مطالعات علوم زیستی و زیست فناوری

دوره ۳، شماره ۳، پاییز ۱۳۹۶

صفحات ۵۵- ۴۱

Online ISSN: ٢٤٧٦-٤٤٩٣

Print ISSN: ٢٥٣٨-٢٠٢٠

www.irijournals.com

تشخیص سرطان پستان با استفاده از الگوریتم فرا ابتکاری

فاطمه قائمی[۱]، ایرج مهدوی [۲]

[۱] دانشجوی کارشناسی‌ارشد مهندسی فناوری اطلاعات، دانشگاه علوم فنون، بابل، ایران

[۲] استاد گروه صنایع، دانشگاه علوم و فنون، بابل، ایران

چکیده

با پیشرفت علم و تکنولوژی در فناوری اطلاعات روش‌های متعددی در حوزه سلامت مطرح شده‌اند و مورد آزمایش و پیاده سازی قرار گرفته‌اند. داده‌کاوی تکنیک و ابزار بسیار متداولی است که امروزه در زمینه‌های مختلفی کاربرد پیدا کرده است. تشخیص بیماری‌های مختلف در علم پزشکی، یکی از زمینه‌های پرکاربرد داده‌کاوی محسوب می‌شود که در سال‌های اخیر تحقیقات و مطالعات زیادی پیرامون آن انجام شده است. داده کاوی توانسته با تکنیک های مختلف از قبیل شبکه‌ی عصبی در این راستا گام موثری بردارد. سرطان بیماری وحشتناکی می باشد. میلیون‌ها انسان هر ساله به دلیل مبتلا شدن به این بیماری جان خود را از دست می دهند. سرطان پستان امروزه در میان افراد به صورت گسترده‌ای در حال پیشرفت بوده است و انتخاب روش درمانی مناسب و تشخیص زودهنگام در روند درمان آن، بسیار ضروری می باشد. سلول‌های سرطانی باید به درستی تشخیص داده شود. با توجه به ضرورت تشخیص صحیح و به موقع موجود، ما در این راستا گام برداشته و توانسته‌ایم با استفاده از شبکه‌ی عصبی و الگوریتم فراابتکاری ملخ دقت تشخیص را افزایش دهیم. در این پژوهش توانستیم با مقایسه‌ی تشخیص توسط الگوریتم ژنتیک که در مطالعات قبلی وجود داشت ثابت نماییم الگوریتم ملخ با ساختار شبکه عصبی مشابهی دقت تشخیص تومورها را بهبود بخشد. در مورد مطالعاتی مورد مقایسه دقت در تشخیص تومورهای خوش خیم وبدخیم به طور میانگین به میزان ۹۸٪ بوده و ما توانستیم با استفاده از الگوریتم ملخ دقت تشخیص را ۱/۷۵٪ افزایش داده و به میزان ۹۹/۷۵٪ برسیم.

واژه‌های کلیدی: سرطان پستان، ماموگرافی-داده کاوی، شبکه های عصبی، الگوریتم های فرا ابتکاری، الگوریتم ملخ.

۱- مقدمه

در حال حاضر با توجه به رشد روز افزون اطلاعات و حجم بالای داده‌های ذخیره شده در سیستم‌های پایگاهی نیاز به ابزاری است تا بتوان داده‌های ذخیره شده را پردازش نمود و اطلاعات حاصل از این پردازش را در اختیار تصمیم گیرندگان قرار داد. در دهه های اخیر، توانایی بشر برای تولید و جمع‌آوری داده‌ها به سرعت افزایش یافته و این رشد انفجاری در داده‌های ذخیره شده نیاز مبرم به ابزارهای خودکارسازی را ایجاد کرده است تا به صورت هوشمند به انسان یاری رساند تا این حجم عظیم داده‌ها را به دانش مفید و مورد نیاز خود تبدیل نماید. فرایند طبقه بندی داده ها در مسائل پزشکی، مدیریتی و مهندسی به کار گرفته می شود. روش های هوشمندانه از جمله محاسبات تکاملی به صورت گسترده‌ای در این زمینه استفاده می‌شود که از جمله برای تحلیل وضعیت بیماری سرطان و دیگر بیماری‌ها مورد استفاده قرار گرفته و به کمک پزشکان برآمده‌است. این سیستم‌ها ابزاری ارزشمند برای پیش‌بینی، تشخیص و کنترل این گونه بیماری‌ها و همچنین تحلیل بقای بیماران مبتلا می باشند. با استفاده از این روش‌ها و تشخیص زودهنگام سرطان می‌توان از این بیماری رها شد.

۲- ادبیات تحقیق

پیدایش شبکه های عصبی مصنوعی در علم پزشکی از جایی شروع شده که پزشکان با حجم زیادی از داده های بیماران رو به رو شدند و قدرت تشخیص بیماری برای آن ها سخت و تا حدی غیرممکن بود، از این رو پزشکان و دانشمندان به این فکر افتادند که نرم افزاری طراحی نمایند که در تشخیص راحت‌تر بیماری‌ها به آن‌ها کمک کند. در عین حال آن ها به نرم‌افزاری احتیاج داشته که قدرت پردازش و تصمیم‌گیری آن بالا باشد و مانند مغز انسان قابلیت تصمیم‌گیری داشته باشد از این رو به فکر ساختن شبکه‌های عصبی مصنوعی افتادند که شبیه ساز مغز انسان است. (صادق پور و خاکسار،۱۳۹۴).

در دنیای امروزی بیماری ها رو به افزایش می باشد. در این میان سرطان یکی از خطرناک ترین انواع بیماری می باشد. سرطان در حالی تبدیل به یکی از بیماری‌های عصر جدید شده که با وجود پیشرفت‌های شگرف در علم و دانش بازهم درمان و راه حل صد درصدی برای آن وجود ندارد. سرطان یا به اصطلاح پزشکی نئوپولیسم بدخیم[1] مجموعه‌ای از بیماری‌هاست که شامل رشد غیرطبیعی سلول می باشد. در سرطان، تقسیم و رشد سلول غیر قابل کنترل بوده و آن هایی که تومورهای بدخیم را شکل می دهند به طور تهاجمی در کنار اندام های بدن قرار میگیرند. سرطان می تواند در تمامی بدن از طریق سیستم لنفاوی و همچنین گردش خون پخش شود. تمامی تومورها سرطانی نمی باشند، تومورهای خوش خیم در کنار بافت ها به طور تهاجمی قرار نگرفته و در سراسر بدن پخش نمی شوند. بیش از ۲۰۰ نوع سرطان وجود دارد که انسان را تهدید می کند. علت این بیماری پیچیده می باشد و دلایل جزئی و محدودی تا کنون شناخته شده است. انواع مختلف سرطان وجود دارد که امروزه سرطان پستان یکی از شایع‌ترین و خطرناک‌ترین این بیماری می باشد که در سال‌های اخیر در جهان یکی از عوامل موثر مرگ‌ومیر شناخته شده است. سرطان پستان در میان زنان بسیار شایع می باشد.

