

Prediction of Breast Tumor Malignancy Using Neural Network and Whale Optimization Algorithms (WOA)

Sharifi A¹, Alizadeh K^{2*}

¹ Ph.D. Student in Analytical Chemistry, Department of Chemistry, Lorestan University, Khorramabad, Iran

²Department of Chemistry, Faculty of Basic Science, Lorestan University, Khorramabad, Iran

Receive: 15/5/2019
Accepted: 14/8/2019

*Corresponding Author:
alizadeh.k@lu.ac.ir

Ethics Approval:
Not applicable

Abstract

Introduction: Breast cancer is the most prevalent cause of cancer mortality among women. Early diagnosis of breast cancer gives patients greater survival time. The present study aims to provide an algorithm for more accurate prediction and more effective decision-making in the treatment of patients with breast cancer.

Methods: The present study was applied, descriptive-analytical, based on the use of computerized methods. We obtained 699 independent records containing nine clinical variables from the UCI machine learning. The EM algorithm was used to analyze the data before normalizing them. Following that, a combination of neural network model based on multilayer perceptron structure with the Whale Optimization Algorithm (WOA) was used to predict the breast tumor malignancy.

Results: After preprocessing the disease data set and reducing data dimensions, the accuracy of the proposed algorithm for training and testing data was 99.6% and 99%, respectively. The prediction accuracy of the proposed model was 99.4%, which would be a satisfying result compared to different methods of machine learning in other studies.

Conclusion: Considering the importance of early diagnosis of breast cancer, the results of this study may have highly useful implications for health care providers and planners so as to achieve the early diagnosis of the disease.

Keywords: Breast Cancer, EM Algorithm, Whale Optimization Algorithm (WOA), Artificial Neural Network, Multilayer Perceptron

پیش‌بینی بدخیمی تومور پستان با استفاده از شبکه عصبی و الگوریتم بهینه‌سازی وال (WOA)

علی شریفی^۱، کمال علیزاده^{۲*}

^۱ دانشجوی دکتری تخصصی شیمی تجزیه، گروه شیمی، دانشکده علوم پایه، دانشگاه لرستان، لرستان، ایران

^۲ گروه شیمی، دانشکده علوم پایه، دانشگاه لرستان، لرستان، ایران

چکیده

مقدمه: سرطان پستان به عنوان یکی از شایع‌ترین علل مرگ و میر در میان زنان در نظر گرفته می‌شود. تشخیص زودهنگام سرطان پستان شانس زنده ماندن را افزایش می‌دهد. مطالعه حاضر جهت پیش‌بینی دقیق‌تر و تصمیم‌گیری مؤثرتر در درمان بیماران مبتلا به سرطان پستان صورت گرفته است.

روش بررسی: مطالعه حاضر که از نوع کاربردی و توصیفی-تحلیلی بر اساس بهره‌گیری از روش‌های کامپیوتری است، جامعه‌ی هدف آن متشکل از ۶۹۹ مورد بیماران مبتلا به سرطان پستان خوش‌خیم و بدخیم با ۹ متغیر ورودی روی مجموعه داده‌های بیماری سرطان پستان پایگاه UCI انجام شده است، قبل از نرمال‌سازی داده‌ها از الگوریتم EM برای داده‌کاوی استفاده شده است. سپس از مدل ترکیب شبکه عصبی مبتنی بر ساختار پرسپترون چند لایه با الگوریتم بهینه‌سازی وال (WOA) برای پیش‌بینی بدخیمی تومور پستان استفاده شده است و دقت و پیش‌بینی آن مورد بررسی و مقایسه قرار گرفته است.

یافته‌ها: نتایج مطالعه حاضر نشان می‌دهد که پس از پیش‌پردازش مجموعه داده‌های بیماری، دقت الگوریتم پیشنهادی برای داده‌های آموزش و آزمون به ترتیب برابر با ۹۹/۶ و ۹۹ بوده است و همچنین دقت پیش‌بینی مدل پیشنهادی برابر ۹۹/۴ به دست آمد که با مقایسه صورت گرفته نسبت به روش‌های مختلف یادگیری ماشین در مطالعات دیگر نتیجه خوبی می‌باشد.

نتیجه‌گیری: با توجه به اهمیت تشخیص زودهنگام بیماری سرطان پستان، یافته‌های این مطالعه می‌تواند به برنامه‌ریزان و ارائه‌کنندگان خدمات سلامت در برنامه‌های تشخیص به موقع این بیماری کمک شایانی نماید.

واژه‌های کلیدی: سرطان پستان، الگوریتم EM، الگوریتم بهینه‌سازی وال WOA، شبکه عصبی مصنوعی، ساختار پرسپترون چند لایه

تاریخ ارسال: ۱۳۹۸/۲/۲۵

تاریخ پذیرش: ۱۳۹۸/۵/۲۳

* نویسنده مسئول:

alizadeh.k@lu.ac.ir

مقدمه

سرطان پستان، شایع‌ترین سرطان در بین خانم‌ها و یکی از عوامل عمده مرگ ناشی از سرطان در زنان ۲۱ الی ۵۹ سال می‌باشد. بر اساس آمار وزارت بهداشت، درمان و آموزش پزشکی، این بیماری در سال‌های اخیر به بیماری شماره یک برای زنان در ایران تبدیل شده است (۱). سرطان پستان، تهدید بزرگی بر سلامت زنان بوده و از عوامل شایع در کاهش عمر زنان به شمار می‌رود (۲). وجود شباهت زیاد در علائم بالینی و آزمایشگاهی سرطان پستان احتمال تشخیص نادرست را افزایش می‌دهد (۳). بنابراین پیش‌بینی و تشخیص سریع این سرطان شانس درمان را بالا می‌برد و 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 پیش‌بینی بر اساس آنها دارای کمترین خطا و بیشترین دقت، صحت و اطمینان باشد.

