



Improving Diagnosis of Breast Cancer Disease Using Adaptive Neuro-fuzzy Inference System

Ali Maroosi^{1*}, Iman Zabbah², Maedeh Mogharebi³, Ehsan Yasrebi⁴, Kamran Layeghi⁵

¹Assistant Professor, Department of Computer Engineering, Faculty of Technical and Engineering, University of Torbat Heydarieh, Torbat Heydarieh, Iran.

²Faculty Member, Department of Computer Engineering, Torbat Heydariyeh Branch, Islamic Azad University, Torbat Heydariyeh, Iran.

³M.Sc, Department of Computer Engineering, Faculty of Technical and Engineering, University of Torbat Heydarieh, Torbat Heydarieh, Iran.

⁴PhD Student, Department of Computer Engineering, Faculty of Technical and Engineering, University of Torbat Heydarieh, Torbat Heydarieh, Iran.

⁵Assistant Professor, Department of Computer Engineering, North Tehran Branch, Islamic Azad University, Tehran, Iran.

ARTICLE INFO

ABSTRACT

Article Type:

Original Research

Received: 05.31.2021

Revised: 08.31.2021

Accepted: 01.03.2022

Keyword:

Breast Cancer
Data Mining
Fuzzy Clustering
Neural Network
Neuro-fuzzy

*Corresponding Author:

Ali Maroosi

Email: ali.maroosi@torbath.ac.ir

Cancer and particularly breast cancer is one of the most common diseases among women worldwide. Early detection of breast cancer is a major challenge for physicians and is key in successful treatment and patient survival. This study introduces some data mining methods for the prediction of breast cancer based on a dataset containing 683 independent records with 9 features from the UCI machine learning repository. The models were used to diagnose benign and malignant breast cancer. Results showed that the accuracy of Multi-Layer Perceptron Neural Network (MLP), Learning Vector Quantization (LVQ) Neural Network, Radial Basis Function (RBF), Fuzzy Clustering (KFC), Adaptive Neuro-Fuzzy Inference System Model (ANFIS) were 97.5%, 97.5%, 98.3%, 75% and 99.2%, respectively. Early diagnosis of breast cancer disease reduces the cost of treatment and increases the chance of successful treatment. This study demonstrated that neuro-fuzzy inference system performed better than other models for breast cancer diagnosis. In this study, while diagnosing breast cancer, it was illustrated that models based on fuzzy neural inference had a more acceptable performance than other methods in diagnosing breast cancer. The proposed model can assist the medical community, particularly mammography specialists.



EXTENDED ABSTRACT

Introduction

One of the most common cancers among women is breast cancer, affecting approximately 10% at different stages of life. In this disease, malignant (cancerous) cells are formed in the breast tissue. Although the prevalence of this disease is rare before the age of 25 to 30 years, the occurrence of this cancer at younger ages has been reported, making breast cancer to be one of the most common cancers among women. Early diagnosis of this disease is one of the greatest challenges for physicians and plays a very key role in successful treatment and patient longevity.

Methodology

The present research was a diagnostic study, categorizing the state of breast cancer in terms of benign or malignant based on the input variables. The data used in this research was obtained from the dataset of breast cancer patients of Wisconsin Hospital, available in the California Machine Learning Database. In this research, models were presented based on data mining so that they have the ability to predict breast cancer. Its database consists of 683 records with 9 clinical variables collected from breast cancer patients at Wisconsin Hospital which is available in the UCI repository dataset.

Table 1 shows breast cancer diagnosis variables and their ranges, descriptive statistics and the Pearson Correlation Coefficients stated for 9 variables.

Table.1. The features of breast cancer diagnosis and their characteristics.

Feature number	Feature	Range	Mean± standard deviation	Correlation coefficient
1	Clump Thickness	1-10	4.44±2.82	0.715
2	Uniformity of Cell Size	1-10	3.65±3.15	0.821
3	Uniformity of Cell Shape	1-10	3.22±2.98	0.822
4	Marginal Adhesion	1-10	3.83±2.86	0.706
5	Single Epithelial Cell Size	1-10	3.23±2.22	0.691
6	Bare Nucleoli	1-10	3.64±3.54	0.823
7	Bland Chromatin	1-10	3.45±2.45	0.758
8	Normal Nucleoli	1-10	3.053±2.87	0.719
9	Mitoses	1-10	1.73±1.60	0.423

In this study, three different types of neural network including multilayer perceptron (MLP) neural network, learning vector neural network (LVQ) and radial basis neural network (RBF) were investigated with the aim of classifying breast cancer. Another suitable method in predicting whether breast cancer is benign or malignant is the use of the neuro-fuzzy system or the adaptive fuzzy system, which was used for analysis and investigation in this study. Fuzzy clustering was also used as an unsupervised learning method. The main purpose of clustering is to identify common structures in the data to create clusters. In classical clustering, each sample belongs to only one cluster (for example, malignant cancer cluster) and cannot be a member of two or more clusters. However, in fuzzy clustering, one sample can belong to two or more clusters with different degrees. Finally, in this article, the

diagnosis of breast cancer was made based on two classes, benign and malignant, and the mean square error criterion was used to evaluate the models.

Results and discussion

In the present study, 683 patients with breast cancer were investigated. The total clinical variables of the patients included nine risk factors. In all models, 80% of the samples (547 samples) were used for training the networks, 10% (55 cases) were used for validation, and the remaining data (56 cases) were used for testing the networks. In the current research, two methods of supervised and unsupervised data mining were evaluated. In the supervised method, three types of artificial neural networks were used. Furthermore, neural-fuzzy network was studied in the second part of the research. The best-obtained accuracy of diagnosing whether cancer was benign or malignant for the multilayer perceptron neural networks (MLP), learning vector neural network (LVQ), radial basis neural network (RBF), unsupervised k-means fuzzy clustering (KFC), and adaptive neuro-fuzzy inference system (ANFIS) were 97.5%, 97.6%, 98.3%, 75%, and 99.2%, respectively.

Conclusion

One of the biggest issues in medical science is disease diagnosis based on various clinical tests, which sometimes makes the final diagnosis difficult for expert physicians. Data mining discovers hidden information in big data, which sometimes remains hidden from the view of medical science specialists. Various methods have been used in the field of prediction, survival, and recurrence of breast cancer patients, and sometimes the obtained results have been in support of physician's decisions. Since the health of women in a society has a direct role in the health of the family and the mental health of society, a timely diagnosis of breast cancer increases the chances of successful treatment while reducing the patient's treatment costs. The results of the study show that the use of the neural-fuzzy method had a better performance in modeling breast cancer than the artificial neural network method and also compared to the fuzzy classification method. In this study, multilayer perceptron neural network (MLP), learning vector neural network (LVQ), radial basis function network (RBF), as well as unsupervised method of k-means fuzzy clustering (KFC) and adaptive neuro-fuzzy inference system (ANFIS) were used in order to classify and model breast cancer patients. Furthermore, the results indicate that although different data mining methods can predict disease, according to the powerful concept of fuzzy, it is possible to use a neural-fuzzy network by combining neural network and fuzzy logic. Fuzzy clustering was used to discover which sample relatively belonged to the benign or malignant cancer cluster.