۲-۱- سرطان سینه[2]

امروزه همواره بیماری ها روبه افزایش می باشد. در این میان سرطان یکی از خطرناک ترین این بیماری ها بوده است. در حال حاضر انواع مختلفی از سرطان وجود دارد که در این میان سرطان پستان یکی از خطرناک ترین آن ها و دلیل مرگ و میر در سال های اخیر دلیل بوده است و در آینده نیز دلیل مرگ و میر های بیشتری خواهد بود. این بیماری اغلب در میان زنان متداول می باشد (سئوسا و بیندو،۲۰۱۵). سرطان پستان یکی از انواع سرطان‌هاست که در آن سلول‌های غیرطبیعی به صورت

[1] Malignant neoplasm

[2] Breast Cancer

غیرقابل کنترل در یک یا هر دو سینه رشد می کنند. این سلول‌ها می‌توانند بافت‌های اطراف را مورد تهاجم قرار داده و موجب تشکیل یک توده شوند که اغلب تومور نامیده می‌شود. همچنین این سلول‌ها می‌توانند انتشار یافته و به بافت‌ها و غدد لنفاوی اطراف گسترش یابند که به آن متاستاز[1] گویند. در حالی که میزان شیوع این بیماری در کشورهای در حال توسعه کمتر می باشد، میزان مرگ و میر بیشتر می باشد. بهبود این بیماری به دو عامل اصلی تشخیص در مراحل ابتدائی و انتخاب روش درمان مطلوب وابسته می باشد (ییپ و تیب،۲۰۱۴). علت اصلی این بیماری در اکثر بیماران مبهم می باشد، ولی عوامل خطرزای متعددی مربوط به پیشرفت این بیماری می شوند که شامل افزایش سن، سابقه ی خانوادگی، چاقی، مصرف الکل، قرارگرفتن در معرض استروژن، وراثت ژن های مستعد، مشخصا ژن های RBCA1 و RBCA2 دو ژن از مجموعه ژن‌های انسان هستند که وظیفه اصلی آنها کنترل سلامت DNA و ترمیم جهش‌ها و آسیب‌های وارده به DNA است وظیفه اصلی دو ژنBRCA1 و BRCA2ترمیم جهش و آسیب های وارده به DNA ااست. این دو ژن عضو گروهی از ژن‌ها هستند که به خانواده ژن‌های سرکوب‌گر تومور معروف هستند. برخلاف پیشرفت در شناسایی عوامل خطرزا و نشان های ژنتیکی این بیماری، حدودا ۷۰ تا ۸۰ درصد مواردی که در زنان اتفاق می‌افتد بدون دلیل شناخته شده‌ی پیش بینی شده ای می باشد (کلی و همکارانش،۲۰۰۹). این بیماری در ابتدا علائم و نشانه ای ندارد. غده شاید آنقدر کوچک باشد که قابل حس کردن نباشد و نشانه های تغییرات غیرمعمولی را در خود احساس نکنیم. اغلب ناحیه ی غیر معمول در غربالگری ماموگرافی مشخص شده که پس از آن به تست های گوناگون منجر خواهد شد. در برخی موارد نیز توده توسط خود شخص یا دکتر قابل تشخیص می باشد. توده‌ای که بدون درد، سخت و دارای لبه می باشد بیشتر احتمال توده ی سرطانی بودن را دارد ولی گاها توده نازک و نرم و دایره ای می باشد. به این دلیل است که هرساله ضروری می باشد که هرساله جهت معاینه به دکتر مراجعه شود. از نشانه های آن می توان به ورم کل یا قسمتی از سینه، خارش یا کم رنگ شدن پوست، درد در سینه، درد یا فرورفتگی در نوک سینه، وجود توده در ناحیه ی زیربغل و ... باشد. روش های متعددی در تشخیص این بیماری مطرح می شوند; معاینه توسط خود شخص، معاینه توسط پزشک، ام آر آی[2]، سونوگرافی، ترموگرافی[3]، بیوپسی[4]، ماموگرافی[5].

طبقه بندی پستان در ماموگرافی انجام می شود و اکثر آن ها طبقه بندی خوش خیم می باشد، ولی تشخیص خصوصا در طبقه بندی کوچک تر از ۱ میلیمتر دقیق ترین ماموگرافی جهت تشخیص زودهنگام می باشد.۷۰ درصد از سرطان ها خود را از طریق میکرو طبقه بندی آشکار می کنند؛ بنابراین تشخیص طبقه بندی در ازیابی ماموگرافی به عنوان موضوع مهمی مطرح می شود (اسماعیل سریتاس،۲۰۱۲). ماموگرافی در زنان بالای ۳۵ سال انجام میگیرد (حسین قیومی زاده و همکارانش،۲۰۱۲)

۲-2- داده کاوی

داده کاوی فرآیندی جهت شناسایی الگوها و مدل های صحیح، جدید و به صورت بالقوه مفید، در حجم وسیعی از داده می باشد، به طریقی که این الگوها و مدل ها برای انسان ها قابل درک باشند .کاوش داده به معنی کنکاش داده های موجود در پایگاه داده و انجام تحلیل های مختلف بر روی آن و استخراج اطلاعات می باشد. به صورت دقیق تر می توان گفت:

" کاوش داده‌ها، شناسایی الگوهای صحیح، بدیع، سودمند و قابل درک از داده های موجود در یک پایگاه داده است که با استفاده از پردازش های معمول قابل دستیابی نیستند"(شهرابی، ۱۳۹۰).