بهره‌گیری از روش‌های هوشمند تشخیص بیماری سرطان پستان می‌تواند گام مؤثری در جهت تشخیص زود هنگام این بیماری باشد. در زمینه تشخیص بیماری سرطان پستان با استفاده از روش‌های داده‌کاوی طی سال‌های اخیر کارهای زیاد انجام شده که در این قسمت به چند نمونه آن اشاره می‌کنیم.

پندهارکر و همکارانش در سال ۱۹۹۹ برای بررسی الگوهای موجود در سرطان پستان از چندین روش داده‌کاوی (Data Mining) استفاده کردند. نتیجه مطالعه آنها نشان داد داده‌کاوی می‌تواند به نحو مطلوبی در شناسایی الگوها در تشخیص سرطان پستان به کار رود (۴). سدهی و همکارانش در سال ۲۰۱۴ تحلیل بقای بیماران مبتلا به سرطان پستان با استفاده از مدل‌های شبکه عصبی مصنوعی و رگرسیون کاکس را مورد بررسی قرار داده و در بین مدل‌های شبکه عصبی، مدل‌های با الگوریتم آموزش OSS، SCG و LM به ترتیب با صحت پیش‌بینی ۹۴، ۹۰ و ۷۸ درصد برای داده‌های اعتبارسنجی، بیشترین کارایی را داشتند. سطح زیر منحنی مشخصه عملکرد برای مدل‌های مذکور به ترتیب ۰/۹۹۱، ۰/۹۷۲ و ۰/۸۳۷ و برای مدل رگرسیون کاکس ۰/۸۶۹

به‌دست آمد. چنانچه معماری مناسبی برای مدل شبکه عصبی مصنوعی انتخاب گردد، این مدل نسبت به مدل رگرسیون کاکس کارایی بیشتری برای پیش‌بینی وضعیت بقا دارد (۵).

اگر چه نتایج تحقیقات پیشین گامی مؤثر در جهت تشخیص بیماری سرطان پستان محسوب می‌شود اما هر یک از مقالاتی که از داده‌های مورد استفاده در این مقاله بهره جستند دارای نقاط ضعفی بوده که از آن جمله می‌توان به حذف داده‌های گزارش نشده در تجزیه و تحلیل داده‌ها اشاره کرد. از نقاط قوت مطالعه حاضر علاوه بر کاهش هزینه‌های تشخیص، بهره‌گیری از طبقه‌بندی با کارایی بهتر بدون وجود عوارض ناشی از روش‌های تهاجمی، دقت تشخیص بالا نسبت به مقالات ذکر شده، انتخاب عنوان متناسب با دیتای در دسترس و مقایسه کامل با پژوهش‌های صورت گرفته تا امروز است. در این مقاله، روشی جدید مبتنی بر الگوریتم EM و ترکیب شبکه عصبی مصنوعی با ساختار پرسپترون چند لایه با الگوریتم بهینه‌سازی وال (WOA) به عنوان الگوریتم فراابتکاری برای پیش‌بینی اتوماتیک بدخیمی تومور پستان پیشنهاد شده است که با توجه اهمیت روزافزون تشخیص زودهنگام بیماری سرطان پستان تلاش شده است که سرعت و دقت تشخیص بیماری را نسبت به پژوهش‌های مشابه افزایش دهد.

مواد و روش‌ها

این مطالعه از نوع کاربردی و توصیفی-تحلیلی و روش پیشنهادی با استفاده از مجموع داده مربوط به بیماران مبتلا به سرطان پستان بیمارستان ویسکانسین، موجود در انبار داده یادگیری ماشین دانشگاه ایروین، کالیفرنیا (UCI: University of California, Irvine) (۶) مورد ارزیابی قرار گرفت. داده‌های در دسترس شامل ۶۹۹ مورد بیمار با ۹ پارامتر بود که ۱۶ مورد دارای ویژگی‌های کامل نبودند. همان طور که در جدول ۱ نشان داده شده است متغیرهای ورودی شامل فاکتورهای ضخامت انبوه، یکنواختی اندازه سلول، یکنواختی شکل سلول، چسبندگی لبه‌ها، حجم سلول بافت اپیتلیال، هسته‌های عربان، کروماتین بلاند، هسته عادی و تقسیم سلول به دو قسمت بود.

