور می‌دهند که به دو نوع خوش‌خیم و بدخیم تقسیم می‌شود. سرطان پستان با ایجاد تومور در بافت پستان به وجود می‌آید. امکان رخ دادن این سرطان در هر دو جنس (مذکر، مؤنث) وجود دارد، اضافه شود شیوع آن در مردان بسیار کمتر از زنان است.

به گزارش انجمن سرطان آمریکا[1] در سال ۲۰۱۵ حدود ۲۳۱۸۴۰ مورد جدید سرطان پستان در میان زنان ایالت متحده تشخیص داده شده است که تعداد ۴۰۲۹۰ نفر منجر به مرگ شده است. همچنین ۲۳۵۰ مرد مبتلابه این نوع سرطان تشخیص داده شد که ۴۴۰ مورد آن منجر به مرگ شده است (۲). بر اساس آمار ذکرشده، در این گزارش آمده است که سرطان پستان بعد از سرطان ریه بالاترین آمار مرگ‌ومیر را به خود اختصاص داده است. از آنجا که تشخیص به‌موقع این سرطان و درمان‌های اولیه شانس زنده ماندن بیمار را افزایش می‌دهد نیاز به ابزاری که در تشخیص زودهنگام به پزشک کمک کند احساس می‌شود. ریسک فاکتورهای ابتلا به سرطان پستان در زنان شامل سن بالا، جهش‌های ژنتیکی، سابقه خانوادگی، مصرف مشروبات الکلی و دخانیات، زمان اولین قاعدگی قبل از ۱۲ سالگی و شروع یائسگی بعد از ۶۶ سالگی، بارداری پس از ۳۰ سالگی یا عدم بارداری، اضافه وزن و چاقی بعد از یائسگی و زندگی بی‌تحرک است (۳).

مطالعات نشان می‌دهد روش‌های هوش مصنوعی و داده‌کاوی در تشخیص انواع بیماری‌ها ازجمله سرطان به کمک پزشک آمده است. به‌طور خاص در تحقیقات سرطان پستان، این روش‌ها در تشخیص زودهنگام سرطان پستان، تشخیص نوع توده پستان، پیش‌بینی بقای افراد مبتلا به سرطان پستان و پیش‌بینی عود مجدد این سرطان به‌کاررفته است. به‌عنوان نمونه محمودیان و همکاران از ترکیب روش SVR و سیستم فازی نوع ۲ برای پیش‌بینی زمان عود سرطان پستان استفاده کردند (۴). در مطالعه دیگری با استفاده از ترکیب رویکرد تکاملی و روش Boosting یک سیستم خودکار جهت درجه‌بندی بدخیمی سرطان پستان طراحی‌شده است (۵). لاندین و همکارانش از مدل‌های شبکه عصبی مصنوعی و رگرسیون لجستیک برای پیش‌بینی بقای بیماران مبتلا به سرطان پستان استفاده کرده‌اند (۶). رحمانی سریاست و همکاران، یک سیستم کامپیوتری خودکار جهت تشخیص نوع سرطان پستان طراحی و بر روی پایگاه داده‌های MIAS و DDSM سیستم خود را ارزیابی کردند (۷). درزی و همکاران برای طبقه‌بندی داده‌های ۹۱۸ نفر مبتلا به سرطان پستان سه الگوریتم AdaBoost.M1، KNN و شبکه عصبی احتمالی را به‌کار برده‌اند (۸). شیخ‌پور و همکاران برای افزایش کارایی سیستم‌های تشخیص سرطان پستان روش کاهش ویژگی دو مرحله‌ای را پیشنهاد کردند و عملکرد روش‌های درخت تصمیم J48، بیزین ساده، طبقه‌بندی کننده درجه دوم، ماشین بردار پشتیبان و روش KNN را مورد بررسی قراردادند (۹). کاروالو و همکاران یک مدل ترکیبی بر اساس شبکه بیزی و MCDA[2] برای تشخیص زودهنگام سرطان پستان پیشنهاد کردند. در این مقاله زمانی که شبکه بیزی قادر به پاسخگویی نیست، از تصمیم‌گیری چند معیاره استفاده شده است. نویسندگان بیان کردند ابزار طراحی‌شده می‌تواند بر مجموعه داده با هزاران مورد در زمان کوتاهی (چند دقیقه) اجرا شود و نوع سرطان پستان را تشخیص دهد (۱۰).

به نقل از (۱۱) امروزه سیستم‌های تشخیص کامپیوتری (CAD[3]) به‌منظور تشخیص خودکار نوع توده‌های پستان موردتوجه بیشتری قرارگرفته است. این سیستم‌ها می‌توانند در شرایط خستگی، اضطراب و عدم تجربه کافی به پزشک کمک کنند.

هدف از این مقاله طراحی و ارزیابی یک سیستم تصمیم‌گیری گروهی با ترکیب فازی روش‌های رگرسیون PCA[4]، RSM[5] و SVR- Firefly[6] در پیش‌بینی خوش‌خیم یا بدخیم بودن نوع توده پستان است تا بتواند با پیش‌بینی خودکار در انتخاب روش درمانی مناسب به پزشک و پیش‌آگهی به بیماران کمک کند.

مواد و روش‌ها

[1] American Cancer Society

[2] Multi Criteria Decision Analysis
[3] Computer Aided Diagnosis
[4] Principal Components Analysis
[5] Response Surface Methodology
[6] Support Vector Regression -Firefly

www.SID.ir

Archive of SID

در این مقاله از مجموعه داده مربوط به بیماران مبتلا به سرطان پستان بیمارستان ویسکانسین (WBCD[7])، موجود در انبار داده یادگیری ماشین دانشگاه ایروین کالیفرنیا (UCI[8]) استفاده‌شده است (۱۲). این مجموعه داده شامل ۶۹۹ مورد زن مبتلا به سرطان پستان است که دارای ۹ ویژگی (مطابق با جدول ۱) و دو کلاس خوش‌خیم و بدخیم مربوط به هر رکورد است.