فرآیند طبقه بندی داده‌ها با استفاده از دانشی که از داده‌های شناخته شده به دست می آید، امروزه یکی از پرطرف‌دارترین موضوعات مورد مطالعه در آمار، علوم تصمیم گیری، علوم کامپیوتری و مسائل پزشکی می باشد. مسائل متنوعی نظیر تشخیص تصاویر، تشخیص بیماری و ... از تکنیک های طبقه بندی استفاده می کنند .اخیراً مدل های هوشمند نظیر شبکه های عصبی

[1] Metastasis
[2] Magnetic Resonance Imaging
[3] Thermography
[4] biopsy
[5] Mammography

مصنوعی، در این حوزه، به فراوانی مورد استفاده قرار گرفته اند. در این گونه مسائل استفاده از مدل های کلاسیک آماری با محدودیت ها و مشکلاتی همراه است. همواره مشکل برقراری شرایط اولیه مورد نیاز این مدل ها مانند توزیع نرمال متغیرهای پاسخ، یکسان بودن وار یانس خطا و ... که بر پایه آن بتوان مدل مناسبی بر اساس داده های مورد نظر برازش نمود، وجود داشته است. شبکه های عصبی مصنوعی که از روش های نوین مدلسازی و پیش بینی هستند، محدودیت های روش های کلاسیک را ندارند (دارائی و همکارانش،۲۰۱۵).

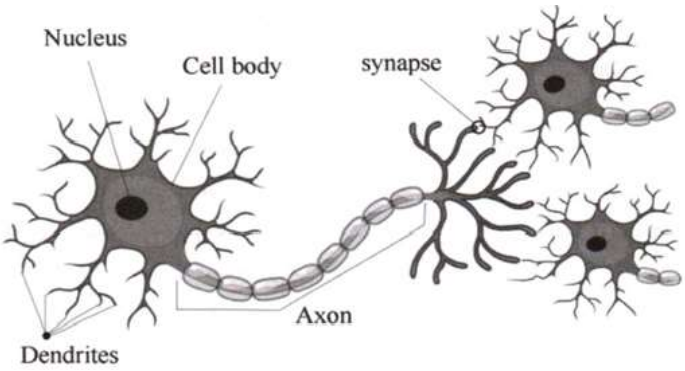
۲-۳- شبکه های عصبی

شبکه های عصبی مصنوعی تا حدودی از مغز انسان الگو برداری شده اند و همان گونه که مغز انسان می تواند با استفاده از تجربیات قبلی و مسائل از پیش یاد گرفته، مسائل جدید را تجزیه و تحلیل نمایند، شبکه های عصبی نیز در صورت آموزش قادرند بر مبنای اطلاعاتی که به اعضای آن ها آموزش داده شده، جواب های قابل قبول ارائه دهند و نیز می توان از آن ها به طور نامحدود در ارائه ی جواب به اطلاعاتی که قبلا با آن ها مواجه نبوده اند استفاده نمود (صادق پور و خاکسار حقانی،۱۳۹۵). شبکه های عصبی در سال های اخیر به طور گسترده ای در زمینه های گوناگون به عنوان ابزار هوشمند از قبیل هوش مصنوعی، تشخیص الگو، تشخیص پزشکی، یادگیری ماشین و ... مورد استفاده قرار گرفته است. شبکه عصبی به عنوان نگاشتی از ورودی به خروجی می باشد. شبکه عصبی در به کارگیری در طبقه بندی سرطان پستان طبق الگوی زیر عمل می کند:

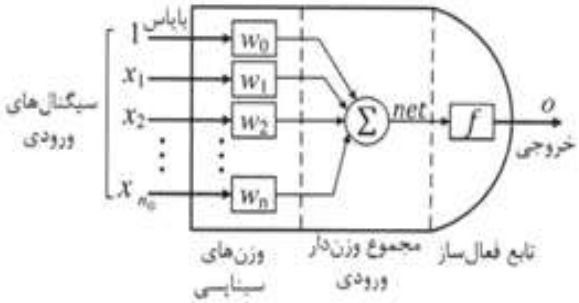
برای شروع، از طریق نمونه های آموزشی و کلاس های نمونه ها، پیش بینی شبکه از هر نمونه با برچسب کلاس شناخته شده ی واقعی مقایسه می شود و وزن هر نمونه ی آموزشی برای رسیدن به هدف طبقه بندی دیگر داده های نمونه مشخص می‌شود (سین و تون،۲۰۱۵). مجهز شدن علم پزشکی به ابزارهای هوشمند در تشخیص و درمان بیماری ها می تواند اشتباهات پزشکان و خسارات جانی و مالی را کاهش می دهد. یک روش غیر پارامتری برای طبقه بندی است که در حیطه پزشکی بر اساس متغیرهای ورودی نسبت به طبقه بندی افراد به بیمار یا سالم اقدام می کند. طبقه بندی و پیشگویی وضعیت بیمار بر اساس عوامل خطر یکی از کاربردهای شبکه های عصبی مصنوعی است. شبکه های عصبی، با قابلیت قابل توجه آنها در استنتاج معانی از داده های پیچیده یا مبهم، می تواند برای استخراج الگوها و شناسایی روش هایی که آگاهی از آنها برای انسان و دیگر تکنیک های کامپیوتری بسیار پیچیده و دشوار است به کارگرفته شود از مزیت های استفاده از شبکه های عصبی می‌توان به یادگیری انطباق پذیر[1]، سازمان دهی توسط خود[2]، عملکرد به هنگام[3]، تحمل اشتباه بدون ایجاد وقفه در هنگام کد گذاری اطلاعات اشاره نمود.

طرح واره‌ای از سلول عصبی نورن به عنوان واحد پردازنده اطلاعات در سیستم‌های زیستی، در شکل زیر نشان داده شده است.

[1] Adaptive learning
[2] Self-organizing
[3] Real time



شکل ۱- طرح واره ای از نورون زیستی



شکل ۲ - مدل ریاضی نورون زیستی

در شکل ۱ چهار پارامتر اصلی تحریک نورن نشان داده شده اند که عبارت‌اند از وزن سیناپسی[۱]، دندریت یا ورودی نورن[۲]، هسته مرکزی[۳] یا مرکز تصمیم گیری و آکسون[۴] یا خروجی نورن. مدل ریاضی نورن در شکل ۲ نشان داده شده است.