جدول ۱: متغیرهای ورودی در تشخیص سرطان پستان

متغیرهای بالینی (فارسی)	متغیرهای بالینی (انگلیسی)	محدوده	Means±SD
ضخامت انبوه	Clump thickness	۱ - ۱۰	۴/۴۴ ± ۲/۸۲۱
یکنواختی اندازه سلول	Uniformity of cell size	۱ - ۱۰	۳/۱۵ ± ۳/۶۵۰
یکنواختی شکل سلول	Uniformity of cell shape	۱ - ۱۰	۳/۲۲ ± ۲/۹۸۹
چسبندگی لبه‌ها	Marginal adhesion	۱ - ۱۰	۲/۸۳ ± ۲/۸۶۵
حجم سلول بافت اپیتلیال	Single epithelial cell size	۱ - ۱۰	۳/۲۳ ± ۲/۲۲۳
هسته‌های عریان	Bare nuclei	۱ - ۱۰	۳/۵۴ ± ۳/۶۴۴
کروماتین بلاند	Bland chromatin	۱ - ۱۰	۳/۴۵ ± ۲/۴۵۰
هسته عادی	Normal nucleoli	۱ - ۱۰	۲/۸۷ ± ۳/۰۵۳
تقسیم هسته سلول به دو قسمت	Mitoses	۱ - ۱۰	۱/۶۰ ± ۱/۷۳۳

حداکثرسازی (M-step). با تکرار الگوریتم با توجه به اینکه مقدار درست‌نمایی در هر مرحله افزایش می‌یابد، می‌توان نسبت به همگرایی مطمئن بود (۸، ۹).

نرمال‌سازی داده‌ها: پس از جایگزینی داده‌های مفقودی با مقادیر مناسب، 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 (۱۰). لذا در گام بعدی مقادیر هر یک از این متغیرها در جدول ۱ جهت طبقه‌بندی و دقت در تشخیص، نرمال‌سازی شد. جهت نرمال‌سازی به عنوان یکی از گام‌های پیش‌پردازش داده از روش کمینه-بیشینه استفاده شد. این روش در فرمول شماره ۱ ارائه شده است.

$$X_{\text{new}} = \left(2 \times \frac{(X - X_{\min})}{(X_{\min} - X_{\max})} \right) - 1 \quad (1)$$

($X_{\min}$)، ($X_{\max}$) به ترتیب کمترین و بیشترین مقدار در هر متغیر هستند. در این روش مقادیر اندازه‌گیری شده به محدوده‌ی جدیدی تبدیل شدند. این محدوده در این مقاله ± 1 تعریف شد.

الگوریتم بهینه‌سازی وال (WOA): این الگوریتم بهینه‌سازی فراابتکاری به عنوان یکی از جدیدترین الگوریتم‌های بهینه‌سازی مبتنی بر جمعیت (Population Based Optimization) که نخستین بار در سال ۲۰۱۶ توسط میرجلیلی و لویس در مقاله‌ای با همین عنوان مطرح گردید و توضیحات مدل ریاضی و الگوریتم بهینه‌سازی در آن ذکر گردیده (۱۱) با الهام

در این پژوهش ابتدا برای پیش‌پردازش داده‌ها از الگوریتم EM با استفاده از نرم‌افزار WEKA 3.9.2 استفاده گردید و سپس داده‌ها نرمال‌سازی شده و سپس از ترکیب شبکه هوش مصنوعی مبتنی بر پرسپترون چندلایه با الگوریتم بهینه‌سازی وال به منظور تشخیص نوع سرطان پستان استفاده گردید. برای انجام این تحلیل‌ها از نرم‌افزار MATLAB R2018b استفاده شده است.

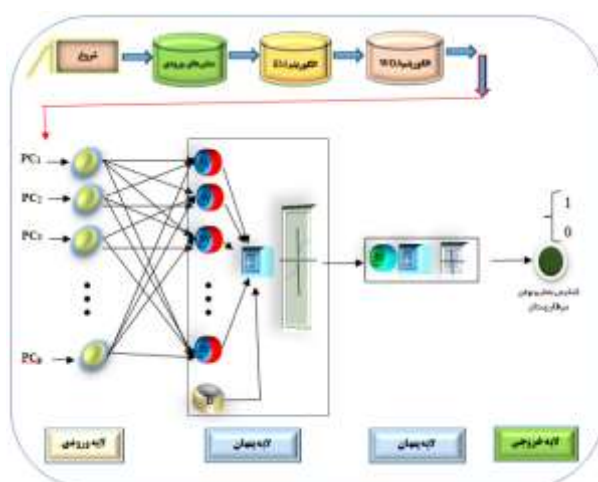
پیش‌پردازش و نرمال‌سازی داده‌ها: تکنیک‌های پیش‌پردازش و پاک‌سازی داده با هدف ارتقا کیفیت داده‌ها به کار بسته می‌شوند. وجود داده‌های گمشده در تحقیقات علوم پزشکی امری اجتناب‌ناپذیر است از آنجاییکه در دیتاست ما ویژگی‌هایی با مقادیر مفقودی (۱۶ مورد) وجود دارد و در پژوهش‌های پیشین، کل اطلاعات بیماری که دارای حتی یک مورد داده مفقود شده باشد کنار گذاشته شد و ما نمی‌خواهیم این اطلاعات را از دست دهیم بنابراین پیش از مدل‌سازی نیاز داریم که این مقادیر را با مقادیر مناسبی پر کنیم لذا مقادیر مفقودی با استفاده از الگوریتم EM مقدارگذاری می‌گردد.