جدول ۱: متغیرهای ورودی در دسته‌بندی نوع توده پستان

متغیرهای بالینی	محدوده	میانگین	انحراف معیار
ضخامت انبوه	۱-۱۰	۴/۴۴	۲/۸۳
یکنواختی اندازه سلول	۱-۱۰	۳/۱۵	۳/۰۷
یکنواختی شکل سلول	۱-۱۰	۳/۲۲	۲/۹۹
چسبندگی لبه‌ها	۱-۱۰	۲/۸۳	۲/۸۶
حجم سلول بافت اپیتلیال	۱-۱۰	۲/۲۳	۲/۲۲
هسته‌های عریان	۱-۱۰	۳/۵۴	۳/۶۴
کروماتین بلاند	۱-۱۰	۳/۴۵	۲/۴۵
هسته عادی	۱-۱۰	۲/۸۷	۳/۰۵
تقسیم هسته سلول به دو قسمت	۱-۱۰	۱/۶۰	۱/۷۳

محدوده تمامی متغیرهای بالینی ۱-۱۰ در نظر گرفته‌شده که عدد ۱ خوش‌خیم‌ترین و عدد ۱۰ بدخیم‌ترین وضعیت را نشان می‌دهد. به‌منظور استفاده از این مجموعه داده در شبیه‌سازی‌های این مقاله پس از بررسی داده‌ها در مرحله پیش‌پردازش ازآنجاکه شماره شناسایی بیماران جزء فاکتورهای تشخیصی این بیماری نبود ستون مربوطه در مجموعه داده‌ها حذف شد. علاوه بر آن با بررسی رکوردهای این مجموعه داده مشاهده شد که مقادیر برخی از ویژگی‌ها در دسترس نیست و به‌اصطلاح داده گم‌شده است. به‌منظور کم کردن خطا در روند شبیه‌سازی، رکوردهای حاوی داده‌های گم‌شده که ۱۶ مورد بود حذف شد و شبیه‌سازی با ۶۸۳ داده باقیمانده انجام شد که از این تعداد ۴۴۴ نمونه مربوط به کلاس خوش‌خیم و ۲۳۹ نمونه دیگر مربوط به کلاس بدخیم بود.

نرمال‌سازی داده‌ها: نرمال‌سازی داده‌ها یکی از مراحل پیش‌پردازش داده‌های واقعی به‌منظور بهبود کیفیت آنهاست. نرمال‌سازی به‌ویژه برای الگوریتم‌های دسته‌بندی مفید است. در پژوهش قبلی نویسندگان این مقاله (۱۳) نشان داده شد که نرمال کردن داده‌های موجود در مجموعه داده ویسکانسین موجب کاهش خطای دسته‌بندی‌شده است. این امر به‌این‌علت است که تبدیل داده‌ها به بازه [۰،۱] از اینکه داده‌های با مقیاس بزرگ نتیجه را به سمت خویش منحرف کنند جلوگیری می‌کند لذا در این مقاله برای نرمال‌سازی داده‌های آموزش و آزمون از روش مشهور Min-Max استفاده شد و مقادیر تمام ویژگی‌ها به بازه [۰،۱] نگاشت شد. فرض کنید مجموعه داده به‌صورت $D = [X_1, X_2, \dots, X_n]^T$ تعریف شود که n تعداد رکوردها، d تعداد ویژگی‌ها و هر رکورد به‌صورت $X_i = (X_{i1}, X_{i2}, \dots, X_{id})$ باشد. همچنین فرض کنید $\check{X} = (\check{X}_1, \check{X}_2, \dots, \check{X}_d)$ مینیمم D، $\hat{X}$ ماکزیمم D و $\Delta = \hat{X} - \check{X} = (\delta_1, \delta_2, \dots, \delta_d)$ باشد.

برای نرمال کردن، D به مجموعه داده جدید $D' = [X'_1, X'_2, \dots, X'_n]^T$ با $X'_i = (X'_{i1}, X'_{i2}, \dots, X'_{id})$ تبدیل می‌شود، به صورتی که:

$$X'_{ij} = \frac{x_{ij} - \check{X}_j}{\delta_j} \qquad i = 1, \dots, n \quad , j = 1, \dots, d. \tag{1}$$

روش‌های رگرسیون

روش تحلیل مؤلفه‌های اصلی (PCA: Principal Components Analysis): روش تحلیل مؤلفه‌های اصلی (PCA) (۱۴) یکی از روش‌های تحلیل داده‌های چند متغیره است. ایده اصلی این روش کاهش ابعاد یک مجموعه داده شامل تعداد زیادی متغیر است، به‌نحوی‌که تنوع داده‌های مجموعه حفظ شود. این روش از نظر ریاضی یک تبدیل خطی متعامد است که داده‌ها را به دستگاه مختصات جدید منتقل می‌کند. این انتقال به صورتی انجام می‌شود که اولین محور مختصات بیشترین واریانس داده‌ها را در برمی‌گیرد و محورهای بعدی نیز به‌نوبه خود بیشترین واریانس ممکن داده را در برمی‌گیرند. یکی از کاربردهای مهم این روش در رگرسیون است (۱۴). روش PCA در پیش‌بینی سرطان (۱۵-۱۷) به کار گرفته‌شده است.

[7] Wisconsin Breast Cancer Database
[8] UC Irvine Machine Learning Repository

www.SID.ir

Archive of SID

روش متدولوژی سطح پاسخ (RSM: Response Surface Methodology): متدولوژی سطح پاسخ (RSM) (۱۸) مجموعه‌ای از روش‌های آماری و ریاضی است که در فرایندهای بهینه‌سازی کاربرد دارد. در این روش چگونگی روابط بین یک یا چند پاسخ با استفاده از تأثیر عوامل مستقل اندازه‌گیری می‌شود (۱۹). در حالت کلی مدل RSM با فرمول زیر نمایش داده می‌شود:

$$y = f(x_1, x_2, \dots, x_d) + \varepsilon \qquad (۲)$$

که $x_1, x_2, \dots, x_d$ متغیرهای ورودی هستند و ε میزان خطای مدل تعریف می‌شود. در این مقاله از مدل درجه‌دو برای تابع f با فرمول زیر استفاده می‌شود.

$$y = \beta_0 + \sum_{i=1}^{d} \beta_i x_i + \sum_{i=1}^{d} \beta_{ii} x_i^2 + \sum_{i<j} \sum \beta_{ij} x_i x_j + \varepsilon \qquad (۳)$$

$$f(x) = < w, x > + b \quad w \in X \ , b \in R \qquad (۴)$$

که β_i ها ضرایب ثابت هستند (۱۸).