این مدل پرکاربرد چهار پارامتر اصلی دارد که عبارت اند از:

- سیگنال های ورودی: معادل با دندریت در نورن زیستی هستند و به عنوان ورودی های نورن در نظر گرفته می شوند.
- وزن های سیناپسی: مجموعه ای از لینک های ارتباطی هستند که مسئولیت ایجاد ارتباط بین سیگنال‌های ورودی و درون نورن را برعهده دارند. به هر وزن، مقادیری مثبت یا منفی اختصاص داده شده و سیگنال های ورودی در آن‌ها ضرب می شوند. البته توجه داشته باشید که سیناپس‌های مغز انسان فقط مقادیر مثبت را اختیار می‌کنند.
- جمع کننده: وظیفه جمع کردن سیگنال های وزن دار شدهٔ ورودی را بر عهده دارند.
- تابع فعال ساز[۵] خروجی: برای انجام عملیات ریاضی خطی یا غیرخطی و محدودسازی دامنهٔ خروجی نورون بکار می-رود. معمولاً دامنه‌ی خروجی محدود شده در بازه 0.1 یا 1,1- در نظر گرفته می‌شوند. نتیجهٔ کار جمع کننده و تابع فعال سازی، ایجاد خروجی توسط هسته مرکزی به ازای سیگنال‌های ورودی است.

در مدل بالا یک بایاس نیز وجود دارد که با عدد ۱ نشان داده شده است. در مجموع، شکل ریاضی سیگنال خروجی نورون مصنوعی را می توان به شکل زیر نوشت:

[1] Synapse
[2] Dendrite
[3] Nucleus
[4] Axon
[5] Activation function

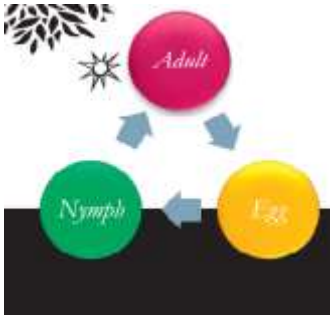
$net=\sum_{i=1}^{n} w_i x_i + w_0$ (معادله ۱)

$o=f(net)$ (معادله ۲)

x_iها سیگنال های ورودی و w_iها وزن های سیناپسی هستند. خروجی جمع کننده (net) ترکیب خطی سیگنال های وزن دار و بایاس، یا به بیان دیگر، مجموع ورودی‌های وزن دار نورون است. تابع f، تابع فعال سازی سیگنال خروجی نورون است. تأثیر رفتار w(وزن بایاس) در خروجی جمع کننده خطی، درشکل شماره‌ی۳ نشان داده شده است.

۲-۴- الگوریتم فرا ابتکاری ملخ (سرامی و همکارانش، ۲۰۱۷)

ظهور فرا ابتکاری ها برای حل مسائل بهینه سازی ترکیبی یکی از قابل توجه ترین دستاوردهای دو دهه ی گذشته شده است. در واقع الگوریتم‌های فراابتکاری، یکی از انواع الگوریتم‌های بهینه‌سازی تقریبی هستند که دارای راهکارهای برونرفت از نقاط بهینه محلی هستند و قابلیت کاربرد در طیف گسترده‌ای از مسائل را دارند. رده‌های گوناگونی از این نوع الگوریتم در دهه‌های اخیر توسعه یافته‌است که همه این ها زیر مجموعه الگوریتم فراابتکاری می باشند. از رده ی این الگوریتم‌ها می‌توان به الگوریتم ژنتیک، الگوریتم ذرات پراکنده، بهینه سازی کلونی مورچه ها، الگوریتم وال، الگوریتم کلونی زنبورها، الگوریتم ملخ و ... نام برد. ملخ ها حشراتی اند که به دلیل آسیب رساندن به محصولات کشاورزی به عنوان آفت شناخته شده اند. چرخه ی عمر این حشرات در شکل زیر نشان داده شده است



شکل ۳- چرخه ی عمر ملخ

جنبه ی منحصربه فرد دسته ی ملخ ها این است که رفتار دسته‌ای در دوران بلوغ و پوره مشاهده می شود. میلیون ها پوره ی ملخ مانند استوانه ی غلطان حرکت می کنند و می‌پرند. در مسیرشان تقریبا تمامی سبزیجات را می خورند. پس از این رفتار، زمانی که بالغ شدند یک دسته در هوا تشکیل می دهند. ملخ ها این گونه در فواصل بلند مهاجرت می‌کنند. شاخصه‌ی اصلی دسته در فاز لارو بودن حرکات آهسته و قدم های کوچک ملخ ها می باشد. در مقابل، مسافت طولانی و حرکات ناگهانی شاخصه ی ضروری دسته در بلوغ می باشد. جست و جو به دنبال منبع غذا، ویژگی مهم دیگری در دسته های ملخ ها می باشد. ملخ های پوره روی زمین می نشینند و موقعیت آن ها نباید زیر حد آستانه ی باشد. این معادله را در شبیه سازی دسته ای و الگوریتم بهینه سازی به کار نمی بریم به خاطر این که الگوریتم را از جست و جو و بهره برداری ناحیه ی جست و جو در اطراف جواب منع می کند. در جقیقت این مدل برای دسته ها در فضای آزاد به کار می رود؛ بنابراین معادله‌ی ۳ می‌تواند به کار رود و تعامل بین ملخ ها در دسته را شبیه سازی نماید.

(معادله ۳)

$$X_i = \sum_{\substack{j=1 \\ j\neq i}}^{N} s(|x_j - x_i|)\frac{x_i - x_j}{d_{ij}} - g\widehat{e_g} + u\widehat{e_g}$$

هرچند این مدل ریاضی نمی تواند به طور مستقیم در حل مسائل بهینه سازی مورد استفاده قرار گیرد، به دلیل این که ملخ ها به سرعت به ناحیه ی آسودگی می رسند و به نقطه ی مشخصی همگرا نمی شوند. نسخه ی تغییر یافته ی این معادله در نظر گرفته شده است تا مسائل بهینه سازی را حل کند:

$$X_i^{\text{۲}} = C\left(\sum_{\substack{j=1 \\ j\neq i}}^{N} C\,\frac{ub_d - lb_d}{2}\, s(\left|x_j^d - x_i^d\right|)\,\frac{x_i - x_j}{d_{i,j}}\right) + T^{\text{۲}}_d \quad \text{(معادله ۴)}$$