الگوریتم EM: Expectation-maximization (algorithm) یکی از روش‌های مدرن و پیشرفته در حل مسئله گم‌شدگی تلقی می‌شود که از دیدگاه تئوری از پیچیدگی‌های خاصی برخوردار می‌باشد، اما از نظر کارایی عملکرد بهتری نسبت به روش‌های کلاسیک از جمله میانگین سوالات، میانگین فردی، نمای فردی و رگرسیون دارد (۷). این الگوریتم یک روند تکراری مؤثر به منظور محاسبه برآورد حداکثر درست‌نمایی در حضور داده‌های گم شده به حساب می‌آید. هر تکرار الگوریتم شامل دو مرحله می‌باشد: مرحله امید ریاضی (E-step) و مرحله

گرفته شده‌اند. برای تحلیل داده‌ها به کمک شبکه عصبی مصنوعی ابتدا داده‌های به دست آمده از مراحل قبل را به دو گروه آموزشی و آزمون تقسیم می‌کنیم. سپس انتخاب ویژگی توسط الگوریتم بهینه‌سازی وال (WOA) با استفاده از دو مدل ریاضی محاصره کردن و حمله شبکه حباب انجام گرفت. طبقه‌بندی داده‌ها توسط ویژگی‌های انتخاب شده توسط الگوریتم بهینه‌سازی وال در شبکه عصبی مصنوعی پرسپترون سه لایه آموزش داده شدند. این سه لایه شامل لایه ورودی متشکل از متغیرهای مستقل، لایه میانی یا پنهان و لایه خروجی متشکل از متغیر پاسخ هستند. برای حصول بهترین کلاس‌بندی، از معماری‌های مختلف برای شبکه عصبی از ۵ تا ۲۰ نرون در لایه میانی و بهینه‌سازی وزن ویژگی‌ها از طریق WOA در ۴۰۰ تکرار (Epoch) استفاده گردید. از شاخص میانگین مربعات خطا (MSE) به عنوان شاخص محاسبه عملکرد (Performance) استفاده شد که در تکرارهای اول کاهش خطا محسوس از ۰/۱۱۵ به ۰/۰۸ بدست آمد و در نهایت کمترین مقدار خطای به دست آمده برای آموزش و آزمون شبکه عصبی به ترتیب برابر ۰/۰۰۲۴۳ و ۰/۰۲۱۸ و در بهترین وضعیت، کمترین خطا با مقدار ۰/۰۰۲۳۶۵۹ به دست آمد. معماری شبکه پرسپترون سه لایه ۱-۱۲-۱۰-۹ با دارا بودن دو لایه میانی یا پنهان توانست بهترین دقت پیش‌بینی را نشان دهد. در شکل ۱ مدل ریاضی ساختار ترکیب پرسپترون چند لایه با الگوریتم بهینه‌سازی وال (WOA) بکار رفته برای پیش‌بینی بدخیمی تومور پستان نشان داده شده است.

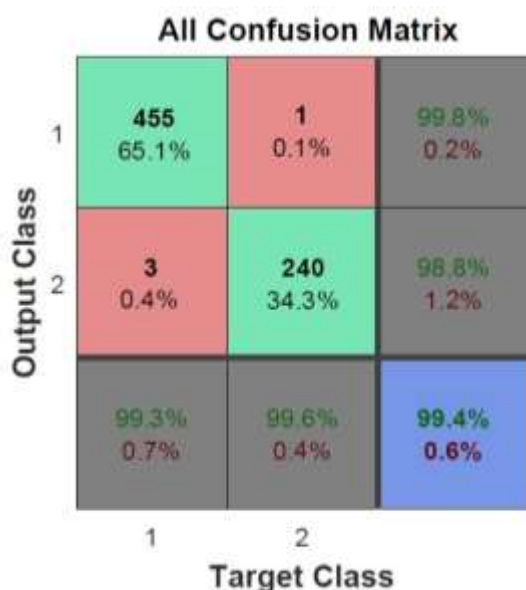
گرفتن از طبیعت و رفتار اجتماعی نهنگ‌های کوهان‌دار از استراتژی شکار شبکه حبابی جهت اکتشاف و بهره‌برداری بهره گرفته و با اجتناب از نقاط بهینه محلی، با تکنیک انطباقی یکپارچه بتواند به راه‌حل بهینه با هدر رفت زمان محاسباتی کمتر دست یابد (۱۱، ۱۲). کدنویسی این الگوریتم فراابتکاری در <http://www.alimirjalili.com/WOA.html> در دسترس است.

شبکه عصبی مصنوعی (ANN: Artificial Neural Network): یک شبکه عصبی مصنوعی، 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 (۱۳، ۱۴). 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 (۱۵). تفسیر اپیدمیولوژیک شبکه‌های عصبی در مقایسه با مدل‌های آماری مرسوم پیچیده‌تر است. با این وجود، این گونه مدل‌ها در زمینه‌های گوناگون علوم پزشکی از جمله اپیدمیولوژی (۱۶) پیش‌بینی سرطان پروستات (۱۷) پیش‌بینی حاملگی ناخواسته (۱۸) پیش‌بینی مرگ پس از جراحی قلب باز (۱۹) و پیش‌بینی هم زمان سندرم متابولیک و شاخص مقاومت به انسولین (۲۰) به کار



شکل ۱: مدل ریاضی ساختار ترکیب پرسپترون چند لایه با الگوریتم بهینه‌سازی وال برای پیش‌بینی بدخیمی تومور پستان