روش Support Vector Regression- Firefly:
مدل رگرسیون بردار پشتیبان (SVR) (۲۰) ازجمله روش‌های یادگیری ماشین است که در مسائل پیش‌بینی کاربرد دارد. فرض کنید مجموعه آموزشی $\{(x_1, y_1), \dots, (x_n, y_n)\} \subset X \times R$ داده‌شده است که $X \in R^d$ نشان‌دهنده فضای الگوهای ورودی است. هدف پیدا کردن تابع f(x) برای تخمین تابع رگرسیون است که به‌صورت زیر تعریف می‌شود:
در فرمول ۴، <.,.> ضرب داخلی تعریف می‌شود. تابع بهینه رگرسیون به‌صورت زیر تعریف می‌شود.

$$\text{Minimize} \quad \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*) \qquad (۵)$$

$$\text{Subject to} \begin{cases} y_i - < w, x_i > - b \le \varepsilon + \xi_i \\ < w, x_i > + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases}$$

که C یک ثابت از پیش تعیین‌شده است و ξ_i و ξ_i^* متغیرهایی هستند که حد بالا و پایین خروجی سیستم را تعیین می‌کنند (۲۱).

برای مدل رگرسیون بردار پشتیبان از کرنل‌های مختلفی استفاده می‌شود که معمولا تابع کرنل گوسی[9] (RBF) برای پیش‌بینی عملکرد بهتری دارد (۲۱). کارایی روش SVR کاملا به مقدار پارامترهای مدل سازنده آن بستگی دارد. در این مقاله برای تنظیم این پارامترها، از الگوریتم فرا ابتکاری کرم شب‌تاب[10] که ازجمله الگوریتم‌های فرا ابتکاری الهام گرفته از طبیعت است و در مسایل بهینه‌سازی کاربرد دارد، استفاده‌شده است. این الگوریتم در مسایل رگرسیون هم به‌طور گسترده مورداستفاده قرارگرفته است (۲۲، ۲۳). روش حاصل که ترکیب SVR و الگوریتم کرم شب‌تاب است SVR-Firefly نامیده می‌شود.

ماتریس کانفیوژن: به‌طورکلی برای بررسی میزان موفقیت و کارایی سیستم‌های یادگیری، پیش‌بینی و تشخیص از ماتریس کانفیوژن استفاده می‌شود. تحلیل‌های ماتریس کانفیوژن در دسته‌بندی و تشخیص بیماران منجر به چهار حالت مثبت حقیقی[11] (TP)، منفی حقیقی[12] (TN)، مثبت کاذب[13] (FP) و منفی کاذب[14] (FN) می‌شود. از نتایج ماتریس کانفیوژن چهار شاخص صحت[15]، دقت[16]، حساسیت[17] و اختصاصیت[18] به‌صورت زیر به دست می‌آید:

$$\text{صحت} = \frac{TP+TN}{TP+TN+FP+FN} \qquad (۶)$$

$$\text{دقت} = \frac{TP}{TP+FP} \qquad (۷)$$

$$\text{حساسیت} = \frac{TP}{TP+FN} \qquad (۸)$$

$$\text{اختصاصیت} = \frac{TN}{FP+TN} \qquad (۹)$$

طراحی سیستم تصمیم‌گیری گروهی پیشنهادی: به‌منظور طراحی سیستم پیشنهادی، ابتدا هر یک از روش‌های رگرسیون PCA، RSM و SVR-Firefly بر

[9] Gaussian Radial Basis Kernel Function (RBF)
[10] Firefly algorithm
[11] True Positive
[12] True Negative
[13] False Positive
[14] False Negative
[15] Accuracy
[16] precision
[17] Sensitivity
[18] Specificity

www.SID.ir

Archive of SID

مجموعه داده ویسکانسین پیاده‌سازی شده است. داده‌های ورودی شامل یک ماتریس با ۶۸۳ سطر (تعداد رکوردهای موجود) و ۹ ستون (تعداد متغیرهای بالینی) است. ماتریس دیگری به‌عنوان ماتریس هدف در نظر گرفته شد که شامل ۶۸۳ سطر و یک ستون دارای دو مقدار صفر و یک است درصورتی‌که توده پستان از نوع خوش‌خیم باشد مقدار متناظر صفر و درصورتی‌که نوع توده بدخیم باشد مقدار متناظر یک در نظر گرفته شد. برای پیاده‌سازی روش‌های رگرسیون ذکرشده، داده‌های مرحله آموزش و مرحله آزمون در هر بار شبیه‌سازی به‌صورت تصادفی از کل داده‌های موجود، انتخاب شد. در مجموع با استفاده از روش cross-validation شش دسته داده شامل داده‌های آموزش و آزمون متفاوت تولید شد و هر یک از روش‌های رگرسیون شش مرتبه (یک مرتبه به ازای هر دسته داده) شبیه‌سازی گردید.

در طراحی سیستم پیشنهادی، هرکدام از روش‌های رگرسیون به‌عنوان یک متخصص در نظر گرفته می‌شوند که در پیش‌بینی نوع توده پستان اظهارنظر کنند. ازآنجایی‌که این روش‌ها هرکدام نقاط ضعف و قوت خود را دارند و در برخی موارد ممکن است قادر به پیش‌بینی صحیح نباشند، یک سیستم تصمیم‌گیری گروهی[19] با ترکیب خروجی روش‌های رگرسیون با دو روش رأی‌گیری فازی و میانگین‌گیری فازی طراحی شد تا نقاط قوت روش‌ها را حفظ و تقویت کند و نقاط ضعف آن‌ها را بپوشاند. در یک مسئله تصمیم‌گیری گروهی، نظرات چند متخصص برای رسیدن به یک تصمیم مشترک باهم ترکیب می‌شود (۲۴). دلیل استفاده از رویکردهای فازی در طراحی سیستم پیشنهادی این است که عدم قطعیت در خروجی روش‌ها در نظر گرفته شود.