بهبود تشخیص بیماری سرطان پستان با استفاده از سیستم استنتاج عصبی - فازی تطبیقی

علی ماروسی^{۱*}، ایمان ذباح^۲، مائده مقربی^۳، احسان یثربی^۴، کامران لایقی^۵

- ۱- استادیار، گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه تربت حیدریه، تربت حیدریه، ایران.
- ۲- عضو هیأت علمی، گروه کامپیوتر، واحد تربت حیدریه، دانشگاه آزاد اسلامی، تربت حیدریه، ایران.
- ۳- کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه تربت حیدریه، تربت حیدریه، ایران.
- ۴- دانشجوی دکترا، گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه تربت حیدریه، تربت حیدریه، ایران.
- ۵- استادیار، گروه مهندسی کامپیوتر، واحد تهران شمال، دانشگاه آزاد اسلامی، تهران، ایران.

چکیده

اطلاعات مقاله

امروزه سرطان یکی از مهم‌ترین عوامل مرگ‌ومیر در دنیا است و سرطان پستان از شایع‌ترین آن‌ها در میان زنان می‌باشد. تشخیص زودهنگام این بیماری، یکی از معضلات مهم پزشکان می‌باشد و نقش بسیار کلیدی در درمان موفق و حیات بیمار ایفا می‌کند. در این پژوهش مدل‌های مبتنی بر داده‌کاوی، ارائه شده و پایگاه داده آن مشتمل بر ۶۸۳ رکورد با ۹ متغیر بالینی است که از بیماران مبتلا به سرطان پستان بیمارستان ویسکانسین جمع‌آوری شده است و در مرجع داده UCI قابل دسترس است. تشخیص سرطان پستان در دو خوشه خوش‌خیم و بدخیم صورت گرفته است. در بهترین حالت خوش‌خیم یا بدخیم بودن سرطان با روش‌های نظارت‌شده شبکه‌های عصبی پرسپترون چندلایه (MLP)، شبکه عصبی بردار یادگیر (LVQ)، شبکه عصبی پایه شعاعی (RBF)، خوشه‌بندی بدون نظارت (KFC) و همچنین عصبی- فازی (ANFIS) به ترتیب دقت‌های ۹۷/۵ درصد، ۹۷/۶ درصد، ۹۸/۳ درصد، ۷۵ درصد، ۹۹/۲ درصد به دست آمد. تشخیص به‌موقع سرطان پستان، ضمن کاهش هزینه‌های درمانی بیمار، شانس درمان موفقیت‌آمیز را افزایش می‌دهد. در این مطالعه ضمن تشخیص بیماری سرطان پستان، نشان داده شده است که مدل‌های مبتنی بر استنتاج عصبی- فازی، عملکرد قابل‌قبول‌تری نسبت به سایر روش‌های مورد مطالعه، در تشخیص بیماری سرطان پستان دارد. مدل ارائه شده می‌تواند به‌عنوان دستیار پزشک در خدمت جامعه پزشکی قرار گیرد.

نوع مقاله: مقاله پژوهشی

دریافت مقاله: ۱۴۰۱/۰۳/۱۰

بازنگری مقاله: ۱۴۰۱/۰۶/۰۹

پذیرش مقاله: ۱۴۰۱/۱۰/۱۳

کلید واژگان:

سرطان پستان
داده‌کاوی
خوشه‌بندی فازی
شبکه عصبی
عصبی- فازی

*نویسنده مسئول: علی ماروسی

پست الکترونیکی:

ali.maroosi@torbath.ac.ir

مقدمه

یکی از شایع‌ترین سرطان‌ها در میان زنان، سرطان پستان است که حدود ۱۰ درصد از آنان را در مراحل مختلف زندگی، تحت تأثیر قرار می‌دهد [۱]. در این بیماری، سلول‌های بدخیم (سرطانی) در بافت پستان تشکیل می‌شوند. اگرچه شیوع این بیماری در سنین قبل از ۲۵ تا ۳۰ سالگی نادر است اما بروز این سرطان در سنین کمتر حتی در سن جوانی نیز گزارش شده است [۲]. این مسئله موجب شده است که حدود ۴/۴ میلیون زن در جهان به سرطان پستان مبتلا باشند. این سرطان یکی از رایج‌ترین سرطان‌هایی است که زنان، بعد از سرطان پوست به آن مبتلا می‌شوند. در سال ۲۰۱۳ تقریباً ۲۳۲۰۰۰ مورد جدید سرطان پستان در میان زنان آمریکایی تشخیص داده شده است [۳]. بر اساس اطلاعات ملی ثبت سرطان و آمار مرگ‌ومیر گزارش شده توسط وزارت بهداشت، درمان و آموزش پزشکی کشور میزان مرگ‌ومیر کلی سرطان پستان در طول سال‌های تحت مطالعه از ۰/۹۶ به ۴/۳۳ در هر ۱۰۰۰۰۰ مورد و بروز از ۱۶ به ۲۸/۳ در هر ۱۰۰۰۰۰ مورد افزایش داشته است. در ایران سرطان پستان اولین نوع سرطان تشخیص داده شده در میان زنان است که ۲۴/۴ درصد از همه انواع بدخیمی‌ها را به خود اختصاص می‌دهد [۴]. تمام تومورها، سرطانی نیستند و ممکن است خوش‌خیم یا بدخیم باشند. تومورهای خوش‌خیم، رشد غیرطبیعی دارند ولی به‌ندرت مرگ‌آور هستند. با این‌حال، تعدادی از توده‌های خوش‌خیم پستان نیز می‌توانند خطر ابتلا به سرطان پستان را افزایش دهند. همچنین در برخی از زنان با سابقه نمونه‌برداری از توده‌های خوش‌خیم پستان نیز خطر سرطان پستان افزایش یافته است اما تومورهای بدخیم، جدی‌تر هستند و سرطانی محسوب می‌شوند. تشخیص زودهنگام این نوع از سرطان‌ها احتمال درمان موفقیت‌آمیز را بالا می‌برد [۵]. با تشخیص زودهنگام این بیماری و پیشرفت‌های به‌دست‌آمده در درمان، میزان بقای بیماران مبتلا به آن در حال افزایش است. تشخیص به‌موقع سرطان پستان (حداکثر ۵ سال پس از اولین تقسیم سلول سرطانی) احتمال زنده بودن بیمار سرطانی را از ۵۶ درصد به ۸۶ درصد افزایش می‌دهد [۶]. عامل‌های خطر^۱ ابتلا به سرطان پستان شامل سن بالا، سابقه فردی ابتلا به سرطان پستان، تغییرات در ژن، نژاد، پرتودرمانی قفسه سینه، تراکم پستان، مصرف مشروبات الکلی و دخانیات، زمان اولین قاعدگی قبل از ۱۲ سالگی و شروع یائسگی بعد از ۵۵ سالگی، زمان بارداری یا عدم بارداری، اضافه‌وزن و چاقی بعد از یائسگی و زندگی بی‌تحرک است [۷]. اخیراً مطالعات متعددی در مورد تشخیص خوش‌خیم یا بدخیم بودن سرطان پستان صورت گرفته است [۸]. برخی از این مطالعات گذشته بر روش‌های آماری استوار بوده است؛ مانند مطالعه ای که از روش تجزیه و تحلیل گسسته خطی در تشخیص سرطان پستان بر روی مجموعه داده‌های ویسکانسین استفاده شده است [۹]. اما اخیراً، تعداد وسیعی از مطالعات متکی بر روش‌های هوش مصنوعی بوده است. برای مثال روزبهانی و همکاران برای بررسی الگوهای موجود در سرطان پستان از چندین روش داده‌کاوی استفاده کردند. در این مطالعه نشان دادند که داده‌کاوی می‌تواند با هدف تشخیص، پیش‌آگهی و درمان بیماری، به‌عنوان یک ابزار ارزشمند در شناسایی شباهت‌ها در مورد سرطان پستان به‌کار رود [۱۰]. توجه به شبکه‌های عصبی مصنوعی به‌عنوان یک سیستم قوی تشخیص الگو در بسیاری از مطالعات حوزه پزشکی دیده می‌شود [۱۱]. فوگن و همکاران برای کشف الگوهای تشخیص سرطان پستان بر روی مجموعه داده بیمارستان ویسکانسین، از ماشین بردار پشتیبان استفاده کردند. نتایج به‌دست‌آمده نشان داد که SVM برای تشخیص الگوهای سرطان پستان روش مناسبی بوده است [۱۲]. در مطالعه‌ای دیگر کینی و آتشی با استفاده از الگوریتم‌های داده‌کاوی (درخت تصمیم) یک مدل پیش‌آگهی مبتنی بر داده‌کاوی برای پیش‌بینی عود مجدد سرطان پستان با توجه به داده‌های موجود در مرکز تحقیقات سرطان پستان، ایجاد کردند [۱۳]. اگرچه تحقیقات مربوط به سرطان، بیشتر بالینی و بیولوژیکی است اما استفاده از تحقیقات آماری در حوزه پزشکی، جایی که داده‌ها و آمار با موفقیت به‌کاربرده می‌شوند، مسیرهای تحقیقاتی جدید و نوآورانه‌ای را برای کاربرد بیشتر در تحقیقات را ارائه می‌دهد [۱۴]. توجه بیشتر به عوامل بالینی و بیولوژیکی بیماری سرطان پستان و تشخیص

