

مقایسه‌ی کارایی الگوریتم‌های داده‌کاوی در تشخیص بیماری تیروئید

محسن رضایی^۱، نازنین زهرا جعفری^۱، حسین غفاریان^۲، مسعود خسروی

فارمد^۳، ایمان ذباح^۴، پروانه دهقان^۵

چکیده

زمینه و هدف: تشخیص به‌موقع عملکرد غیرطبیعی تیروئید و به‌دنبال آن در پیش گرفتن درمان صحیح، می‌تواند باعث کاهش مرگ‌ومیر مرتبط با این بیماری شود. هم‌چنین عدم تشخیص به‌موقع، عوارض جبران‌ناپذیری برای بیمار در پی خواهد داشت. این مطالعه، با هدف تعیین وضعیت غده تیروئید از نظر نرمال بودن، پرکاری یا کم‌کاری با استفاده از تکنیک‌های داده‌کاوی انجام شده است.

روش بررسی: تولید مدل پیش‌بینی کننده به‌منظور طبقه‌بندی بیماری تیروئید، پس از پیش‌پردازش داده‌ها با استفاده از روش‌های نظارت‌شده و بدون ناظر انجام گردید. این مطالعه از نوع تحلیلی بوده و پایگاه داده‌ی آن شامل ۲۱۵ رکورد مستقل مبتنی بر ۵ ویژگی پیوسته و برگرفته‌شده از مرجع داده یادگیری ماشین UCI می‌باشد.

یافته‌ها: در روش نظارت‌شده از شبکه‌های عصبی پرسپترون چندلایه و شبکه عصبی بردار یادگیر و شبکه عصبی فازی و در روش بدون نظارت از خوشه‌بندی فازی استفاده گردید. با روش حداقل مربعات خطا (RMSE) به‌ترتیب دقت‌های ۰/۰۵۵ و ۰/۲۷۴ و ۰/۱۲ و ۱/۰۳۱ حاصل شد.

نتیجه‌گیری: کاهش خطای تشخیص بیماری تیروئید یکی از اهداف محققان بوده است. استفاده از روش‌های مبتنی بر داده‌کاوی می‌تواند به کاهش این خطا کمک کند. در این مطالعه تشخیص بیماری تیروئید به کمک روش‌های مختلف تشخیص الگو صورت گرفت. نتایج نشان می‌دهد که مدل عصبی فازی دارای حداقل میزان خطا و بیشترین دقت است.

واژه‌های کلیدی: بیماری تیروئید، خوشه‌بندی فازی، شبکه عصبی مصنوعی، شبکه‌های عصبی فازی، یادگیری نظارت‌شده

دریافت مقاله: اردیبهشت ۱۳۹۸

پذیرش مقاله: شهریور ۱۳۹۸

* نویسنده مسئول:

ایمان ذباح؛

دانشگاه آزاد اسلامی تربت حیدریه

Email:

Iman.zabbah@iautorbat.ac.ir

۱ دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات، گروه کامپیوتر، موسسه آموزش عالی اترک قوچان، قوچان، ایران

۲ استادیار گروه کامپیوتر، دانشکده مهندسی، دانشگاه اراک، اراک، ایران

۳ استادیار گروه کامپیوتر، موسسه آموزش عالی اترک قوچان، قوچان، ایران

۴ مربی، گروه کامپیوتر، واحد تربت حیدریه، دانشگاه آزاد اسلامی، تربت حیدریه، ایران

۵ استادیار گروه رادیوترایی انکولوژی، دانشگاه علوم پزشکی تربت حیدریه، تربت حیدریه، ایران