در روش رأی‌گیری فازی برای هر دسته داده، خروجی روش‌های PCA، RSM و SVR-Firefly با استفاده از توابع عضویت، فازی سازی شد. توابع عضویت در هر سه کلاس بدخیم[20]، نامعلوم[21] و خوش‌خیم[22] به‌صورت زیر تعریف شده است:

$$\mu_{c_{i,\text{Malignant}}}(r) = \begin{cases} 1 & \alpha \le x < \theta - \delta \\ \frac{\theta - x}{\delta} & \theta - \delta \le x < \theta \\ 0 & \theta \le x \le \beta \end{cases} \quad (10)$$

$$\mu_{c_{i,Unknown}}(r) = \begin{cases} 0 & \alpha \le x < \theta - \delta \\ \frac{x+\delta-\theta}{\delta} & \theta - \delta \le x < \theta \\ \frac{\theta+\delta-x}{\delta} & \theta \le x < \theta + \delta \\ 0 & \theta + \delta \le x \le \beta \end{cases} \quad (11)$$

که c_i روش رگرسیون مورد استفاده، r بردار ویژگی، $x = P_{c_i}(r)$ خروجی روش رگرسیون c_i برای ورودی r مرز θ ، $\beta = \max P_{c_i}(r)$ ،$\alpha = \min P_{c_i}(r)$ ، جداکننده کلاس‌های خوش‌خیم و بدخیم در حالت غیر فازی و δ طول محدوده نامعلوم است که توسط کاربر تعیین می‌شود. پس از مرحله فازی سازی، نظر هرکدام از روش‌ها به‌صورت زیر تعیین شده است:

که class یکی از سه حالت بدخیم، نامعلوم و خوش‌خیم و c_i روش رگرسیون i-ام مورداستفاده را نشان می‌دهد. در نهایت تصمیم‌گیری در مورد خوش‌خیم یا بدخیم بودن

$$V_i = \arg\max_{class}(\mu_{c_{i,class}}(r)) \, , i = 1, \dots, 3 \quad (12)$$

توده سرطان پستان با در نظر گرفتن رأی اکثریت (دو رأی از سه رأی) محاسبه شد.

در روش میانگین‌گیری فازی نیز مانند روش رأی‌گیری فازی ابتدا خروجی روش‌های PCA، RSM و SVR-Firefly با استفاده از توابع عضویت ذکرشده در بخش پیشین فازی سازی شد و درجه عضویت یک رکورد ورودی r در هر سه کلاس بدخیم، نامعلوم و خوش‌خیم تعیین شد. پس از مرحله فازی سازی، تصمیم‌گیری در مورد خوش‌خیم یا بدخیم بودن تومورهای پستان با روش میانگین‌گیری فازی به‌صورت زیر انجام شد:

$$V = \arg\max_{class}\left(\frac{1}{L}\sum_{i=1}^{L}\mu_{c_{i,class}}(r)\right) \quad (13)$$

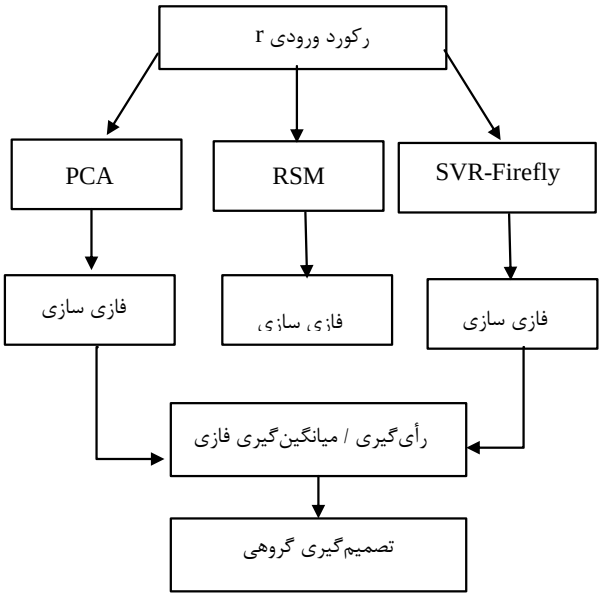
[19] group decision making
[20] Malignant
[21] Unknown
[22] Benign

www.SID.ir

Archive of SID

که class یکی از سه حالت بدخیم، نامعلوم و خوش‌خیم، L تعداد روش‌های رگرسیون مورد استفاده و c_i روش رگرسیون i-ام را نشان می‌دهد.

شکل ۱ روند به‌کار گرفته‌شده در سیستم تصمیم‌گیری گروهی پیشنهادی در این مقاله را نشان می‌دهد.



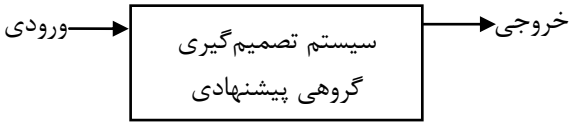
شکل ۱: ترکیب روش‌های رگرسیون با روش رأی‌گیری / میانگین‌گیری فازی در طراحی سیستم پیشنهادی

پس از طراحی سیستم پیشنهادی، عملکرد آن با استفاده از مجموعه داده ویسکانسین مورد ارزیابی قرار گرفت.

یافته‌ها

ابتدا، شاخص‌های صحت، دقت، حساسیت و اختصاصیت در شبیه‌سازی هر یک از روش‌های رگرسیون برای مجموعه داده به‌کاررفته در این مقاله محاسبه‌شده و در جدول ۲ نشان داده‌شده است. همان‌طور که در جدول ۲ مشاهده می‌شود، روش‌های رگرسیون به‌تنهایی قابلیت رقابت با سایر روش‌های طراحی‌شده در تشخیص و پیش‌بینی نوع توده‌های پستان را دارد.