¹ Risk Factor

دقیق‌تر این بیماری حتی به‌اندازه ۱ درصد باعث خواهد شد که شانس بقای بیمار افزایش یابد و روند دوره درمان با سرعت بیشتری طی شود؛ از این رو در این مطالعه که با هدف بهبود تشخیص خوش‌خیم یا بدخیم بودن بیماری سرطان پستان صورت گرفته است، تلاش شده است با استفاده از مدل‌های مختلف داده‌کاوی به‌خصوص شبکه‌های عصبی مصنوعی و شبکه‌های عصبی- فازی در کنار یکدیگر، دقت تشخیص بیماری نسبت به مطالعات مشابه بهبود یابد.

مواد و روش‌ها

تحقیق حاضر حاصل یک مطالعه تشخیصی است که براساس متغیرهای ورودی به تفکیک وضعیت بیماری سرطان پستان از نظر خوش‌خیم یا بدخیم بودن می‌پردازد. داده‌های مورداستفاده در این پژوهش از مجموعه داده مربوط به بیماران مبتلا به سرطان پستان بیمارستان ویسکانسین، موجود در انبار داده یادگیری ماشین کالیفرنیا تأمین شده است [۱۵]. بانک اطلاعاتی این منبع شامل ۶۹۹ نمونه با ۹ پارامتر می‌باشد. در این مجموعه ۱۶ نمونه، ویژگی‌های کامل را ندارند. در نهایت داده‌های این منبع را می‌توان یک ماتریس 9×683 در نظر گرفت. از این تعداد ۶۵ درصد این بیماران یعنی ۴۴۴ نفر مبتلا به سرطان خوش‌خیم و ۳۵ درصد یعنی ۲۳۹ نفر مبتلا به سرطان بدخیم سینه بوده‌اند. جدول ۱ پارامترهای تشخیص بیماری سرطان پستان و دامنه تغییرات هریک از آن‌ها، آماره‌های توصیفی و نیز ضرایب همبستگی پیرسون برای ۹ متغیر ورودی را نشان می‌دهد که بیش‌ترین ضریب همبستگی مربوط به متغیر نوکلئالی لخت با مقدار $0/823$ می‌باشد که این نتیجه‌گیری در مطالعات دیگر نیز مشاهده می‌شود [۱۶].

جدول ۱. متغیرهای ورودی در تشخیص سرطان پستان.

شماره ویژگی	ویژگی	دامنه تغییرات	میانگین و انحراف معیار	ضریب همبستگی	
۱	Clump Thickness	(ضخامت غده)	$10-1$	$2/821 \pm 4/44$	$0/715$
۲	Uniformity of Cell Size	(یکنواختی اندازه سلول)	$10-1$	$3/150 \pm 3/65$	$0/821$
۳	Uniformity of Cell Shape	(یکنواختی شکل سلول)	$10-1$	$3/220 \pm 2/989$	$0/822$
۴	Marginal Adhesion	(چسبندگی لبه‌ها)	$10-1$	$3/83 \pm 2/865$	$0/706$
۵	Single Epithelial Cell Size	(اندازه سلول)	$10-1$	$3/230 \pm 2/223$	$0/691$
۶	Bare Nucleoli	(نوکلئالی لخت)	$10-1$	$3/540 \pm 3/644$	$0/823$
۷	Bland Chromatin	(رنگینه ملایم)	$10-1$	$3/450 \pm 2/450$	$0/758$
۸	Normal Nucleoli	(نوکلئالی نرمال)	$10-1$	$2/870 \pm 3/053$	$0/719$
۹	Mitoses	(تقسیم هسته)	$10-1$	$1/600 \pm 1/733$	$0/423$

همچنین با توجه به اینکه بازه تغییرات هریک از این ۹ عامل خطر بین ۱ تا ۱۰ است، تحلیل آماری هریک از آن‌ها در جدول ۲ محاسبه شده است.

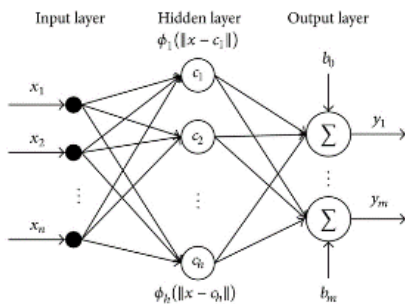
جدول ۲. تحلیل آماری ۹ عامل خطر بیمار سرطان پستان.