اولین قسمت معادله ی ۴، مجموع، متمرکز بر موقعیت دیگر ملخ‌ها می باشد و بر تعامل ملخ ها در طبیعت اشاره دارد. دومین قسمت، $\hat{T}_d$ مایل آن ها به حرکت به سمت منبع غذا را شبیه سازی می کند. همچنین پارامترc، کاهش سرعت ملخ هایی که به سمت منبع غذا حرکت می کنند و به تدریج تحلیل می روند را بیه سازی می نماید. برای ارائه ی فتار تصادفی بیشتر و به عنوان جایگزین، دو عبارت در معادله ۴ با مقادیر تصادفی ضرب می شوند. همچنین عبارات تکی میتوانند با مقادیر تصادفی ضرب شوند تا رفتار های تصادفی در تعامل با ملخ ها و یا تمایل به سمت منبع غذا را ارائه دهند. فرمول های در نظرگرفته شده قادر خواهند بود تا فضای جست و جو را کشف و بهره برداری نمایند. با این حال باید مکانیسمی وجود داشته باشد تا عامل جست و جو را ملزم به تنظیم سطح جست‌وجو به بهره‌برداری نماید در طبیعت ملخ ها ابتدا به حرکت و جست و جوی محلی می پردازند چرا که در مرحله ی لارو بودن بالی ندارند. سپس آزادانه در هوا پرواز کرده و ناحیه ی وسیع تری را کشف می کنند. در الگوریتم های بهینه سازی تصادفی، اکتشاف اولیه به سبب نیاز برای یافتن ضای جست و جو می باشد. پس از یافتن فضای جست و جو هره برداری عامل جست و جو را موظف می نماید تا به صورت محلی جست و جو کند و تقریب دقیقی از بهینه ی سراسری را بیابد. برای متعادل نمودن جست و جو و بهره برداری لازم می باشد که پارامتر c متناسب یا تعداد تکرارها کاهش یابد. اینمکانیسم همان طور که تعداد تکرار را افزایش می دهد. بهره برداری را افزایش می دهد. ضریب c ناحیه ی آسودگی را متناسب با تعداد تکرار کاهش می دهد و به صورت زیر محاسبه می گردد:

$c=cmax-l\frac{cmax-cmin}{L}$ معادله ۵

شبه کد الگوریتم GOA در شکل ۵ نشان داده شده است. GOA بهینه سازی را با ایجاد مجموعه ای از راه حل های تصادفی آغاز می کند. عامل های جست و جو موقعیت خود را بر اساس معادله ۴ به روز می نماید. موقعیت بهترین هدف به دست آمده تا اینجا در هر تکرار آپدیت شده است. به علاوه فاکتور c با استفاده از معادله ی شماره ی ۵ محاسبه می شود و فاصله ی بین ملخ ها در بازه ی بین [1,4] در هر تکرار نرمال می شود. به روز رسانی موقعیت به صورت تکراری تا ده دست آمدن یک معیار رضایت بخش انجام میگیرد. موقعیت و شایستگی بهترین هدف به عنوان بهترین تقریب برای بهینه ی سراسری به دست می آید. شبه کد الگوریتم GOA در شکل ۴ نشان داده شده است. GOA بهینه سازی را با ایجاد مجموعه ای از راه حل های تصادفی آغاز می کند. عامل های جست و جو موقعیت خود را بر اساس معادله ۴ به روز می نماید. موقعیت بهترین هدف به دست آمده تا اینجا در هر تکرار آپدیت شده است. به علاوه فاکتور c با استفاده از معادله ۵ محاسبه می شود و فاصله ی بین ملخ ها در بازه ی بین [1,4] در هر تکرار نرمال می شود. به روز رسانی موقعیت به صورت تکراری تا ده دست آمدن یک معیار رضایت بخش انجام میگیرد. موقعیت و شایستگی بهترین هدف به عنوان بهترین تقریب برای بهینه ی سراسری به دست می-آید.

Initialize the swarm X_i (i = 1, 2, ..., n)
Initialize cmax, cmin, and maximum number of iterations
Calculate the fitness of each search agent
T=the best search agent
***while** (l < Max number of iterations)*
    Update c using Eq. (2.8)
    ***for** each search agent*
        Normalize the distances between grasshoppers in [1,4]
        Update the position of the current search agent by the equation (2.7)
        Bring the current search agent back if it goes outside the boundaries
    end for
    Update T if there is a better solution
    l=l+1
end while
Return T

شکل ۴ - شبه کد الگوریتم ملخ

۳- مدل و فرضیه‌های تحقیق

در این تحقیق بر آن شدیم تا به بهبود خطای پیش بینی سرطان پستان، با استفاده از الگوریتم فراابتکاری ملخ نسبت به تحقیق انجام شده در این زمینه (دارائی و همکاران،۲۰۱۵) که با استفاده از ترکیب شبکه ی عصبی و الگوریتم انجام شده پرداخته شود. داده های مربوط به سرطان پستان که در قسمت آموزش شبکه ی عصبی در این تحقیق مورد استفاده قرار گرفته است، توسط دکتر William H.Wolberg از دانشگاه ویسکونسین ایالات متحده آمریکا گرد آوری شده است. این مجموعه، دارای ۶۹۹ نمونه (رکورد) داده‌ای با مقادیر متناسب ویژگی ها می‌باشد که از طریق سیتولوژی تومورهای ناحیه ی پستان به دست آمده‌اند. هر نمونه داده‌ای دارای نه ویژگی می باشد که مقادیر عددی صحیح، در بازه ی [1,10] دارند. برای ارزیابی اولیه این ویژگی ها آزمون T را در نرم افزار SPSS برروی این متغیرها اجرا نمودیم که نتایج این آزمون به همراه شرح ویژگی ها در جدول ۱ ارائه می شود.

جدول ۱ - نتایج آزمون T بر روی ویژگی‌های داده‌ها

ردیف	نام ویژگی (متغیر مستقل)	کلاس (متغیر وابسته)	تعداد	انحراف معیار+میانگین	سطح معنی دار*
1	Clump Thickness	خوش خیم (۱) بد خیم (۲)	۴۴۴ ۲۳۹	۲/۹۶ ± ۱/۶۷۳ 7/19 ± 2/438	0/000
2	Uniformity of Cell Size	خوش خیم (۱) بدخیم (۲)	۴۴۴ ۲۳۹	1/31 ± 0/856 ۶/۵۸ ± ۲/۷۲۴	0/000
3	Uniformity of Cell Shape	خوش خیم (۱) بدخیم (۲)	۴۴۴ ۲۳۹	۱/۴۱ ± ۰/۹۵۷ 6/58 ± 2/569	0/000

۴	Marginal Adhesion	خوش خیم (۱) بدخیم (۲)	۴۴۴ ۲۳۹	۱/۳۵ ± ۰/۹۱۷ 3/197±/59	0/000
۵	Single Epithelial Cell Size	خوش خیم (۱) بدخیم (۲)	۴۴۴ ۲۳۹	2/11±۰/۸۷۷ 5/33±2/443	0/000
۶	Bare Nuclei	خوش خیم (۱) بدخیم (۲)	۴۴۴ ۲۳۹	1/35 ± 1/178 7/63 ± 3/117	0/000
۷	Bland Chromatin	خوش خیم (۱) بدخیم (۲)	۴۴۴ ۲۳۹	2/08 ± 1/062 5/97 ± 2/282	0/000
۸	Normal Nucleoli	خوش خیم (۱) بدخیم (۲)	۴۴۴ ۲۳۹	۱/۲۶ ± ۰/۹۵۵ 5/86 ± 3/349	0/000
9	Mitoses	خوش خیم (۱) بدخیم (۲)	۴۴۴ ۲۳۹	1/07 ± 0/510 2/60 ± 2/564	0/000

*سطح معنی دار در فاصله اطمینان 95 % محاسبه شده است.