همان‌طور که پیش‌تر اشاره شد سیستم تصمیم‌گیری گروهی پیشنهادشده با ترکیب روش‌های رگرسیون، یک‌بار با روش رأی‌گیری فازی و بار دیگر با روش میانگین‌گیری فازی طراحی‌شده است. شکل ۲ شمای کلی از سیستم طراحی‌شده را نشان می‌دهد.



شکل ۲: سیستم تصمیم‌گیری گروهی پیشنهادی

پس از آموزش سیستم پیشنهادی، از داده‌های دسته آزمون به‌عنوان ورودی سیستم استفاده شد. درواقع ورودی یک ماتریس با ۹ ستون (تعداد متغیرهای بالینی) و به تعداد داده‌های دسته آزمون سطر (رکورد مربوط به هر بیمار) است. سیستم پیشنهادی هر رکورد ورودی را به دو دسته خوش‌خیم و بدخیم دسته‌بندی می‌کند و نوع توده را به‌عنوان خروجی می‌دهد. در روش رأی‌گیری فازی، این خروجی با رأی اکثریت به دست آمد. به‌این‌ترتیب که اگر در خروجی سه روش رگرسیون ذکرشده پس از فازی سازی، دو روش (یا بیشتر) نوع توده را بدخیم تشخیص دادند خروجی سیستم پیشنهادی برای آن رکورد بدخیم و در غیر اینصورت خروجی سیستم برای آن رکورد خوش‌خیم در نظر گرفته شد. در روش میانگین‌گیری فازی خروجی سیستم پیشنهادی با استفاده از فرمول ۱۰ به دست آمد.

لازم به ذکر است که در روش‌های رگرسیون، خروجی حاصل با حد آستانه استاندارد ۰/۵ به دو نوع خوش‌خیم و بدخیم دسته‌بندی‌شده است.

به‌منظور ارزیابی عملکرد این سیستم بر مجموعه داده به‌کاررفته در این مقاله، خروجی سیستم پیشنهادی با دو روش رأی‌گیری فازی و میانگین‌گیری فازی در مرحله آزمون بر اساس شاخص‌های به‌دست‌آمده از ماتریس کانفیوژن محاسبه‌شده و به ترتیب در جداول ۳ و ۴ نشان داده‌شده است.

پس از انجام تصمیم‌گیری گروهی بر اساس نتایج خروجی هر یک از روش‌ها مشاهده شد که سیستم پیشنهادی در داده‌های دسته ششم بهترین عملکرد را دارد. شاخص حساسیت نشان‌دهنده توانایی سیستم در تشخیص نوع بدخیم و شاخص اختصاصیت نشان‌دهنده توانایی سیستم در تشخیص نوع خوش‌خیم است.

www.SID.ir

Archive of SID

جدول ۲: شاخص‌های عملکردی در بررسی روش‌های رگرسیون

	PCA				RSM				SVR-Firefly			
	صحت	دقت	حساسیت	اختصاصیت	صحت	دقت	حساسیت	اختصاصیت	صحت	دقت	حساسیت	اختصاصیت
دسته اول	۰/۹۲۲۱	۰/۹۵۵۶	۰/۸۱۱۳	۰/۹۸۰۲	۰/۹۷۴۰	۰/۹۴۵۴	۰/۹۸۱۱	۰/۹۷۰۳	۰/۹۷۴۰	۰/۹۴۵۵	۰/۹۸۱۱	۰/۹۷۰۳
دسته دوم	۰/۹۱۴۳	۰/۹۴۱۲	۰/۷۶۱۹	۰/۹۷۹۶	۰/۹۷۸۵	۰/۹۵۳۴	۰/۹۷۶۱	۰/۹۷۹۵	۰/۹۷۸۶	۰/۹۵۳۵	۰/۹۷۶۲	۰/۹۷۹۶
دسته سوم	۰/۹۲۵۲	۰/۹۷۴۴	۰/۷۹۱۷	۰/۹۸۹۹	۰/۹۷۲۷	۰/۹۷۸۲	۰/۹۳۷۵	۰/۹۸۹۹	۰/۹۷۹۶	۰/۹۷۸۷	۰/۹۵۸۳	۰/۹۸۹۹
دسته چهارم	۰/۹۴۰۰	۰/۹۳۷۵	۰/۸۱۰۸	۰/۹۸۲۳	۰/۹۶۶۶	۰/۸۸۰۹	۱	۰/۹۵۵۷	۰/۹۶۶۷	۰/۸۸۱۰	۱	۰/۹۵۵۸
دسته پنجم	۰/۹۶۰۵	۰/۹۸۰۸	۰/۹۱۰۷	۰/۹۸۹۶	۰/۹۶۷۱	۰/۹۴۷۳	۰/۹۶۴۲	۰/۹۶۸۷	۰/۹۷۳۷	۰/۹۴۸۳	۰/۹۸۲۱	۰/۹۶۸۸
دسته ششم	۰/۹۷۱۴	۱	۰/۹۲	۱	۱	۱	۱	۱	۱	۱	۱	۱

جدول ۳: عملکرد سیستم تصمیم‌گیری گروهی پیشنهادی با روش رأی‌گیری فازی در مرحله آزمون

	صحت	دقت	حساسیت	اختصاصیت
دسته اول	۰/۹۷۴۰	۰/۹۴۵۵	۰/۹۸۱۱	۰/۹۷۰۳
دسته دوم	۰/۹۷۸۶	۰/۹۵۳۵	۰/۹۷۶۲	۰/۹۷۹۶
دسته سوم	۰/۹۸۶۴	۰/۹۷۹۲	۰/۹۷۹۲	۰/۹۸۹۹
دسته چهارم	۰/۹۸۰۰	۰/۹۲۵۰	۱	۰/۹۷۳۵
دسته پنجم	۰/۹۸۰۳	۰/۹۴۹۲	۱	۰/۹۶۸۸
دسته ششم	۱	۱	۱	۱
میانگین دسته‌ها	۰/۹۸۳۲	۰/۹۵۸۸	۰/۹۹۰۰	۰/۹۸۳۲

جدول ۴: عملکرد سیستم تصمیم‌گیری گروهی پیشنهادی با روش میانگین‌گیری فازی در مرحله آزمون