در نرم‌افزار اعمال گردید. نتایج آزمون شبکه عصبی با الگوریتم وال نشان می‌دهد که ۱۴۱ بیمار از ۱۴۲ بیمار مبتلا به تومور خوش‌خیم یعنی ۹۹/۳٪ و همچنین ۶۷ بیمار از ۶۸ بیمار مبتلا به تومور بدخیم یعنی ۹۸/۵٪ را در کلاس مربوطه به درستی دسته‌بندی نماید و با ۰/۱٪ خطای در پیش‌بینی به دقت ۹۹٪ رسیده است. نتایج بررسی کلی نشان می‌دهد که شبکه عصبی پرسپترون چندلایه آموزش دیده با الگوریتم وال توانسته است ۴۵۵ بیمار از ۴۵۸ بیمار مبتلا به تومور خوش‌خیم یعنی ۹۹/۳٪ و همچنین تعداد ۲۴۰ بیمار از ۲۴۱ بیمار مبتلا به تومور بدخیم یعنی ۹۹/۶٪ را در کلاس مربوطه به درستی دسته‌بندی نموده و به دقت ۹۹/۴٪ دست یابد. همچنین این شبکه آموزش دیده با الگوریتم وال، ۳ بیمار مبتلا به تومور خوش‌خیم یعنی ۰/۴٪ را به اشتباه در کلاس تومور بدخیم و تنها یک بیمار مبتلا به تومور بدخیم یعنی ۰/۱٪ را به اشتباه در کلاس تومور خوش‌خیم قرار داده است. ماتریس آشفتگی بهترین معماری در شکل ۲ نشان داده شد.



شکل ۲: ماتریس آشفتگی ترکیب پرسپترون چند لایه با الگوریتم بهینه‌سازی وال

بحث

پژوهش حاضر، با هدف پیش‌بینی بدخیمی تومور پستان با استفاده از روش‌های کامپیوتری انجام گردید و از مزایای آن می‌توان به دقت تشخیص بالا بدون وجود عوارض ناشی از روش‌های تهاجمی نام برد. در این پژوهش از مجموع

یافته‌ها

برای یافتن ۱۶ مورد داده‌های مفقودی از الگوریتم EM در این مقاله استفاده شده است. جهت انتخاب ویژگی و انجام فرآیند بهینه‌سازی از الگوریتم فراابتکاری بهینه‌سازی وال (WOA) استفاده شد.

مرحله آموزش شبکه عصبی: از مهم‌ترین اهداف شبکه عصبی یافتن وزن‌های متناسب با نرون‌ها در لایه‌های مختلف است. در این مرحله از ۷۰ درصد (۴۸۹ بیمار) داده‌های پیش‌پردازش شده و نرمال‌شده توسط الگوریتم EM، جهت آموزش شبکه عصبی مصنوعی با الگوریتم بهینه‌سازی وال استفاده شده است. شبکه عصبی پرسپترون چند لایه ساده، آموزش و بهینه‌سازی شبکه را بر اساس گرادیان و الگوریتم پس انتشار خطا (BP) انجام داده که دارای معایبی از جمله افتادن در دام مینیمم محلی و وابستگی به ساختار معین شبکه بوده در حالی که الگوریتم بهینه‌سازی وال (WOA) با قابلیت فرار از دام محلی و عدم وابستگی به ساختار معین شبکه، با تکنیک انطباقی یکپارچه می‌تواند به راه‌حل بهینه و تنظیم اوزان اتصال شبکه با هدر رفت زمان محاسباتی کمتر دست یابد. برای پیاده‌سازی شبکه عصبی در نرم‌افزار MATLAB از یک ماتریس ورودی شامل ۹ سطر و ۶۹۹ ستون و ماتریس دیگری با ۱ سطر (۱ کلاس خوش‌خیم و بدخیم) و ۶۹۹ ستون به‌عنوان ماتریس هدف استفاده شد و سپس به شبکه عصبی مصنوعی مبتنی بر پرسپترون چند لایه وارد شده‌اند. ماتریس هدف نشان‌دهنده دو کلاس خوش‌خیم و بدخیم بود.

در صورتی که نوع سرطان متعلق به کلاس مربوطه بود سطر مربوط به آن با عدد چهار و سطر دیگر با عدد دو پر شد. نتایج آموزش شبکه عصبی با الگوریتم وال نشان می‌دهد که ۳۱۴ بیمار از ۳۱۶ بیمار مبتلا به تومور خوش‌خیم یعنی ۹۹/۴٪ و همچنین تمامی ۱۷۳ بیمار مبتلا به تومور بدخیم یعنی ۱۰۰٪ را در کلاس مربوطه به درستی دسته‌بندی نماید و با ۰/۴٪ خطای در پیش‌بینی به دقت ۹۹/۶٪ رسیده است.

مرحله آزمون شبکه عصبی: در این مرحله ۳۰٪ (۲۱۰ بیمار) داده‌های پیش‌پردازش شده، نرمال‌شده توسط الگوریتم EM که در مرحله آموزش استفاده نشده بودند، به‌صورت بردار به شبکه عصبی مصنوعی پیاده‌سازی شده

داده‌های بیماری سرطان پستان موجود در انبار داده یادگیری ماشین دانشگاه ایروین کالیفرنیا استفاده گردید و با انتخاب الگوریتم EM و سپس ترکیب مدل شبکه عصبی مبتنی بر پرسپترون چندلایه با الگوریتم بهینه‌سازی وال (WOA) طراحی و ارزیابی صورت گرفت که نتیجه آن رسیدن به دقت ۹۹/۴ بوده است. در این بخش کارایی الگوریتم پیشنهادی با تعدادی از پژوهش‌های صورت گرفته تا امروز مقایسه می‌شود.