داده ها در دو کلاس (گروه)، خوش خیم و بدخیم، طبقه بندی شده اند. از آن جایی که ۱۶ رکورد، هر کدام دارای یک فیلد خالی بودند، ما آن ها را از مجموعه داده ای حذف نمودیم و در نتیجه ۶۸۳ نمونه باقی ماندند که از این تعداد، ۴۴۴ نمونه متعلق به بیمارانی با تومور خوش خیم و ۲۳۹ نمونه متعلق به بیماران با تومور بدخیم می باشد. همان طورکه در جدول شماره ۱، مطابق با نتایج مشاهده می شود، میانگین و انحراف معیار کلیه T آزمون ویژگی ها در دو گروه سالم و بیمار تفاوت معنی داری دارند و این موضوع حاکی از ارتباط مقادیر همه این ویژگی ها با خوش خیمی یا بدخیمی تومورهای ناحیه پستان است و بر این اساس همه ویژگی ها را به عنوان متغیرهای ورودی، در مدل شبکه عصبی استفاده نمودیم.

ساختار شبکه ی عصبی ما مطابق با پژوهش انجام شده در این زمینه می‌باشد. ما از ساختار شبکه عصبی ((9,6.8,1 به دست آمده در پژوهش آقای دارائی و همکارانش استفاده نمودیم و وزن های شبکه عصبی را با استفاده از الگوریتم ملخ بهینه نمودیم. حوزه استفاده از الگوریتم های فراابتکاری (در این تحقیق الگوریتم ملخ) در شبکه های عصبی آن جایی است که آموزش یک شبکه و به عبارتی دیگر تعیین وزن یال های آن به یک مساله بهینه سازی ختم می شود.

۴- روش تحقیق

برای به کارگیری داده ها در مدل شبکه عصبی مصنوعی، ابتدا باید آن ها را نرمال نمود. بدین منظور کلیه داده های هر ستون داده ای (ویژگی) را بر مقدار بیشینه هر ستون که عدد ۱۰ می باشد، تقسیم نمودیم که در نتیجه کلیه داده ها در بازه [0,1] قرار گرفتند. برای محاسبه دقیق کارایی مدل پیشنهادی در مراحل 5-fold cross validationآموزش و آزمون از روش استفاده نمودیم. بر این اساس، مجموعه داده ای را به پنج قسمت تقریباً مساوی تقسیم نموده، به گونه‌ای که نسبت داده های دو کلاس (بدخیم به خوش خیم) در هر قسمت، برابر با نسبت آن ها در کل داده ها یعنی (444÷239) باشد .بر اساس این روش، در هر اجرا، چهار قسمت از پنج قسمت از داده ها جهت آموزش شبکه و یک قسمت باقیمانده برای آزمودن شبکه استفاده گردید. این عمل برای هر پنج قسمت، انجام شده، میانگین پارامترهای مربوط به کارایی شبکه در مراحل آموزش و آزمون به طور جداگانه، محاسبه گردید.

در ابتدا ویژگی ها با وزن فرضی وارد شبکه می شوند و خطای شبکه محاسبه می شود، سپس در هر مرحله ی آموزش شبکه الگوریتم ملخ وزن ها را آپدیت کرده تا نهایتا وزن مناسب ویژگی ها به دست آید.

۴-۱- ساختار شبکه عصبی مورد نظر

در گام اول مقدار دهی اولیه به نورون ها و لایه ها انجام می‌گیرد. در این گام تعداد ویژگی ها، تعداد لایه های میانی، تعداد گره های لایه ی میانی و تعداد گره های لایه ی خروجی و تعداد عامل و تکرار های الگوریتم بهینه سازی ملخ را مشخص می نماییم. تعداد ویژگی ها در مجموعه داده های ما ۹ عدد در نظر گرفته شده است. تعداد لایه های میانی بر اساس پژوهش عنوان شده ی ما ۲ عدد در نظر گرفته شده. تعداد گره های لایه ی میانی اول ۸ عدد و تعداد گره های لایه ی دوم ۶ عدد در نظر گرفته شده. مقدار دهی اولیه به ملخ; در واقع در این قسمت تعداد عامل جست و جو که در این جا تعداد ملخ ها می باشند را مشخص می کنیم که ما به صورت انتخابی۳ عامل را در نظر گرفتیم. تعداد تکرار الگوریتم ملخ را ۲۰ مرتبه در نظر می گیریم. در گام بعدی وزن دهی اولیه به متغیرهایی که قرار است بهینه شوند انجام می‌گیرد. در این گام از الگوریتم ملخ در جهت وزن دهی استفاده نمودیم. تعداد متغیر هایی که قرار است بهینه شوند طبق معادله ی زیر به دست می آیند:

تعداد نورون های لایه‌ی پنهان [تعداد ورودی ها (ویژگی ها)+BIOS]