چسبندگی لیبه‌ها	نوکلئلای نرمال		رنگینه ملایم		نوکلئلای لخت		اندازه سلول مخاطی منفرد		تقسیم هسته سلول		یکنواختی اندازه سلول		ضخامت غده		یکنواختی شکل سلول																											
	فراوانی	درصد	فراوانی	درصد	فراوانی	درصد	فراوانی	درصد	فراوانی	درصد	فراوانی	درصد	فراوانی	درصد	فراوانی	درصد																										
۸۰/۵	۵۶۳	۶۱/۸	۳۲۲	۲۱/۵	۱۵۰	۲۱/۵	۵۷/۵	۴۰۲	۶/۳	۴۴	۶/۳	۲۷۲	۵۲/۴	۳۹۳	۵۶/۲	۲۷۲	۵۲/۴	۳۹۳	۵۶/۲	۲۷۲	۵۲/۴	۳۹۳	۵۶/۲	۲۷۲	۵۲/۴	۳۹۳	۵۶/۲	۲۷۲	۵۲/۴	۳۹۳	۵۶/۲											
۵	۳۵	۵/۲	۳۶	۵/۲	۱۶۰	۲۲/۹	۴/۳	۳۰	۴/۳	۳۷۶	۵۳/۸	۵۸	۸/۳	۸/۳	۳۷۶	۵۳/۸	۳۰	۴/۳	۳۷۶	۵۳/۸	۵۸	۸/۳	۳۷۶	۵۳/۸	۵۸	۸/۳	۳۷۶	۵۳/۸	۵۸	۸/۳	۳۷۶	۵۳/۸	۵۸	۸/۳	۳۷۶	۵۳/۸						
۴/۷	۳۳	۶	۴۲	۶	۱۶۱	۲۳	۴	۲۸	۴	۷۱	۱۰/۲	۵۸	۸/۳	۸/۳	۷۱	۱۰/۲	۲۸	۴	۷۱	۱۰/۲	۵۸	۸/۳	۵۲	۷/۴	۵۸	۸/۳	۷۱	۱۰/۲	۵۸	۸/۳	۷۱	۱۰/۲	۵۸	۸/۳	۷۱	۱۰/۲	۵۸	۸/۳	۷۱	۱۰/۲		
۱/۷	۱۲	۲/۶	۱۸	۲/۶	۲۹	۵/۶	۲/۷	۱۹	۲/۷	۴۸	۶/۹	۳۳	۴/۷	۴/۷	۴۸	۶/۹	۱۹	۲/۷	۴۸	۶/۹	۳۳	۴/۷	۳۸	۵/۴	۳۳	۴/۷	۴۸	۶/۹	۳۳	۴/۷	۴۸	۶/۹	۳۳	۴/۷	۴۸	۶/۹	۳۳	۴/۷	۴۸	۶/۹		
۰/۹	۶	۲/۷	۱۹	۲/۷	۲۴	۴/۹	۴/۳	۳۰	۴/۳	۳۹	۵/۶	۲۳	۳/۳	۳/۳	۳۹	۵/۶	۳۰	۴/۳	۳۹	۵/۶	۲۳	۳/۳	۳۰	۴/۳	۲۳	۳/۳	۳۹	۵/۶	۲۳	۳/۳	۳۹	۵/۶	۲۳	۳/۳	۳۹	۵/۶	۲۳	۳/۳	۳۹	۵/۶		
۰/۴	۳	۳/۱	۲۲	۳/۱	۹	۱/۳	۱/۳	۲۲	۳/۱	۴۰	۵/۷	۲۱	۳	۳	۴۰	۵/۷	۲۱	۳	۴۰	۵/۷	۲۱	۳	۲۵	۳/۶	۲۱	۳	۴۰	۵/۷	۲۱	۳	۴۰	۵/۷	۲۱	۳	۴۰	۵/۷	۲۱	۳	۴۰	۵/۷		
۱/۳	۹	۲/۳	۱۶	۲/۳	۷۱	۱۰/۲	۱/۱	۸	۱/۱	۱۱	۱/۶	۱۳	۱/۹	۱/۹	۱۱	۱/۶	۱۳	۱/۹	۱۱	۱/۶	۱۳	۱/۹	۱۹	۲/۷	۱۳	۱/۹	۱۱	۱/۶	۱۳	۱/۹	۱۱	۱/۶	۱۳	۱/۹	۱۱	۱/۶	۱۳	۱/۹	۱۱	۱/۶		
۱/۱	۸	۳/۳	۲۳	۳/۳	۲۸	۴	۳	۲۱	۳	۲۱	۳	۲۵	۳/۶	۳/۶	۲۱	۳	۲۵	۳/۶	۲۱	۳	۲۱	۳	۲۸	۴	۲۵	۳/۶	۲۱	۳	۲۱	۳	۲۱	۳	۲۱	۳	۲۱	۳	۲۱	۳	۲۱	۳		
۲	۱۴	۲/۱	۱۵	۲/۱	۲۰	۱/۶	۱/۳	۱۵	۲/۱	۲	۳	۴	۰/۶	۴	۰/۶	۲	۳	۴	۰/۶	۲	۳	۴	۰/۹	۶	۰/۹	۴	۰/۶	۲	۳	۴	۰/۹	۶	۰/۹	۴	۰/۶	۲	۳	۴	۰/۹	۶	۰/۹	
۰	۰	۸/۶	۶۰	۸/۶	۱۱	۲/۹	۱۱	۲/۹	۶۰	۸/۶	۱۱	۲/۹	۱۱	۲/۹	۶۰	۸/۶	۱۱	۲/۹	۶۰	۸/۶	۱۱	۲/۹	۶۷	۹/۶	۵۵	۷/۹	۳۱	۴/۴	۱۳۲	۱۸/۹	۱۱	۲/۹	۶۰	۸/۶	۱۱	۲/۹	۶۰	۸/۶	۱۱	۲/۹	۶۰	۸/۶
۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷	۶۸۳	۹۷/۷

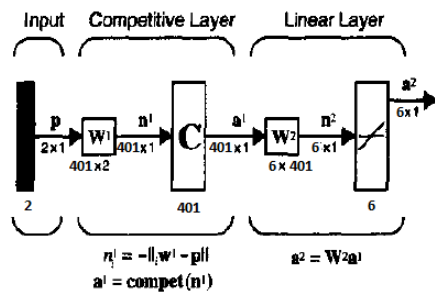
به عبارت دقیق‌تر در مورد ردیف اول جدول، از تعداد ۶۸۳ مورد، ۲۴۶ نفر یعنی معادل ۴۹/۵ درصد، دارای یکنواختی شکل سلول با درجه ۱ بوده‌اند. همچنین ضخامت غده ۱۳۹ نفر یعنی ۱۹/۹ درصد از درجه ۱ می‌باشد. به همین ترتیب برای سایر ویژگی‌ها محاسبات انجام شده است. همان‌طور که در ردیف ۱۰ جدول مشاهده می‌شود فقط ۸/۳ درصد از موارد مورد مطالعه دارای درجه یکنواختی ۱۰ بوده‌اند.