احمد و همکارانش در سال ۲۰۱۴ با استفاده از انتخاب ویژگی‌های مبتنی بر الگوریتم ژنتیک و بهینه‌سازی پارامترهای شبکه هوش مصنوعی توانستند با استفاده از بهترین الگوریتم یعنی GAANN_BP بهترین طبقه‌بندی و طبقه‌بندی متوسط صحیح را به ترتیب با دقت ۹۹/۲۴ و ۹۸/۲۹ درصد بر روی مجموعه داده‌های سرطان پستان به دست آورند که نسبت به روش پیشنهادی با توجه به معیار دقت عملکرد پایین‌تری دارد (۲۱).

خسروانیان و آیت در سال ۲۰۱۵ با استفاده از داده‌های مورد استفاده در همین مقاله با پیاده‌سازی سیستم تصمیم‌یار مبتنی بر شبکه عصبی احتمالی جهت تشخیص نوع سرطان پستان توانستند پارامترهای حساسیت، اختصاصیت و صحت به کمک این سیستم به ترتیب معادل اعداد ۱، ۰/۹۸ و ۰/۹۹ به دست آورند که نسبت به روش پیشنهادی علاوه بر کنار گذاشتن داده‌های مفقودی در مقاله عملکرد پایین‌تری را نشان می‌دهد (۲۲).

در مطالعه‌ای دیگر، اتمو و همکارانش در سال ۲۰۱۴ توانستند با استفاده از روش شبکه هوش مصنوعی مبتنی بر یادگیری ماشین با استفاده از روش‌های ELMANN و BPANN به تشخیص بیماری سرطان پستان بپردازند و در نهایت با مقایسه دقت و صحت نتایج به دست آمده روش ELMANN را نسبت به روش دیگر در مجموع بهتر دانسته‌اند که هر دو روش به کار رفته عملکرد ضعیف‌تری نسبت به روش پیشنهادی را نشان می‌دهد (۲۳). بادویچ و تیواری در سال ۲۰۱۵ توانستند با استفاده از الگوریتم جدیدی از شبکه عصبی بهینه‌سازی شده ژنتیکی GONN به تشخیص بیماری سرطان پستان بپردازند و دقت طبقه‌بندی را با استفاده از روش‌های کلاسیک و BP ۹۸/۲۴، ۹۹/۶۳ و ۱۰۰ درصد برای مقادیر آزمون و آموزش به ترتیب ۵۰-۵۰، ۶۰-۴۰، ۷۰-۳۰

به‌دست آورند که علی‌رغم دقت‌های خوب بدست آمده، روش‌های بکاررفته در پژوهش‌های مشابه انجام و نوآوری در آن دیده نمی‌شد (۲۴).

آسری و همکارانش در سال ۲۰۱۶ با استفاده از الگوریتم یادگیری ماشین‌های مختلف: ماشین بردار پشتیبانی (SVM)، درخت تصمیم‌گیری (C4.5)، Bayes Naive (NB) و K نزدیک‌ترین همسایگان به بررسی بیماری سرطان پستان پرداخته و بهترین دقت (۹۷/۱۳٪) را با استفاده از روش ماشین بردار پشتیبان به دست آوردند که نسبت به روش پیشنهادی با توجه به معیار دقت عملکرد پایین‌تری دارد (۲۵).

فوگارتی و همکاران در سال ۲۰۰۱ الگوریتم ژنتیک را برای تشخیص سرطان پستان بر روی همین مجموعه داده به کار بردند و پس از مقایسه خروجی با خروجی واقعی به دقت ۹۶/۳۲٪ رسیدند که نسبت به روش پیشنهادی با توجه به معیار دقت عملکرد پایین‌تری دارد (۲۶).

دهقان و همکاران در سال ۲۰۱۸ با استفاده از داده‌های مورد استفاده در همین مقاله و با روش‌های شبکه‌های عصبی پرسپترون چند لایه و شبکه عصبی LVQ و بیزین با میانگین ۱۰ بار تست به ترتیب با دقت‌های ۹۷/۵ و ۹۷/۶ و ۹۸/۳ درصد سرطان پستان را پیش‌بینی کرده و بررسی‌های مطالعه آنها نشان داد که شبکه عصبی بیزین در تشخیص بیماری موفق‌تر است که نسبت به روش پیشنهادی علاوه بر کنار گذاشتن داده‌های مفقودی در مقاله عملکرد پایین‌تری را با توجه به معیار دقت نشان می‌دهد (۲۷).

جدول ۲ ادامه مقایسه روش پیشنهادی با روش‌های دیگر با ذکر پایگاه داده مورد استفاده است که نتایج این مقایسه نشان می‌دهد که مدل پیشنهادی در جهت پیش‌بینی تومور بدخیم پستان توانسته است نسبت به سایر روش‌ها موفق‌تر عمل کرده و روش پیشنهادی بالاترین دقت را دارد.

از محدودیت‌های پژوهش‌های پیشین می‌توان به حذف مواردی اشاره کرد که در پایگاه داده دارای ویژگی‌های ناقص بودند. علاوه بر آن در پژوهش‌های پیشین که در مجموعه داده‌های خود اطلاعات افراد سالم را در دسترس نداشته‌اند عنوان مقاله خود را تشخیص سرطان پستان قید کرده‌اند که در این مقاله به‌درستی از عنوان پیش‌بینی بدخیمی تومور پستان یاد شده است.