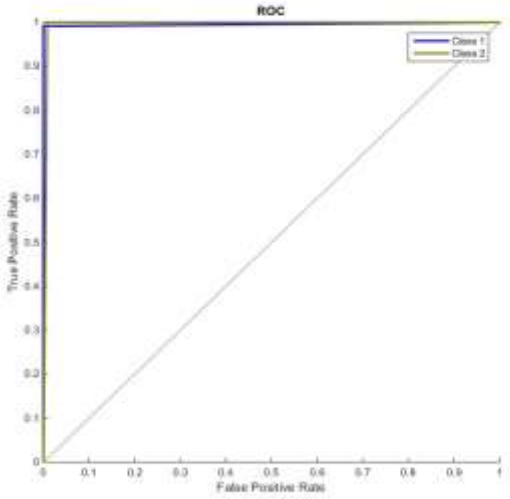
در ابتدای کار به هر ویژگی یک مقدار تصادفی در بازه‌ی[1,1-] می دهیم. این بازه در الگوریتم ملخ پیشنهاد شده بود و ما نیز جهت انتخاب تصادفی اولیه ی اوزان [1,1-] در نظر گرفته شد. ما تعداد عامل ها را در گام اول مشخص نمودیم. در اینجا هر عامل دارای یک ماتریس ۱۰۱۴ می باشد که نیاز به بهینه سازی آن دارد (هر ملخ باید متغیر ها را بهینه نماید و طبق رابطه‌ی عنوان شده در بالا ۱۰ ورودی و ۱۴ لایه‌ی میانی داریم). در گام بعدی به صورت تصادفی از مجموعه داده های ما یک رکورد انتخاب می شود. پس از آن مقدار پیش بینی شده از تابع فعال ساز دریافت شده و با دانستن مقدار واقعی بر اساس معادله ی زیر میزان خطا را محاسبه می کند. در مرحله‌ی بعد به روز رسانی وزن ها توسط الگوریتم ملخ انجام می گیرد. این به روز رسانی بر اساس معادله ۴ که به آن اشاره شده است، انجام می‌پذیرد. به روز رسانی به تعداد تعریف شده ی ۱۰ مرتبه انجام می‌گیرد و نهایتا بهترین جواب انتخاب می شود.

۵- یافته‌های تحقیق

پس از آنکه بر اساس روش پیشنهادی مطرح شده در فصل قبل، پیاده سازی الگوریتم ملخ و شبکه عصبی را با استفاده از نرم افزار متلب انجام دادیم، داده های آموزشی را به داده و مدل آموزش دیده را با داده‌های آزمایشی آزمایش نمودیم که در نهایت نتایج زیر به دست آمده است:

۵-۱- نمودار کارائی الگوریتم ملخ

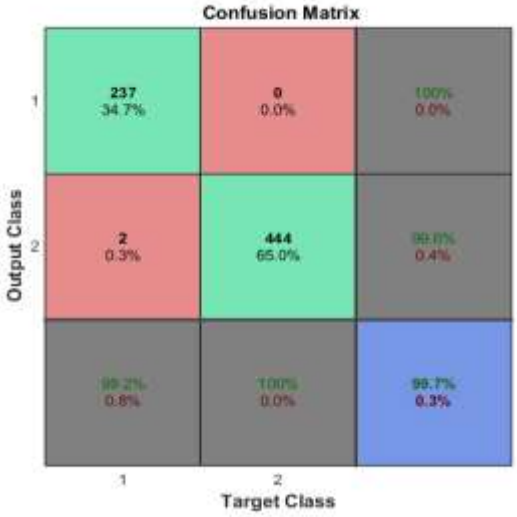
معیار مهمی که برای تعیین میزان کارایی یک دسته بندی اضافه می شود معیار AUC است. AUC نشان دهنده سطح زیر نمودار ROC می‌باشد. هرچه مقدار این عدد مربوطه به یک دسته بندی بزرگتر باشد کارایی نهایی دسته‌بند مطلوب‌تر ارزیابی می‌شود. نمودار ROC روشی برای بررسی کارایی دسته‌بندها می‌باشد. همانطور که در شکل شماره‌ی ۵ مشاهده می‌شود کلاس ۲ که مربوط به بدخیم می‌باشد توسط سیستم پیشنهادی نسبت به کلاس ۱ (که مربوط به تشخیص خوش‌خیم است)، بهتر تشخیص داده می‌شود زیرا به خط افقی بالای نمودار و خط عمود سمت چپ نمودار نزدیک‌تر است.



شکل ۵- نمودارROC

۵-۲- ماتریس درهم‌ریختگی و دیگر معیارهای ارزیابی

ماتریس درهم ریختگی[1] به ماتریسی گفته می‌شود که در آن عملکرد الگوریتم‌های مربوطه را نشان می‌دهند



شکل ۶- ماتریس‌های درهم ریختگی روش پیشنهادی

[1] confusion matrix

هر یک از عناصر ماتریس درهم‌ریختگی برای کلاس ۱ به شرح ذیل می‌باشد:

جدول ۲ - میزان پارامترهای tn tp, fp, fn, برای هریک از دو کلاس

	False Negative	False Positive	True Positive	True Negative
Class1	۰/۰۰۴۵	0	1	۰/۹۹۵۵
Class2	0	۰/۰۰۴۵	۰/۹۹۵۵	1

مهمترین معیار برای تعیین کارایی یک الگوریتم دسته‌بندی دقت یا نرخ دسته‌بندی است که این معیار دقت کل یک دسته‌بند را محاسبه می‌کند و در واقع این معیار مشهورترین و عمومی‌ترین معیار محاسبه کارایی الگوریتم دسته‌بند است که نشان می دهد دسته‌بند طراحی شده چند درصد از کل مجموعه رکوردهای آزمایشی را به درستی دسته‌بندی کرده است.

همچنین در مسائل واقعی معیارهای دیگری نظیر DR و FAR از اهمیت ویژه‌ای برخوردارند. این معیارها که توجه بیشتری به دسته‌بند مثبت نشان می‌دهند توانایی دسته‌بند را در تشخیص دسته مثبت نشان می‌دهند. به طور مشابه تاوان این توانایی تشخیص را تعیین می‌کند. معیار DR نشان می‌دهد که دقت تشخیص دسته مثبت چه مقدار است و معیار FAR نرخ هشدار غلط را با توجه به دسته منفی بیان می‌کند. میانگین اعداد بدست آمده برای این معیارها در جدول شماره‌ی۳ آمده است.