الف- شبکه‌های عصبی مصنوعی

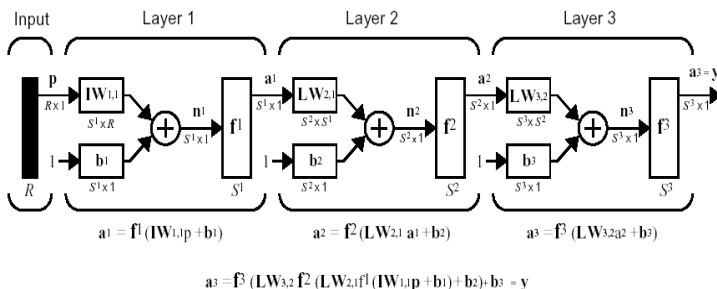
در این پژوهش ابتدا، سه نوع شبکه عصبی مختلف پرسپترون چندلایه (MLP)، شبکه عصبی بردار یادگیر (LVQ) و شبکه عصبی پایه شعاعی (-RBF) با هدف کلاس‌بندی بیماری سرطان پستان بررسی شد. شکل ۱-الف تا ج- ساختار شبکه‌های مورد استفاده در این پژوهش را نشان می‌دهد. از آنجایی که معماری شبکه‌های عصبی می‌تواند نتایج مختلفی به همراه داشته باشد؛ معماری‌های مختلفی بررسی شد. شکل ۲-الف تا ج خطای مربوط به هریک از این معماری‌ها را نشان می‌دهد. همان‌طور که در شکل‌ها نشان داده شده است، محور افقی تعداد نرون‌ها در لایه مخفی هریک از شبکه‌ها و محور عمودی، میزان خطای هریک را نشان می‌دهد.



شکل ۱. ب- ساختار شبکه عصبی RBF

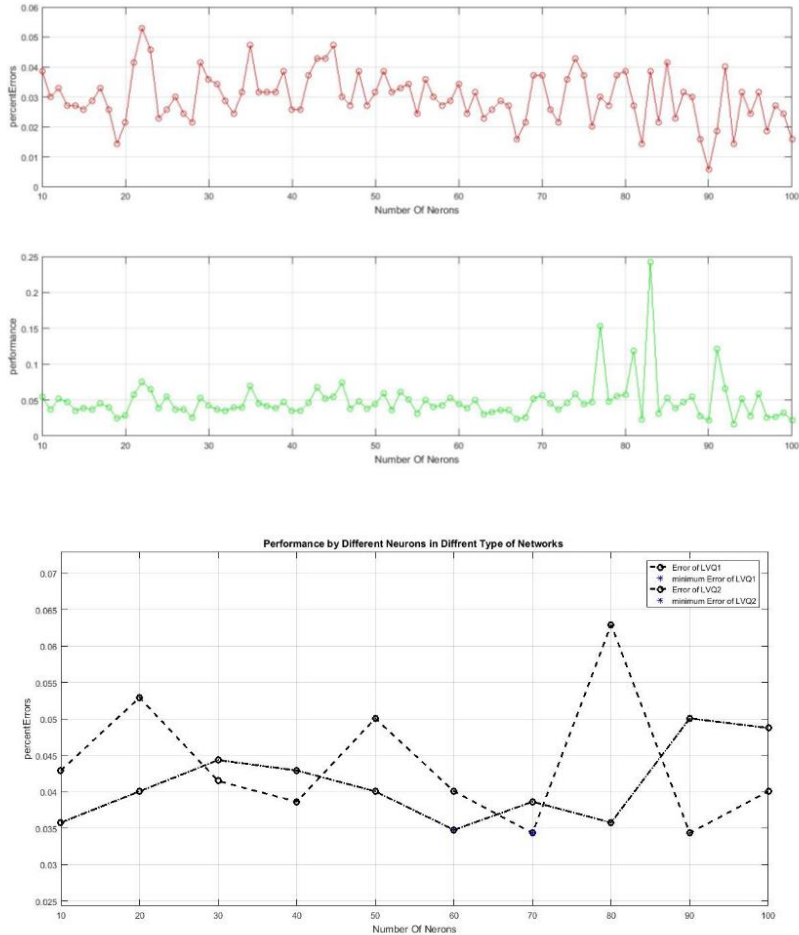


شکل ۱. الف- ساختار شبکه عصبی رقابتی LVQ



شکل ۱. ج- ساختار شبکه عصبی پرسپترون چندلایه

شکل ۱. معماری شبکه‌های عصبی مختلف در تشخیص بیماری سرطان پستان.



شکل ۲. بالا- خطای مربوط به معماری‌های مختلف شبکه عصبی پرسپترون ۲ لایه با تغییر در نرون‌های لایه دوم. شکل ۲- وسط- خطای هریک از شبکه‌های عصبی پایه شعاعی با تغییر تعداد نرون در لایه مخفی. شکل ۲- پایین- خطای مربوط به شبکه عصبی بردار یادگیر با تغییر نرون در لایه رقابت و دو الگوریتم LVQ1 و LVQ2.

ب- شبکه عصبی- فازی

در حال حاضر به دلیل رویکردی که منطق فازی در محاسبات نرم دارد در بسیاری از شاخه‌های علوم اجتماعی و علوم مهندسی و پزشکی مورد توجه پژوهشگران قرار گرفته است [۱۷]. اگرچه شبکه‌های عصبی مصنوعی و سیستم‌های فازی از نظر ساختاری متفاوت هستند اما در واقع این دو دارای ماهیت مکمل نسبت به یکدیگرند. بنابراین، با ترکیب سیستم‌های فازی و شبکه‌های عصبی مصنوعی، توانایی یادگیری وارد مرحله جدیدی می‌شود که سیستم عصبی- فازی^۱ نام دارد. مدل شبکه عصبی- فازی را نخستین بار ژانگ در سال ۱۹۹۳ معرفی کرد. یکی از روش‌های مناسب در پیش‌بینی خوش‌خیم یا بدخیم بودن سرطان پستان استفاده از همین نوع سیستم است که با استفاده از الگوریتم آموزش شبکه

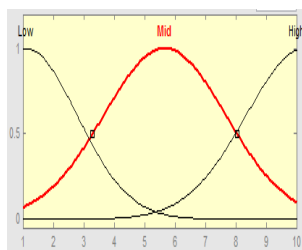
¹ Neuro-Fuzzy model

عصبی، پارامترهای سیستم فازی را تعیین می‌کند. معروف‌ترین نوع سیستم عصبی- فازی، سیستم فازی تطبیق‌پذیر انفیس^۱ نام دارد.

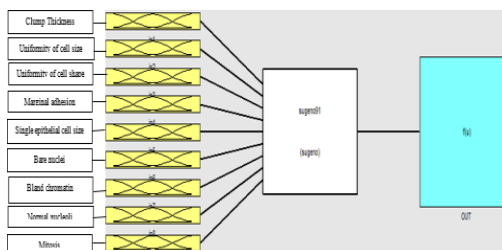
پارامترهای بیماری سرطان پستان طبق جدول ۱ شامل ۹ عامل خطر است؛ بنابراین تعداد ورودی متغیرها در لایه اول ۹ می‌باشد. مدل شبکه عصبی- فازی ارائه‌شده شامل ۵ لایه است که اولین لایه نشان‌دهنده متغیرهای ورودی است. لایه‌های دوم و سوم لایه‌های پنهانی و شامل قواعد فازی است. لایه چهارم، لایه خروجی و در نهایت لایه پنجم، متغیر خروجی نهایی است که نوع سرطان خوش‌خیم یا بدخیم را مشخص می‌کند. به‌منظور آموزش شبکه عصبی- فازی از روش «گرادیان نزولی» و به‌منظور تعیین خطا از روش «حدافل مربعات خطا» استفاده شده است. نحوه آموزش به‌صورت یادگیری با ناظر است. هدف آن است که با آموزش مناسب، شبکه یاد بگیرد و بتواند متناسب با پارامترهای ورودی، خروجی را تعیین کند. شبکه انفیس به‌کار گرفته‌شده در این پژوهش دارای تابع عضویت از نوع خطی^۲ در لایه خروجی و تابع عضویت از نوع گوسین^۳ در لایه ورودی می‌باشد که طبق رابطه ۱ محاسبه می‌شود.