	صحت	دقت	حساسیت	اختصاصیت
دسته اول	۰/۹۷۴۰	۰/۹۴۵۵	۰/۹۸۱۱	۰/۹۷۰۳
دسته دوم	۰/۹۷۸۶	۰/۹۵۳۵	۰/۹۷۶۲	۰/۹۷۹۶
دسته سوم	۰/۹۷۹۶	۰/۹۴۱۲	۱	۰/۹۶۹۷
دسته چهارم	۰/۹۸۰۰	۰/۹۲۵۰	۱	۰/۹۷۳۵
دسته پنجم	۰/۹۸۰۳	۰/۹۴۹۲	۱	۰/۹۶۸۸
دسته ششم	۱	۱	۱	۱
میانگین دسته‌ها	۰/۹۸۲۰	۰/۹۵۲۴	۰/۹۹۲۹	۰/۹۸۰۴

همان‌طور که مشاهده می‌شود در هر دو روش رأی‌گیری فازی و میانگین‌گیری فازی شاخص‌های صحت، دقت، حساسیت و اختصاصیت در دسته ششم داده، معادل عدد یک و در حالت متوسط با روش رأی‌گیری فازی به ترتیب معادل اعداد ۰/۹۸۳۲، ۰/۹۵۸۸، ۰/۹۹۰۰ و ۰/۹۸۳۲ و در روش میانگین‌گیری فازی به ترتیب معادل اعداد ۰/۹۸۲۰، ۰/۹۵۲۴، ۰/۹۹۲۹ و ۰/۹۸۰۴ به‌دست‌آمده است.

این امر نشان می‌دهد که سیستم پیشنهادی توانایی مناسبی در دسته‌بندی نوع توده‌های پستان دارد.

بحث

در این مقاله با ترکیب فازی روش‌های رگرسیون PCA، RSM و SVR-Firefly یک سیستم CAD بر اساس تصمیم‌گیری گروهی در پیش‌بینی نوع توده پستان طراحی و ارزیابی شد. نوآوری پژوهش حاضر طراحی یک سیستم تصمیم‌گیری گروهی بر اساس ترکیب فازی روش‌های رگرسیون بود که تاکنون در پیش‌بینی نوع توده پستان به کار نرفته است. به‌منظور ارزیابی سیستم

www.SID.ir

Archive of SID

پیشنهادشده و مقایسه آن با دیگر سیستم‌های ذکرشده، در این پژوهش نیز از مجموعه داده ویسکانسین استفاده شد و شاخص‌های صحت، دقت، حساسیت و اختصاصیت به‌عنوان معیار ارزیابی در نظر گرفته‌شد. بر اساس مطالعات، تاکنون سیستم‌های مختلفی به کمک روش‌های هوش مصنوعی برای دسته‌بندی نوع توده پستان به دو دسته خوش‌خیم و بدخیم طراحی‌شده و بر روی مجموعه داده معیار ویسکانسین ارزیابی‌شده است.

نتایج شبیه‌سازی نشان داد شاخص‌های به‌دست‌آمده نسبت به سایر روش‌های شبیه‌سازی‌شده بر این مجموعه داده بهبودیافته است.

اثری و همکاران الگوریتم‌های یادگیری ماشین ازجمله ماشین بردار پشتیبان (SVM)، درخت تصمیم (C4.5)، نایو بیز (NB) و K-NN را در نرم‌افزار WEKA پیاده‌سازی کردند. نتیجه مطالعه آن‌ها نشان داد ماشین بردار پشتیبان به بالاترین صحت (۹۷/۱۳٪) نسبت به سایر روش‌ها رسیده است (۲۵). در مقاله عماد و همکاران از روش (OWA[۲۳]) برای دسته‌بندی تومورهای سرطان پستان استفاده‌شده است. آن‌ها OWA را در آموزش K-NN، SVM و رگرسیون لجستیک به کار بردند و بیشترین صحت را با استفاده از روش K-NN، ۹۹/۷۱٪ گزارش کردند (۲۶). در مقاله دیگری از شبکه باور عمیق برای تشخیص زودهنگام سرطان پستان استفاده شد و صحت به‌دست‌آمده در بهترین حالت ۹۹/۶۸٪ گزارش شد (۲۷). در مقاله پنگ و همکاران یک الگوریتم نیمه نظارتی برای تشخیص سرطان پستان طراحی‌شده است و دقت به‌دست‌آمده ۹۵/۹٪ گزارش‌شده است (۲۸). کاراباتاک در مقاله خود روش نایو بیز وزن‌دار را طراحی و برای تشخیص بیماری سرطان پستان به‌کاربرده است و شاخص‌های حساسیت، اختصاصیت و صحت را به ترتیب ۹۹/۱۱٪، ۹۸/۲۵٪، ۹۸/۵۴٪ گزارش کرده است (۲۹). ورنر و همکاران الگوریتم ژنتیک را برای تشخیص سرطان پستان به کاربردند و به صحت ۹۶/۳۲٪ رسیدند (۳۰). خسروانیان و همکاران با استفاده از شبکه عصبی احتمالی یک سیستم تصمیم‌یار طراحی کردند و شاخص‌های حساسیت، اختصاصیت و صحت را به ترتیب ۱، ۰/۹۸ و ۰/۹۹ گزارش کردند (۱۳). سلطانی سروستانی و همکاران مقایسه‌ای بین شبکه‌های عصبی مصنوعیMLP، SOM، PNN و RBF بر دو مجموعه داده ویسکانسین و بیمارستان نمازی شیراز در تشخیص سرطان پستان انجام دادند. نتایج این تحقیق حاکی از برتری شبکه‌های PNN و RBF در تشخیص این بیماری بود (۳۱). در مطالعه‌ای دیگر بر مجموعه داده بیمارستان ویسکانسین روش‌های شبکه عصبی RBF درخت تصمیم J48 و CART، نایو بیز و ترکیب SVM و RBF پیاده‌سازی شده و صحت به‌دست‌آمده به ترتیب ۹۶/۶۶، ۹۴/۵۹، ۹۴/۲۷، ۹۶/۵۰ و ۹۶/۸۴ گزارش‌شده است (۳۲). سالاما و همکاران با استفاده از روش‌های SMO، نایو بیز، MLP، J48، IBK بر روی مجموعه داده بیمارستان ویسکانسین به ترتیب به صحت ۹۶/۹۹٪، ۹۵/۹۹٪، ۹۵/۲۷٪، ۹۵/۱۳٪، ۹۴/۵۶٪ رسیدند (۳۳). دلن و همکاران از شبکه‌های عصبی مصنوعی، درخت تصمیم‌گیری و رگرسیون لجستیک برای توسعه مدل‌های پیش‌بینی سرطان پستان بهره جستند. نتایج تحقیقات آن‌ها نشان داد که الگوریتم درخت تصمیم برای استخراج دانش از داده‌های موجود، مقدم بر سایر روش‌ها بود و نتایج به‌دست‌آمده از تحقیق، نزدیک به واقعیت بود (۳۴). همچنین یی و فویانگ از ماشین بردار پشتیبان (SVM) برای کشف الگوهای تشخیص سرطان پستان استفاده کردند. نتایج مطالعه آن‌ها نشان داد که SV برای تشخیص الگوهای سرطان پستان روش مناسبی بود نتایج به‌دست‌آمده با شواهد موجود و واقعی مطابقت داشت (۳۵). صالحی و همکاران به مقایسه مدل شبکه عصبی مصنوعی و رگرسیون کاکس در پیش‌بینی بقای بیماران مبتلا به سرطان پستان پرداختند که نتایج حاصل، نشان‌دهنده برتری نسبی شبکه عصبی مصنوعی درپیش‌بینی بقای بیماران مبتلا به سرطان پستان بود (۳۶).