مقدمه

آن غده تیروئید به اندازه کافی هورمون تولید نمی‌کند و اگر درمان نشود ممکن است موجب کلاسترول بالا، افزایش فشارخون، عوارض قلبی و عروقی، کاهش باروری و افسردگی شود (۵). پرکاری تیروئید یک بیماری است که در آن بیش از حد هورمون تیروکسین تولید می‌شود، این امر باعث از دست دادن ناگهانی وزن، ضربان قلب نامنظم، عصبی بودن و غیره می‌شود. افزایش فعالیت غده‌ی تیروئید هم موجب افزایش مقدار T4 یا T3 و یا هر دو آن‌ها و بروز بیماری پرکاری تیروئید می‌شود. در پرکاری تیروئید، میزان متابولیسم سلول‌های بدن افزایش می‌یابد. اگرچه تظاهرات بالینی این بیماری تا اندازه‌ای وابسته به سن است ولی شایع‌ترین آن‌ها، عبارت است از: افزایش کارایی قلب، تپش قلب، افزایش کارایی روده که منجر به افزایش حرکات روده و اسهال می‌شود و افزایش متابولیسم که موجب کاهش وزن می‌گردد. مطالعات نشان داده است که پرکاری تحت بالینی تیروئید به‌ویژه در افراد مسن ممکن است با عوارض قلبی و یا پوکی استخوان همراه باشد (۶). یکی از مسایل اصلی برای ایجاد چالش در علوم پزشکی درحال توسعه، فن‌آوری تشخیص بیماری با دقت بالا است. داده‌کاوی یکی از گام‌های اساسی در کشف دانش است و نقش حیاتی در زمینه پزشکی برای تشخیص بیماری دارد (۷). تاکنون تحقیقات و مطالعات زیادی در زمینه تشخیص اختلالات تیروئیدی انجام شده است.

در نهایت اینکه، بیماری تیروئید یکی از شایع‌ترین بیماری‌های شناخته شده حال حاضر است. عدم بررسی هر کدام از اختلالات تیروئیدی می‌تواند باعث عوارض خطرناکی شود. این پژوهش، با هدف تشخیص نوع بیماری تیروئید (کم‌کاری و پرکاری) و یا عدم بیماری (نرمال بودن) با استفاده از متدهای مبتنی بر داده‌کاوی انجام شده است. بدین منظور از دو روش نظارت‌شده شبکه‌های عصبی مصنوعی و شبکه‌های عصبی فازی و همچنین روش بدون ناظر خوشه‌بندی فازی استفاده شده است. مطالعه‌ی حاضر نشان می‌دهد که استفاده از روش نظارت شده به‌خصوص زمانی که به عدم قطعیت پارامترهای ورودی مدل توجه می‌شود، منجر به تشخیص دقیق‌تر بیماری تیروئید می‌شود.

روش بررسی

تحقیق حاضر، حاصل یک مطالعه‌ی تشخیصی است که بر اساس متغیرهای ورودی به پیش‌گویی وضعیت بیماری تیروئید می‌پردازد. در این پژوهش از مجموعه داده‌ی موجود در انبار داده یادگیری ماشین دانشگاه

امروزه اختلالات تیروئیدی یکی از شایع‌ترین بیماری‌های مزمنی می‌باشد که جمعیت زیادی را مبتلا کرده است. مطالعات تحقیقاتی مختلف نشان می‌دهد که تعداد زیادی از افراد در جوامع مختلف به این بیماری دچار می‌شوند؛ از این رو ایجاد آگاهی در بین مردم در مورد علائم و نوع این بیماری و تشخیص آن نقش و اهمیت حیاتی دارد (۱). تست‌های آزمایشگاهی افراد مسن یا زنان حامله و یا کسانی که به علت داشتن بیماری‌های دیگر دارو مصرف می‌کنند و یا بیمارانی که در شرایط خاص پزشکی دیگری قرار دارند، مشکل است. گروه‌هایی که بیشترین تأثیر را از عملکرد بد تیروئید می‌پذیرند، سالمندان و زنان هستند که ۵ تا ۱۰ درصد آنان از این بیماری متأثر می‌شوند (۲). تیروئید، یکی از مهم‌ترین غدد درون‌ریز در بدن است و غده‌ای است پروانه‌ای شکل که در زیر گردن یافت می‌شود. این غده مسئول تنظیم فعالیت‌های سوخت‌وساز بدن است. ترشح آن عمدتاً در کنترل هورمون محرک تیروئید (Thyroid-stimulation hormone) است که از غده هیپوفیز قدامی ترشح می‌شود (۳). هنگامی که غده تیروئید به‌درستی عمل کند دو هورمون به نام تری‌یدوتیرونین (Triiodothyronine) و تیروکسین (Thyroxine) تولید می‌کند. هنگامی که هورمون‌های T3 و T4 به اندازه کافی تولید نشوند، کم‌کاری تیروئید رخ می‌دهد. افراد در اثر ابتلا به این بیماری ممکن است دچار اختلالات ذهنی، اختلالات تنفسی، مشکلات قلبی، نوسان در دمای بدن و بالاخره مرگ ناشی از عدم درمان به‌موقع شوند. اگر فعالیت غده‌ی تیروئید دچار اختلال شود، می‌تواند بر هر سیستمی از بدن مانند دستگاه گوارش و حتی سیستم اعصاب مرکزی تأثیر می‌گذارد، به‌همین دلیل اختلال در فعالیت غده‌ی تیروئید با عوارض وسیعی همراه است. بی‌نظمی شدید تیروئید نباید نادیده گرفته شود؛ به‌طور مثال حمله تیروئیدی که یک عارضه جانبی پرکاری شدید است و کمای می‌گردد که نتیجه‌ی درمان نکردن کم‌کاری تیروئید است، حتی ممکن است منجر به مرگ بیمار شود (۳). دو بیماری شایع غده تیروئید، کم‌کاری و پرکاری تیروئید است که بنا بر آمار جهانی تعداد مبتلایان به کم‌کاری تیروئید بیش از سایر بیماری‌های مرتبط با این غده و سهم بانوان در ابتلا به آن بیش از مردان است (۳). علاوه بر بیماری‌های تیروئید شامل کم‌کاری و پرکاری، بیماری‌های شایع دیگر شامل بزرگی غده تیروئید، توده‌های تیروئید، سرطان تیروئید و التهاب تیروئید نیز وجود دارند (۴). کم‌کاری تیروئید یک بیماری است که در