$$f(x, \sigma, c) = \exp\left(\frac{-(x-c)^2}{2\sigma^2}\right) \quad (1)$$

پارامترهای σ و c تعیین‌کننده شکل تابع می‌باشند. شکل ۳-الف سیستم انفیس طراحی شده را نشان می‌دهد. توابع عضویت برای داده‌های خروجی و داده‌های ورودی تعریف می‌شوند. در این مطالعه برای هر یک از ویژگی‌ها یک متغیر زبانی بین ۲ تا ۶ تابع عضویت در نظر گرفته شده است. برای نمونه شکل ۳-ب متغیر Bare Nucleoli با ۳ تابع عضویت «کم»، «متوسط» و «زیاد» تعریف شده است. برای سایر متغیرها نیز به همین ترتیب توابع عضویت ساخته شده است. شکل ۳-ج، تعداد نرون‌ها و تعداد لایه‌های شبکه طراحی شده را نشان می‌دهد.



شکل ۳.ب- تخصیص متغیر زبانی به متغیر Bare Nucleoli

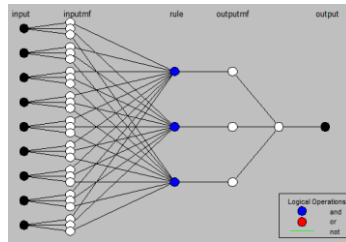


شکل ۳.الف- سیستم انفیس طراحی شده به همراه توابع عضویت

¹ Adaptive Neuro-Fuzzy Inference System (ANFIS)

² Linier

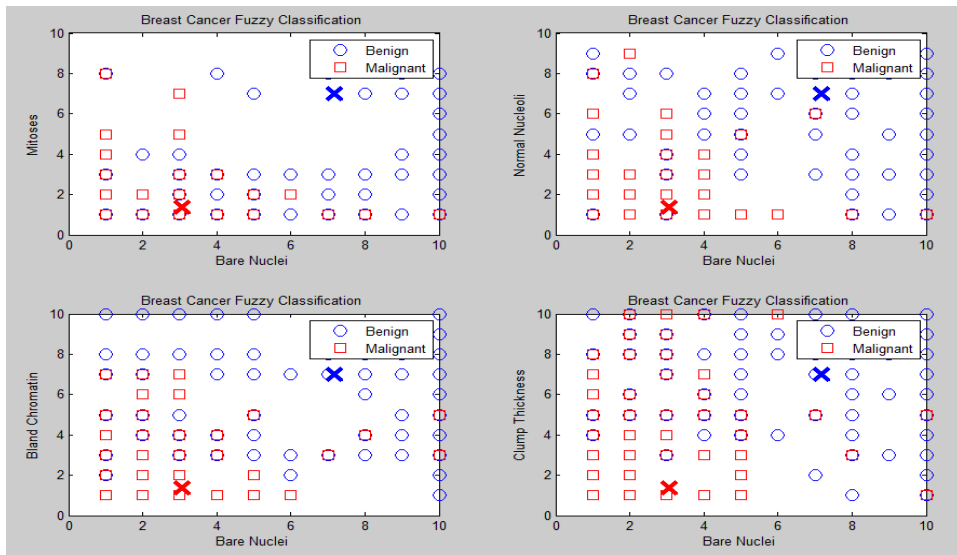
³ Gaussian



شکل ۳. ج- تعداد نرون‌ها و تعداد لایه‌های شبکه عصبی - فازی به منظور تشخیص سرطان پستان.

ج- خوشه‌بندی فازی (KFC)

در بخش سوم این مطالعه، از خوشه‌بندی فازی به‌عنوان روش یادگیری بدون ناظر استفاده شده است. هدف اصلی از خوشه‌بندی، شناسایی ساختارهای مشترک در داده‌ها، برای ایجاد خوشه می‌باشد. در خوشه‌بندی کلاسیک هر نمونه ورودی فقط و فقط متعلق به یک خوشه می‌باشد (برای مثال خوشه سرطان بدخیم) و نمی‌تواند عضو دو خوشه یا بیشتر باشد. اما در خوشه‌بندی فازی یک نمونه می‌تواند با درجات مختلف به دو یا چند خوشه تعلق داشته باشند. با توجه به ضریب همبستگی پیرسون مندرج در جدول ۲ در بین ۹ متغیر، متغیر *Bare Bare Nucleoli* دارای بیشترین ضریب همبستگی است و ۴ ویژگی دیگر براساس آن ترسیم شده‌اند. شکل ۴ پراکندگی متغیر *Bare Bare Nucleoli* براساس متغیرهای *Normal Nucleoli* و *Mitoses* و *Clump Thickness* و *Bland Chromatin* را نشان می‌دهد. این نتایج فقط یک دید بصری برای تحلیل در اختیار ما قرار می‌دهد. اما از آنجایی که کلاس‌بندی فازی با تمام پارامترها و محاسبه فاصله اقلیدسی آن‌ها انجام می‌شود عملاً ترسیم آن‌ها امکان‌پذیر نخواهد بود. در تمام موارد فوق علامت X به معنی مرکز خوشه سرطان خوش خیم و بدخیم می‌باشد.



شکل ۴. میزان پراکندگی متغیر Bare Nucleoli نسبت به متغیرهای Clump ، Mitoses ، Normal Nucleoli و Bland Chromatin و Thickness