نتایج پژوهش حاضر نشان داد سیستم پیشنهادی در این مقاله شاخص‌های عملکردی را نسبت به سایر روش‌ها بهبود داده است. ترکیب روش‌های رگرسیون و تجمیع خروجی‌های آن به روش فازی در پیش‌بینی نوع توده‌های پستان موجب افزایش دقت سیستم پیشنهادی شده است. در واقع بهبود عملکرد با ترکیب فازی خروجی روش‌های رگرسیون نقطه قوت این پژوهش را نشان می‌دهد.

از نقاط ضعف این پژوهش به‌روز نبودن داده‌ها و یکسان بودن وزن ویژگی‌های ورودی است. ضمن اینکه میزان تأثیر این ویژگی‌ها توسط پزشک متخصص در تشخیص

[23] Ordered Weighted averaging Operator

www.SID.ir

Archive of SID

خوش‌خیم یا بدخیم بودن توده پستان مشخص نشده است که در صورت مشخص بودن می‌توان در وزن‌دهی ویژگی‌ها یا تغییر یک ویژگی به ویژگی مؤثرتر، نظر پزشک را اعمال کرد و نتایج حاصل را ارزیابی کرد. همچنین در مشورت با پزشک ممکن است تعداد متغیرهای بالینی تغییر کند، افزایش یابد یا از تعداد آن کاسته شود. این موضوع می‌تواند به‌عنوان پیشنهادی برای پژوهش‌های آتی در نظر گرفته شود. به‌هرحال جنبه پزشکی این موضوع موردنظر نویسندگان این مقاله نبوده و هدف بررسی روشی جدید بر روی این مجموعه داده بوده است که قابلیت مقایسه با سایر تحقیقات انجام‌شده بر این مجموعه داده را داشته باشد.

نتیجه‌گیری

بررسی عملکرد سیستم تصمیم‌گیری گروهی پیشنهادی در این مقاله با شاخص‌های صحت، دقت، حساسیت و اختصاصیت نشان داد این سیستم در پیش‌بینی نوع توده پستان و دسته‌بندی آن به دودسته خوش‌خیم و بدخیم موفق عمل کرده است. این امر نشان می‌دهد که ترکیب فازی روش‌های رگرسیون، قابلیت رقابت با سایر روش‌های طراحی‌شده تاکنون در پیش‌بینی نوع سرطان پستان را دارد و نرم‌افزار طراحی‌شده بر اساس آن می‌تواند به‌عنوان یک سیستم تصمیم‌یار در کمک به پزشک و آموزش دانشجویان پزشکی مفید باشد.

References

1. Piao Y, Piao M, Ryu KH. Multiclass cancer classification using a feature subset-based ensemble from microRNA expression profiles. Comput Biol Med 2017; 80:39-44.
2. American Cancer Society, Breast Cancer: Facts and Figures, ACS, Atlanta, 2015.
3. National Center for Chronic Disease Prevention and Health Promotion, https://www.cdc.gov/cancer/breast/basic_info/risk_factors.htm
4. Mahmoodian H, Ebrahimian L. Using support vector regression in gene selection and fuzzy rule generation for relapse time prediction of breast cancer. Biocybern Biomed Eng 2016; 36: 466-72.
5. Krawczyk B, Galar M, Jelen L, Herrera F. Evolutionary undersampling boosting for imbalanced classificationof breast cancer malignancy. Appl Soft Comput 2016; 38:714-26.
6. Lundin M, Lundin J, Burke HB,Toikkanen S, Pylkkanen L, Joensuu H. Artificial neural networks applied to survival prediction in breast cancer. Oncology 1999; 57(4):281-6.
7. Rahmani Seryasat O, Haddadnia J, Ghayoumi Zadeh H. Assessment of a Novel Computer Aided Mass Diagnosis System in Mammograms .Iranian Journal of Breast Diseases 2016; 9(3): 31-41.
8. Darzi M, Olfat Bakhsh A, Gorgin S, Oveisi F, Hashemi E, Alavi N. Imbalanced Data Classification for Primary Diagnosis of Breast Diseases by AdaBoost. M1, K-Nearest Neighbor and Probabilistic Neural Network .Iranian Journal of Breast Diseases 2016; 9(2): 7-18.
9. Sheikhpour R, Agha Sarram M, Zare Mirakabad M R, Sheikhpour R. Breast Cancer Detection Using Two-Step Reduction of Features Extracted From Fine Needle Aspirate and Data Mining Algorithms. Iranian Journal of Breast Diseases 2015; 7(4): 43-51.
10. Carvalho D, Pinheiro PR, Pinheiro M. A Hybrid Model to Support the Early Diagnosis of Breast Cancer. Procedia Comput Sci 2016; 91: 927-34.
11. Dora L, Agrawal S, Panda R, Abraham A. Optimal breast cancer classification using Gauss–Newton representation based algorithm. Expert Syst Appl 2017; 85: 134-45.
12. UCI Machine Learning Repository. Available at: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original).

www.SID.ir

Archive of SID

13. Khosravanian A, Ayat S. Diagnosing Breast Cancer Type by Using Probabilistic Neural Network in Decision Suport System. Int J Knowl Eng 2016; 2(1):73-6.