مبتلا به پرکاری تیروئید و ۱۳/۹٪ مبتلا به کم کاری تیروئید بوده‌اند. جدول ۱ آمارهای توصیفی پایگاه داده را به تفکیک ویژگی نشان می‌دهد. سطرهای این ماتریس، پارامترهایی هستند که می‌توانند کم کاری و یا پرکاری و یا نرمال بودن بیماری را تعیین نمایند.

ایروین، کالیفرنیا استفاده شده است (۸). بانک اطلاعاتی این منبع شامل ۲۱۵ نمونه با ۵ پارامتر پیوسته و ۳ کلاس کم‌کار و پرکار و سالم می‌باشد. در نهایت، داده‌های این منبع را می‌توان یک ماتریس ۲۱۵×۵ در نظر گرفت. از این تعداد ۶۹/۷٪ این بیماران یعنی ۱۵۰ نفر نرمال و ۱۶/۲٪ یعنی ۳۵ نفر

جدول ۱: پارامترهای بالینی بیماری تیروئید

ویژگی	دامنه تغییرات	معیار انحراف استاندارد	حداقل مقدار	حداکثر مقدار	ضریب همبستگی پیرسون
T3resin	۶۵-۱۴۴	۱۸/۳۳۲۲±۱۰۸/۱۷۸	۱/۲	۱۴۴	۰/۰۴۶
Thyroxin	۰/۵-۲۵/۳	۴/۷۰۸±۹/۸۰۳	۰/۵	۲۵/۳	۰/۰۲۷
Triiodothyronine	۰/۲-۱۰	۲/۰۵۰±۱/۴۲۲ ۲/۰۵۰±۱/۴۲۲	۰/۵	۱۰/۰	۰/۰۱۰
Thyroid Stimulating	۰/۱-۵۶/۴	۲/۹۳۱±۶/۱۴۸	۰/۱	۵۶/۴	-۰/۰۸۹
TSH_Value	۰/۷-۵۶/۳	۴/۲۱۲±۸/۰۸۴	-۰/۷	۵۶/۳	۰/۰۵۶

*N=۲۱۵ ** Significant at p<۰/۰۱

ساختارهایی در بین متغیرهاست که روش خوشه بندی (Clustering) به‌عنوان معمول ترین روش این حالت شناخته می‌شود (۹ و ۱۰).

ب- روش با نظارت شبکه‌های عصبی مصنوعی

شبکه عصبی مصنوعی که الهام گرفته از مغز انسان است به‌عنوان یکی از روش‌های با ناظر مورد استفاده قرار گرفته است. این شبکه، در واقع، پردازش داده‌ها را به عهده پردازنده‌های کوچک و بسیار زیادی می‌سپارد که به‌صورت شبکه‌ای به هم متصل بوده و موازی با یکدیگر رفتار می‌کنند تا یک مسئله را حل نمایند. بعد با ایجاد شبکه‌ای بین این گره‌ها و اعمال یک الگوریتم آموزشی به آن، شبکه را آموزش می‌دهند. در این شبکه عصبی گره‌ها دارای دو حالت فعال و غیرفعال هستند و هر یال (سیناپس یا ارتباط بین گره‌ها) دارای یک وزن می‌باشد.