یافته‌ها

در این مطالعه ۶۸۳ بیمار مبتلا به سرطان پستان که اطلاعات آن از بیمارستان ویسکانسین موجود در مخزن داده یادگیری ماشین UCI جمع‌آوری شده است بررسی شدند. مجموع متغیرهای بالینی بیماران شامل ۹ عامل خطر بود. در تمام مدل‌ها، از ۸۰ درصد نمونه‌ها (۵۴۷ نمونه) برای آموزش شبکه‌ها و از ۱۰ درصد آن‌ها (۵۵ مورد) برای اعتبارسنجی و از باقی‌مانده داده‌ها (۵۶ مورد) برای آزمون شبکه استفاده شده است. در پژوهش حاضر به‌طور کلی دو روش داده‌کاوی نظارت‌شده^۱ و غیرنظارت‌شده^۲ ارزیابی گردید. در روش با ناظر، ۳ نوع شبکه عصبی مصنوعی استفاده شد. همچنین شبکه عصبی-فازی در بخش دوم پژوهش، مطالعه گردید. هر یک از ویژگی‌های تشخیص بیماری سرطان پستان تحلیل شد و براساس آنها مدل‌سازی انجام گردید. به این ترتیب که هر یک از ویژگی‌ها به تنهایی و نیز ترکیب آن‌ها در ساختار شبکه عصبی-فازی بررسی شد و مقدار خطا برای هر یک ثبت گردید. نتایج حاکی از آن است که وقتی تعداد پارامترهای تصمیم‌گیر در تشخیص بیماری افزایش یابد میزان خطای سیستم کاهش می‌یابد؛ به طوری که وقتی فقط از دو ویژگی در تشخیص بیماری سرطان استفاده می‌شود خطای تشخیص حدود ۰/۳ است در حالی که وقتی از تمامی ویژگی‌ها در تشخیص بیماری استفاده می‌شود خطا به حدود ۰/۰۸ کاهش می‌یابد. این نتایج به‌طور کامل تری (براساس تأثیر ترکیب ویژگی‌های مختلف) در جدول ۳ نمایش داده شده است. جدول ۴ صحت عملکرد روش‌های مختلف داده‌کاوی با ۹ ویژگی (براساس ویژگی‌های مندرج در جدول ۱) و تعداد توابع عضویت مختلف را نشان می‌دهد. جدول ۵ نیز صحت عملکرد هر کدام از شبکه‌های عصبی و خوشه‌بندی فازی و عصبی-فازی پس از میانگین ۱۰ بار آزمون را نشان می‌دهد.

جدول ۳. معماری‌های مختلف شبکه عصبی-فازی مورد استفاده.

مقدار خطا	تعداد توابع عضویت	شماره ساختار
۰/۳۰۶۱	۴	۱
۰/۲۵۳۰	۵	۲
۰/۲۱۷۰	۶	۳
۰/۲۰۸۴	۸	۴
۰/۱۶۸۳	۱۰	۵
۰/۰۵۲۴	۲۰	۶
۰/۰۱۸۱	۳۰	۷
۰/۰۰۸۴	۴۰	۸

در جدول ۳، منظور از شماره ساختار، تعداد توابع فازی است که در مدل استفاده شده است.

جدول ۴. بررسی تأثیر ترکیب ویژگی‌های بیماری سرطان پستان، در دقت مدل. نوع تابع عضویت ورودی

gaussmf و خروجی Linear.

¹ Supervised Learning

² Unsupervised Learning

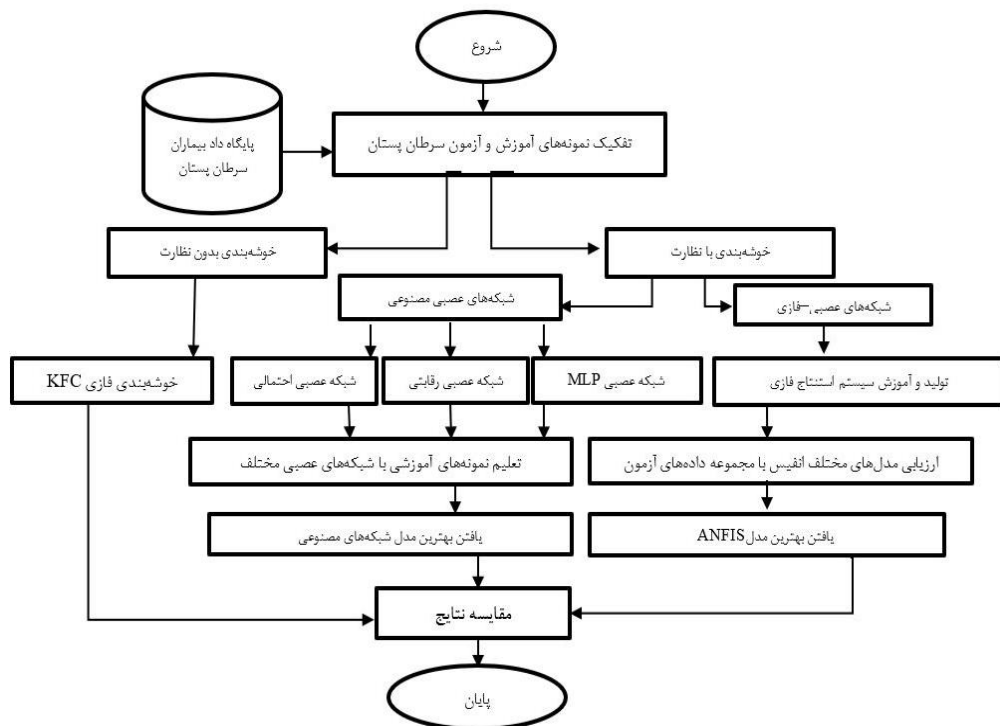
خطا	تعداد توابع عضویت	تکرار	ترکیب ویژگی‌های بیماری
۰/۶۳۷	۵	۳۰	ویژگی اول
۰/۳۳۴	۵	۳۰	ویژگی اول و دوم
۰/۱۸۷	۵	۳۰	ویژگی اول تا سوم
۰/۱۵۰	۳	۱۰	ویژگی اول تا چهارم
۰/۰۵۴۱	۳	۵	ویژگی اول تا پنجم
۰/۰۰۸	۳	۱۰	ویژگی اول تا نهم

جدول ۵. مقایسه صحت عملکرد مدل‌های مختلف داده‌کاوی پس از میانگین ۱۰ بار آزمون.

صحت عملکرد	نوع روش داده‌کاوی
۹۷.۵٪	MLP
۹۷.۵٪	LVQ
۷۵٪	KFC
۹۸.۳٪	RBF
۹۹.۲٪	ANFIS

بحث و نتیجه‌گیری

یکی از بزرگ‌ترین مشکلات در علم پزشکی، تشخیص بیماری، براساس آزمایش‌های مختلف روی بیمار است که گاهی تشخیص نهایی را برای یک پزشک متخصص دشوار می‌کند. علم داده‌کاوی، به کشف الگوهای پنهان در حجم عظیمی از داده‌ها می‌پردازد که گاه از دید متخصصان علوم پزشکی، پنهان می‌مانند. در این میان روش‌های مختلفی در زمینه پیش‌بینی، بقا و عود بیماران مبتلا به سرطان پستان، به‌کاررفته‌اند که گاهی نتایج حاصل از آن‌ها به‌طور باورنکردنی پشتیبان تصمیمات پزشکان بوده است. از آنجایی که سلامت زنان یک جامعه، نقش مستقیمی در سلامت خانواده و سلامت روانی جامعه دارد - [۱۸]؛ توجه به بیماری‌های آنان نیز از اهمیت بالاتری برخوردار است. نتیجه مطالعه نشان می‌دهد که استفاده از روش عصبی-فازی نسبت به روش شبکه عصبی مصنوعی و همچنین نسبت به روش کلاس‌بندی فازی عملکرد بهتری در مدل‌سازی بیماری سرطان پستان دارد. در مطالعه حاضر، از شبکه عصبی پرسپترون چندلایه (MLP)، شبکه عصبی بردار یادگیر (LVQ)، شبکه پایه شعاعی (RBF)، همچنین روش بدون ناظر خوشه‌بندی فازی (KFC) و شبکه‌های عصبی-فازی (ANFIS) به‌منظور کلاس‌بندی و مدل‌سازی بیماران سرطان پستان استفاده گردید. نتایج نشان می‌دهد که اگرچه روش‌های مختلف داده‌کاوی می‌توانند به پیش‌بینی بیماری بپردازند اما با توجه به مفهوم قدرتمند فازی، می‌توان با ترکیب شبکه عصبی و منطق فازی از شبکه عصبی-فازی بهره گرفت، همچنین می‌توان از خوشه‌بندی فازی برای تشخیص اینکه کدام نمونه به‌طور نسبی به خوشه سرطان خوش‌خیم یا بدخیم تعلق دارد پی برد. با استفاده از روش‌های نوین داده‌کاوی می‌توان مدل‌هایی را تولید کرد که ضمن اینکه دستیار پزشک هستند، قابلیت تولید به‌عنوان یک ابزار فناورانه را دارند و امکان تبدیل شدن به یک محصول تجاری را دارا می‌باشند. الگوریتم مورد استفاده در پژوهش حاضر در شکل ۵ نشان داده شده است.