14. Jolliff I. Principal Component Analysis. New York: Springer-Verlag 1986.

15. Flores-Fernandez JM, Herrera-Lopez EJ, Sanchez-Llamas F, Rojas-Calvillo A, Cabrera-Galeana PA, Leal-Pacheco G, Gonzalez-Palomar MG, Femat R, Martinez-Velazquez M. Development of an optimized multi-biomarker panel for the detection of lung cancer based on principal component analysis and artificial neural network modeling. Expert Syst Appl 2012; 39(12):10851-6.

16. Ghosh UN, Chatterjee P, Ghosh DK. Prediction of Prostate Cancer Cells based on Principal Component Analysis Technique. Procedia Technology 2013; 10:37-44.

17. Vesprini D, Sia M, Lockwood G, Moseley D, Rosewall T, Bayley A and et al. Role of Principal Component Analysis in Predicting Toxicity in Prostate Cancer Patients Treated With Hypofractionated Intensity-Modulated Radiation Therapy. Int J Radiat Oncol Biol Phys 2011; 81(4):415-21.

18. Myers RH, Montgomery DC, Anderson-Cook CM, Response surface methodology: process and product optimization using designed experiments. Canada: Wiley 2016.

19. Mirhasani A, Hasanabadi M, Motahari M, Asgary A. Oil Production Optimization in Smart Wells Using Experimental Design Methodology. J Pet Res 2012; 72:38-57.

20. Smola AJ, Scholkopf B. A tutorial on support vector regression. Stat Comput 2004; 14(3):199-222.

21. Shokri S, Sadeghi M, Ahmadi Marvast M. An Integrated Method of Data Pre-processing in Support Vector Regression for the Quality Prediction of Treated Gas-oil. J Pet Res 2013; 75:102-16.

22. Zaji AH, Bonakdari H, Khodashenas SR, Shamshirband S. Firefly optimization algorithm effect on support vector regression prediction improvement of a modified labyrinth side weir's discharge coefficient. Appl Math Comput 2016; 274:14-9.

23. Al- Shammari ET, Keivani A, Shamshirband S, Mostafaeipour A, Yee P, Petkovic D and et al. Prediction of heat load in district heating systems by Support Vector Machine with Firefly searching algorithm. Energy 2016; 95: 266-73.

24. Merigo J, Gil-Lafuente A. Fuzzy induced generalized aggregation operators and its application in multi-person decision making. Expert Syst Appl 2011; 37: 9761-72.

25. Asri H, Mousannif H, Al Moatassime H, Noel T. Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. Procedia Comput Sci 2016; 83:1064-9.

26. Emad. AM, Christopher. TN, Behrouz. HF. Breast tumor classification using a new OWA operator. Expert Syst Appl 2016; 61:302-13.

27. Ahmed MA, Ayman ME. Breast cancer classification using deep belief networks. Expert Syst Appl 2016; 46:139-44.

28. Peng L, Chen W, Zhou W, Li F, Yang J, Zhang J. An immune-inspired semi-supervised algorithm for breast cancer diagnosis. Comput Methods Programs Biomed 2016; 134:259-65.

29. Karabatak M. A new classifier for breast cancer detection based on Naïve Bayesian. Measurement 2015; 72: 32-6.

30. Werner JC, Fogarty TC. Genetic Programming Applied to Severe Diseases Diagnosis. In Proceedings Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP) 2001.

31. Sarvestan Soltani A, Safavi AA, Parandeh MN, Salehi M. Predicting Breast Cancer Survivability using data mining techniques. Software Technology and Engineering (ICSTE). 2nd International Conference 2010; 2: 227-31.

32. Aruna S, Rajagopalan DS, Nandakishore LV. Knowledge based analysis of various statistical tools in detecting breast cancer. Comput Sci Inform Tech 2011; 2: 37-45.

33. Salama GI, Abdelhalim MB, Zeid MAE. Breast Cancer Diagnosis on Three

www.SID.ir

Archive of SID

Different Datasets Using Multi-Classifiers. Int J Comput Sci Inform Tech 2012; 1(1): 36-43.

34. Ghasem Ahmad L. Using Data Mining Techniques for Prediction Breast Cancer Recurrence. Iranian Journal of Breast Disease 2013; 5(4):23-34.

35. Yi W, Fuyong W. Breast cancer diagnosis via support vector machines. In Proc. the Twenty. Control Conference, CCC 2006; 1853-6.

36. Salehi M, gohari M, vahabi N, Zayeri F, yahyazadeh S, kafashian M. Comparison of Artificial Neural Network and Cox Regression Models in Survival Prediction of Breast Cancer Patients. journal of ilam university of medical sciences 2013; 21 (2):120-8.

37. Setiono R. Generating concise and accurate classification rules for breast cancer diagnosis. Artif Intell Med 2000; 18(3):205-19.

38. Chatterjee A, Rakshit A. a new fuzzy pattern classifier. IEEE Trans Knowl Data Eng 2004; 16(8): 881-93.

39. Iranpour M, Almassi S, Analoui M, Breast Cancer Detection from FNA using SVM and RBF Classifier. In 1st Joint Congress on Fuzzy and Intelligent Systems 2007.

40. Cheng C, Wang J, Wu M. OWA-Weighted based clustering method for classification problem. Expert Syst Appl 2009; 36(3): 4988-95.

www.SID.ir