جدول ۲: مقایسه دقت داده‌ها در پیاده‌سازی‌های مختلف در تشخیص سرطان پستان

روش	دقت	سال	پایگاه داده مورد استفاده	نویسندگان و منبع
SVM و RBF	۹۸/۱	۲۰۰۷	مجموعه داده‌های بیماری سرطان پستان پایگاه WDBC	ایران پور و همکاران (۲۸)
RBF	۹۳/۶۷	۲۰۱۱	مجموعه داده‌های بیماری سرطان پستان پایگاه UCI	آرانا و همکاران (۲۹)
درخت تصمیم	۹۲/۹۷			
روش بیزین	۹۲/۶۱			
روش بیزین	۹۲/۹۷	۲۰۱۲	مجموعه داده‌های بیماری سرطان پستان پایگاه WPBC+WDBC+UCI	سالاما و همکاران (۳۰)
درخت تصمیم	۹۳/۱۵			
درخت تصمیم‌گیری	۹۱/۳	۲۰۱۶	Mortaz General Hospital	لطیف و همکاران (۳۱)
نایو بیز	۹۲/۹			
نزدیک‌ترین همسایه	۹۵/۱			
الگوریتم ژنتیک و داده کاوی	۹۷/۳			
ماشین بردار پشتیبان	۹۶/۷	۲۰۰۹	پایگاه داده DDSM	لند و ورهگن (۳۲)
RBF	۹۶/۱۸	۲۰۱۱	مجموعه داده‌های بیماری سرطان پستان پایگاه WDBC	کیان و همکاران (۳۳)
MLP	۹۵/۷۴			
SVM	۹۶/۴	۲۰۱۳	مجموعه داده‌های بیماری سرطان پستان پایگاه UCI	چاراسیا و همکاران (۳۴)
درخت تصمیم	۹۴/۸۴	۲۰۱۲	مجموعه داده‌های بیماری سرطان پستان پایگاه UCI	لاوانیا و همکاران (۳۵)
الگوریتم K-نزدیکترین همسایه	۹۶	۲۰۱۴	بیمارستان ولی عصر (عج)	محمودی و همکاران (۳۶)
شبکه‌های عصبی	۹۵/۱			
شبکه بیزی	۹۵/۹			
درخت تصمیم‌گیری	۹۳/۴			
MLP	۹۷/۵	۲۰۱۸	مجموعه داده‌های بیماری سرطان پستان پایگاه UCI	دهقان و همکاران (۲۷)
LVQ	۹۷/۶			
روش بیزین	۹۸/۳			
الگوریتم EA، WOA و MLP	۹۹/۴	۲۰۱۹	مجموعه داده‌های بیماری سرطان پستان پایگاه UCI	پژوهش حاضر

نتیجه‌گیری

از دلایل بالا بودن حساسیت و اختصاصیت در مقاله حاضر می‌توان به یافتن داده‌های مفقودی، پیش پردازش داده‌های ورودی و انتخاب ویژگی توسط الگوریتم فراابتکاری و انتخاب مناسب شبکه عصبی برای این منظور اشاره کرد. در واقع در صورت پیاده‌سازی شبکه عصبی پرسپترون چند لایه بدون انجام پیش‌پردازش‌های اولیه، انتخاب ویژگی و بهینه‌سازی در این مقاله، دقت پایین‌تری را گزارش می‌کرد. شبکه پیاده‌سازی شده در این مقاله به خاطر سرعت زیاد و تعمیم‌پذیری خوب آن نسبت به دیگر روش‌ها برتر است و به عنوان یک روش غیرتهاجمی به عنوان دستیار پزشک در مراکز درمانی و تشخیص زودهنگام سرطان پستان پیشنهاد می‌شود. در صورت طراحی نرم‌افزار با بهره‌گیری از الگوریتم فراابتکاری پیشنهادی در مراکز درمانی می‌توان گامی مؤثر در جهت

تشخیص بیماری سرطان پستان و کاهش هزینه‌های تشخیص در مراحل اولیه بیماری بدون استفاده از روش‌های تهاجمی برداشت.

تشکر و قدردانی

پژوهش حاضر مستخرج از تحقیقات دکتری تخصصی شیمی آقای علی شریفی به شماره طرح پروپوزال ۳۰۳۳ مصوب شورای تحصیلات تکمیلی با موضوع تشخیص بیماری‌های مختلف با استفاده از روش‌های کمومتریکس و داده‌کاوی می‌باشد که در مورخ ۱۳۹۸/۰۲/۱۵ در دانشکده علوم دانشگاه لرستان ایران به تصویب نهایی رسیده است. در ضمن 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.

تعارض منافع

نویسندگان اعلام می‌دارند که هیچ تعارض منافی در پژوهش حاضر وجود ندارد.