جدول ۳ - معیارهای ارزیابی روش پیشنهادی

الگوریتم پیشنهادی	DR	FAR	Accuracy
GOA + mlp	۰/۹۹۷۷	۰/۰۰۲۲	۹۹/۷۱

۶- بحث و نتیجه‌گیری

همان گونه که در تمامی مراحل عنوان شده است امروزه علم کامپیوتر در تمامی رشته ها نقش مهم و اساسی ایفا می نماید. شبکه های عصبی از جمله کاربرد هایی از شاخه های کامپیوتر می باشند که توانسته اند در تمامی رشته ها در تصمیم گیری و تشخیص به کار آیند. علم پزشکی از جمله علومی می باشد که دقت و صحت اطلاعات و نتایج به صورت مستقیم با جان افراد در ارتباط می باشد. محاسبات با استفاده از شبکه های عصبی توانسته اند نتایج دقیق تری را به نمایش گذارند و صحت اطلاعات را افزایش دهند و به پزشکان در تشخیص و درمان کمک شایانی نمایند. امروزه سرطان پستان در افراد به صورت چشمگیری رو به افزایش می باشد، از این رو بر آن شدیم تا با کمک شبکه عصبی و تلفیق آن با الگوریتم فرا ابتکاری ملخ در تشخیص این بیماری گام برداریم. این الگوریتم به دلیل داشتن دو نوع جست‌وجوی محلی و سراسری برای هر عامل به صورت تکراری می‌تواند در هر تکرار پس از جست‌وجوی سراسری به جست و جوی محلی بپردازد. وجود چندین عامل در هر جست و جو و همچنین متغیر تصادفی C در بهبود جست و جو موثر می باشد. مقایسه ای از دقت به دست آمده در ۵ مرحله‌ی مورد نظر و بر روی داده های آزمون به صورت زیر می باشد:

جدول ۴- مقایسه ی دقت الگوریتم ژنتیک و ملخ

شماره ی اجرا	داده های آزمون	دقت مقاله	دقت پژوهش
۱	F1	۰.۹۷۸	۰.۹۹۳
۲	F2	۰.۹۵۶	۰.۹۹۸
۳	F3	۰.۹۶۴	۰.۹۹۷
۴	F4	۰.۹۶۴	۰.۹۹۸
۵	F5	۰.۹۹۳	۰.۹۹۹
میانگین		۰.۹۷۱	۰.۹۹۷

در پژوهش مورد مقایسه ی ما (دارایی و همکارانش، ۲۰۱۵) از شبکه ی عصبی استفاده شده و با ارائه‌ی ساختار NN(9-8-6-1) و همچنین استفاده از الگوریتم ژنتیک و ارائه‌ی اوزان بهینه‌ی ویژگی‌ها با دقت ۹۸٪، به تشخیص تومور های بدخیم و خوش خیم کمک نموده است. ما در پژوهش خود از ساختار ارائه شده‌ی NN(9-8-6-1) استفاده نمودیم و با استفاده از الگوریتم ملخ در جهت بهبود دقت تشخیص و ایجاد اوزان در وضعیت بهینه تلاش نمودیم و به دقت ۹۹/۷۵٪ رسیدیم.

منابع

۱. صادق پور. ه و خاکسار حقانی. آ.تشخیص بیماری سرطان سینه توسط شبکه های عصبی مصنوعی.۱۳۹۵. دومین کنفرانس ملی رویکرد ها ی نوین در مهندسی کامپیوتر و برق.

۲. شهرابی، ج (۱۳۹۰ب). داده کاوی .انتشارات جهاد دانشگاهی، واحد صنعتی امیرکبیر.

۳. دارائی، م.، وحیدی، ج؛ و علی پور، ع. (۱۳۹۴). ارائه روشی مبتنی بر یک الگوریتم تکاملی برای دستیابی به مدلی کارآمد ازشبکه عصبی مصنوعی جهت پیش بینی وضعیت تومورهای پستان. مجله دانشگاه علوم پزشکی مازندران، ۱۳۰(۲۵), ۱۰۰-۱۱۵.

۴. طلوعی اشلقی. ع و پورابراهیمی. ع و ابراهیمی. م و قاسم احمد. ل. پیش بینی عود مجدد سرطان پستان به کمک سه تکنیک داده کاوی. زمستان ۱۳۹۱. فصلنامهٔ بیماری های پستان ایران.۴. ۲۴-۳۴

5. Swetha, T. L. V. N., & Bindu, C. H... Detection of Breast cancer with Hybrid image segmentation and Otsu's thresholding. (2015, December).In 2015 International Conference on Computing and Network Communications (CoCoNet) 565-570
6. Saritas,I. (2012) Prediction of Breast Cancer Using Artificial Neural Networks. ORIGINAL PAPER
7. Thein, H. T., & Tun, K. M. (2015). AN APPROACH FOR BREAST CANCER DIAGNOSIS. Advanced Computing: An International Journal (ACIJ), 6(1).
8. 8- Saremi, S., Mirjalili, S., & Lewis, A. (2017). Grasshopper optimisation algorithm: Theory and application. Advances in Engineering Software, 105, 30-47.
9. Yip, C. H & ,.Taib, N (2014). Breast health in developing countries . *CLIMACTERIC,17(2), 1-6.*

10. Kelly, K. M., Dean, J., Comulada, S. W., & Lee, S. -J. (2009, march). Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *the European Society of Radiology, 20*(3), 734-742.
11. Ghayoumi Zadeh,H. Haddadnia,J. Hashemian,M. Hassanpour,K. (2012) Iranian Journal of Medical Physics.9.265-274

مطالعات علوم زیستی و زیست فناوری

دوره ۳، شماره ۳، پاییز ۱۳۹۶، صفحات ۵۵- ۴۱

Breast Cancer Detection with meta-heuristic Algorithms

Fateme Ghaemi[1], Iraj Mahdavi[2]

1. Master Student in Industrial Engineering, Mazandaran University of Science and Technology, Iran

2. Associate Professor, Department of Industry, Faculty of Industry, Mazandaran University of Science and Technology, Iran

Abstract

By advancement of science and technology in information technology, several methods have been proposed in health and have been tested and implemented.
Data mining is a very common technique and tool used today in various fields. Diagnosis of various diseases in medical science is considered as one of the most widely used fields of data mining that has been carried out in recent years in many researches and studies. Data mining has been able to take effective steps in this regard with various techniques such as the neural network. Cancer is a terrible disease. Millions of people die every year because of the disease. Breast cancer today is widely evolving among people, and the choice of appropriate treatment and early diagnosis in the treatment process is very necessary. Cancer cells must be detected correctly. Considering the necessity of accurate and timely detection, we have taken steps in this direction and we have been able to increase the accuracy of the diagnosis by using the neural network and metamorphosed algorithm.
In this study we were able to prove by comparing the diagnosis with the genetic algorithm that existed in previous studies. The grasshopper algorithm with the same neural network structure improves the accuracy of tumor detection. In the case study, the accuracy of diagnosis of benign and severe tumors was 98% on average and we were able to increase the diagnostic accuracy by 75.1% using the grasshopper algorithm and reaching 99.75%

Keywords: Breast Cancer- Mammography - Data Mining- Neural Networks - Meta-heuristic algorithms- Grasshopper Optimisation Algorithm