شکل ۵. الگوریتم ارائه شده در این مطالعه.

تضاد منافع

نویسندگان اعلام می‌دارند که در پژوهش حاضر هیچ‌گونه تضاد منافی وجود ندارد.

تقدیر و تشکر

این مقاله مستخرج از نتایج طرح تحقیقاتی اجرا شده به شماره ابلاغیه ۱۳۹۶/۱۰/۰۴۱۰۶۳ UTH از محل اعتبارات پژوهشی دانشگاه تربت حیدریه می‌باشد که بدین وسیله تشکر و قدردانی می‌گردد.

References

- [1] Dumitrescu, R. G. (2018). Early epigenetic markers for precision medicine. In *Cancer Epigenetics for Precision Medicine*. Humana Press. https://doi.org/10.1007/978-1-4939-8751-1_1
- [2] Chean, J., Chen, C.-j., & Shively, J. E. (2017). ETS transcription factor ELF5 induces lumen formation in a 3D model of mammary morphogenesis and its expression is

- inhibited by Jak2 inhibitor TG101348. *Experimental Cell Research*, 359(1), 62-75. <https://doi.org/10.1016/j.yexcr.2017.08.008>
- [3] Harrington, S., Michener, L. A., Kendig, T., Miale, S., & George, S. Z. (2014). Patient-Reported Upper Extremity Outcome Measures Used in Breast Cancer Survivors: A Systematic Review. *Archives of Physical Medicine and Rehabilitation*, 95(1), 153-162. <https://doi.org/10.1016/j.apmr.2013.07.022>
- [4] Sharifian, A., Pourhoseingholi, M. A., Emadedin, M., Rostami-Nejad, M., Ashtari, S., Hajzadeh, N., Firouzei, S., & Hosseini, J. (2015). Burden of Breast Cancer in Iranian Women is Increasing. *Asian Pacific journal of cancer prevention : APJCP*, 16, 5049-5052. <https://doi.org/10.7314/APJCP.2015.16.12.5049>
- [5] Omata, F., McNamara, K. M., Suzuki, K., Abe, E., Hirakawa, H., Ishida, T., Ohuchi, N., & Sasano, H. (2018). Effect of the normal mammary differentiation regulator ELF5 upon clinical outcomes of triple negative breast cancers patients. *Breast Cancer*, 25(4), 489-496. <https://doi.org/10.1007/s12282-018-0842-z>
- [6] Wang, X., Hu, H., & Liu, H. (2018). RNA binding protein Lin28B confers gastric cancer cells stemness via directly binding to NRP-1. *Biomedicine & Pharmacotherapy*, 104, 383-389. <https://doi.org/10.1016/j.biopha.2018.05.064>
- [7] Amadou, A., Ferrari, P., Muwonge, R., Moskal, A., Biessy, C., Romieu, I., & Hainaut, P. (2013). Overweight, obesity and risk of premenopausal breast cancer according to ethnicity: a systematic review and dose-response meta-analysis. *Obesity Reviews*, 14(8), 665-678. <https://doi.org/10.1111/obr.12028>
- [8] Suto, J., Oniga, S., & Sitar, P. P. (2016, May 10-14). *Comparison of wrapper and filter feature selection algorithms on human activity recognition*. 2016 6th International Conference on Computers Communications and Control, Oradea, Romania.
- [9] Bakker, D. P., Busscher, H. J., van Zanten, J., de Vries, J., Klijnstra, J. W., & van der Mei, H. C. (2004). Multiple linear regression analysis of bacterial deposition to polyurethane coatings after conditioning film formation in the marine environment. *Microbiology*, 150(6), 1779-1784. <https://doi.org/10.1099/mic.0.26983-0>
- [10] Roozbahani, Z., Rezaei Noor, J., Yari Eili, M., & Katanforoush, A. (2017). An Analysis of Gene Expression Variations in Lymphoma, Using a Fuzzy Classification Model. *Health Management & Information Science*, 4(1), 1-6. https://jhmi.sums.ac.ir/article_42680_28407bbf900e1255a0e99bee8203638a.pdf
- [11] Rezaeenoor, J., Saadi, G., & Jahani, M. (2019). Prediction of Cardiovascular Diseases Usingan Optimized Artificial Neural Network. *journal of ilam university of medical sciences*, 27(5), 15-23. <https://doi.org/10.29252/sjimu.27.5.15>
- [12] Yi, W., & Fuyong, W. (2006, August 7-11). *Breast Cancer Diagnosis via Support Vector Machines*. 2006 Chinese Control Conference, Harbin, China
- [13] Kiani, B., & Atashi, A. (2014). A prognostic model based on data mining techniques to predict breast cancer recurrence. *Journal of Health and Biomedical Informatics*, 1(1), 26-31. <http://eprints.kmu.ac.ir/id/eprint/24250>
- [14] Zabbah, I., Hassaanzadeh, M., & kohjani, z. (2017). The Effect of Continuous Parameters on the Diagnosis of Coronary Artery Disease Using Artificial Neural Networks. *Journal of Torbat Heydariyeh University of Medical Sciences*, 4(4), 29-39. <http://jms.thums.ac.ir/article-1-369-en.html>
- [15] Repository, U. M. L. (2017, May 4). *Breast Cancer Wisconsin (Diagnostic) Data Set*. UCI Machine Learning Repository. [https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+\(diagnostic\)](https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic))

- [16] Fallahi, A., & Jafari, S. (2011). An expert system for detection of breast cancer using data preprocessing and bayesian network. *International Journal of Advanced Science and Technology*, 34, 65-70.
- [17] Zabbah, I., Yasrebi Naeini, S. E., Ramazanpoor, Z., & Sahragard, K. (2017). The Diagnosis of Thyroid Diseases Using Combinati on of Neural Networks through Hierarchical Method. *Journal of Health and Biomedical Informatics*, 4(1), 21-31. <http://jhbmi.ir/article-1-187-en.html>
- [18] Badzaban, F., Rezaei-Moghaddam, K., & Fatemi, M. (2020). Entrepreneurship Education and its Effects on the Function of Rural Women's Businesses. *Karafan Quarterly Scientific Journal*, 17(3), 79-94. <https://doi.org/10.48301/kssa.2020.124671>