References

1. Fathi T, Khbaz-Zadeh S, Mazloun M. Evaluation of risk factors for breast cancer in women of childbearing age in Mashhad on 2002-2003 years. *Journal of Iran University of Medial Siences*. 2004;42:568-77.
2. Parkin DM, Pisani P, Ferlay J. Estimates of the worldwide incidence of 25 major cancers in 1990. *International journal of cancer*. 1999;80(6):827-41.
3. Milovic B. Prediction and decision making in health care using data mining. *Kuwait chapter of arabian journal of business and management review*. 2012;33(848):1-1.
4. Pendharkar PC, Rodger JA, Yaverbaum GJ, Herman N, Benner M. Association, statistical, mathematical and neural approaches for mining breast cancer patterns. *Expert Systems with Applications*. 1999;17(3):223-32.
5. Sedehi M, Amani F, Momeni DF. Analysis of survival data of patient with breast cancer using artificial neural network and cox regression models. 2014;5(4): 24-32.
6. UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set (cited 2017 May 4). Available from: [http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)).
7. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*. 1977;39(1):1-22.
8. Ghomrawi HM, Mandl LA, Rutledge J, Alexiades MM, Mazumdar M. Is there a role for expectation maximization imputation in addressing missing data in research using WOMAC questionnaire? Comparison to the standard mean approach and a tutorial. *BMC musculoskeletal disorders*. 2011;12(1):109.
9. Downey RG, King CV. Missing data in Likert ratings: A comparison of replacement methods. *The Journal of general psychology*. 1998;125(2):175-91.
10. Dodge Y, Commenges D, editors. *The Oxford dictionary of statistical terms*. Oxford University Press on Demand; 2006.
11. Mirjalili S, Lewis A. The whale optimization algorithm. *Advances in engineering software*. 2016;95: 51-67.
12. Trivedi IN, Pradeep J, Narottam J, Arvind K, Dilip L. Novel adaptive whale optimization algorithm for global optimization. *Indian Journal of Science and Technology*. 2016;9(38): 319-26.
13. B.Yegnanarayana, *Articial Neural Networks*, Prentice-Hall, New Delhi, India, 1999.
14. Jain AK, Mao J, Mohiuddin KM. Artificial neural networks: A tutorial. *Computer*. 1996;29(3):31-44.
15. Anderson JA. *An introduction to neural network*. MIT press. 1995.
16. Wang S. An insight into the standard back-propagation neural network model for regression analysis. *Omega*. 1998;26(1):133-40.
17. Chakraborty S, Ghosh M., Maiti T Tewari A. Bayesian neural networks for bivariate binary data: An application to prostate cancer study. *Statistics in medicine*. 2005;24(23):3645-62.
18. Sadat Hashemi SM, Kazemnejad A, Lucas C. Architect of artificial neural network for modeling of multivariate binary responses and its application to predicting type of unwanted pregnancy. *Tarbiat Modares University*. 2003.
19. Biglarian A, Babaei GHR. Application of ANN in determining important predictors of inpatients mortality after coronary artery bypass graft surgery and its comparison with logistic regression. *MJMS*. 2005;1:21-30.
20. Sedehi M. Design of artificial neural network for modeling of bivariate mixed responses and its application in medical data. Presented for the Degree of PhD in Biostatistics. *Tarbiat Modares University*. 2010.
21. Ahmad F, Isa NA, Hussain Z, Osman MK, Sulaiman SN. A GA-based feature selection and parameter optimization of an ANN in diagnosing breast cancer. *Pattern Analysis and Applications*. 2015;18(4):861-70.
22. Khosravian A, Ayat S. Implementing a decision support system based on probabilistic

- neural network for diagnosing breast cancer type. 2015;8(3):34-41.
23. Utomo CP, Kardiana A, Yuliwulandari R. Breast cancer diagnosis using artificial neural networks with extreme learning techniques. *International Journal of Advanced Research in Artificial Intelligence*. 2014;3(7):10-4.
 24. Bhardwaj A, Tiwari A. Breast cancer diagnosis using genetically optimized neural network model. *Expert Systems with Applications*. 2015;42(10):4611-20.
 25. Asri H, Mousannif H, Al Moatassime H, Noel T. Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*. 2016;83:1064-9.
 26. Werner JC, Fogarty TC. Genetic programming applied to severe diseases diagnosis. In *Proceedings Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP)*. 2001.
 27. Dehghan P, Mogharabi M, Zabbah I, Layeghi K, Maroosi A. Modeling Breast Cancer Using Data Mining Methods. *Journal of Health and Biomedical Informatics* 2018;4(4):266-78.
 28. 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.
 29. Aruna S, Rajagopalan SP, Nandakishore LV. Knowledge based analysis of various statistical tools in detecting breast cancer. *Computer Science & Information Technology*. 2011;2:37-45.
 30. Salama GI, Abdelhalim M, Zeid MA. Breast cancer diagnosis on three different datasets using multi-classifiers. *Breast Cancer (WDBC)*. 2012;32(569):2.
 31. Latif AM, Momeny M, Sarram R, Agha Srram M, Pour Ahmadi A, Haj Ebrahimi Z. Using data mining and genetic algorithm for diagnosis of breast cancer. *ijbd*. 2016;9(1):45-56.
 32. Land Jr WH, Verheggen EA. Multiclass primal support vector machines for breast density classification. *International journal of computational biology and drug design*. 2009;2(1):21-57.
 33. Kiyan T, Yildirim T. Breast cancer diagnosis using statistical neural networks. *IU-Journal of Electrical & Electronics Engineering*. 2004;4(2):1149-53.
 34. Chaurasia S, Chakrabarti P. An Approach with Support Vector Machine using Variable Features Selection on Breast Cancer Prognosis. *International Journal of Advanced Research in Artificial Intelligence*. 2013;2(9):38-42.
 35. Lavanya D, Rani KU. Ensemble decision tree classifier for breast cancer data. *International Journal of Information Technology Convergence and Services*. 2012;2(1):17.
 36. Mahmoodi MS, Mahmoodi SA, Haghighi F, Mahmoodi SM. Determining the stage of breast cancer by data mining algorithms. *ijbd*. 2014;7(2):36-44.