یکی از ساده‌ترین و کارآمدترین ساختارهای شبکه‌های عصبی، مدل پرسپترون چندلایه (Multilayer perceptron) است که از یک لایه ورودی، یک یا چندلایه پنهان و یک لایه خروجی تشکیل یافته است. در این ساختار، تمام نرون‌های یک لایه به تمام نرون‌های لایه بعد متصل هستند. شکل ۱- الف ساختار شبکه عصبی مورد استفاده در این پژوهش را نشان می‌دهد.

در این شبکه از دو نوع سیگنال استفاده می‌شود: نوع اول، سیگنال‌هایی هستند که بر اساس ورودی‌های هر نرون و پارامترهای وزن و تابع محرک نظیرش محاسبه می‌شوند و نوع دوم، سیگنال‌های خطا هستند که با برگشت از لایه خروجی و منشعب شدن به لایه‌های پنهانی دیگر

ستون آخر جدول ضرایب همبستگی پیرسون، نسبت به وضعیت بیماری تیروئید را نشان می‌دهد که بیش‌ترین ضریب همبستگی مربوط به متغیر Triiodothyronine با مقدار ۰/۱ می‌باشد.

با توجه به آنکه ورود داده‌ها به شکل خام، باعث کاهش سرعت و دقت سیستم‌های یادگیر می‌شود، برای پیشگیری از چنین شرایطی و همچنین به لحاظ یکسان کردن ارزش داده‌ها، نرمال کردن آن‌ها، قبل از اعمال به سیستم انجام شده است. عمل نرمال‌سازی داده‌ها مطابق رابطه (۱) انجام گرفته است:

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (\text{رابطه ۱})$$

در این رابطه، x داده اولیه، x_n داده نرمال شده، x_{max} و x_{min} به ترتیب حداقل و حداکثر داده‌های ورودی هستند.

الف- داده‌کاوی

به فرایند استخراج دانش ناشناخته، داده‌کاوی گفته می‌شود. داده‌ها اغلب حجیم، اما بدون ارزش می‌باشند. دانش نهفته در داده‌ها آن‌ها را قابل استفاده می‌کند. در این مطالعه، از هر دو روش بدون ناظر (Unsupervised Learning) و با ناظر (Supervised Learning) به منظور تشخیص بیماری تیروئید استفاده شده است. در روش با ناظر، متغیرهای هدف معلوم است و تلاش برای کشف الگوریتمی است که بتواند ارتباط بین پارامترهای ورودی و خروجی را پیدا کند؛ اما در روش بدون ناظر متغیرهای هدف نامعلوم می‌باشند. در این حالت، تلاش برای کشف

از آنجایی که وزن‌های شبکه عصبی به صورت تصادفی انتخاب می‌شوند و این مسئله موجب تغییر در خروجی شبکه در هر بار اجرا می‌شود لذا در این مطالعه حدود ۴۰۰ نوع شبکه عصبی بررسی شد و میانگین خطای آن‌ها در شکل ۱-د ثبت گردید. میزان خطای هر شبکه از روش میانگین مربعات خطا (Mean Square Error) محاسبه شده است. جدول ۲ نتایج معماری‌های مختلف شبکه عصبی و نیز کمترین خطای مربوط به هر شبکه را نشان می‌دهد.

محاسبه می‌شوند. تعداد نرون‌های لایه پنهان بستگی به نظر طراح شبکه دارد و معمولاً با سعی و خطا به دست می‌آید. خروجی شبکه در لایه آخر با رابطه ۲ محاسبه می‌شود که در آن h و o به ترتیب نشان‌دهنده‌ی لایه‌ی پنهان و لایه‌ی خروجی بوده و منظور از w همان وزن‌های لایه‌هاست.

$$o_i = \text{sgm}(\sum_m \text{sgm}(\sum_i x_i w_{im}^h) w_{mi}^o) \quad (\text{رابطه } 2)$$

تابع انتقال از نوع سیگموئید است و به صورت رابطه‌ی ۳ تعریف

می‌گردد:

$$\text{sgm}(x) = \frac{1}{1+e^{-x}} \quad (\text{رابطه } 3)$$

جدول ۲: بررسی معماری‌های مختلف شبکه عصبی با در نظر گرفتن کلیه پارامترهای بیماری تیروئید

معماری شبکه عصبی	تعداد لایه مخفی	شبکه در بین ۱۰۰ مدل (RMSE) کمترین خطای
۵-۱۰-۱-۱	۲	۰/۰۵۷۹
۵-۲۰-۱-۱	۲	۰/۰۵۴۵
۵-۳۰-۱-۱	۲	۰/۰۵۶۰
۵-۴۰-۱-۱	۲	۰/۰۵۶۷

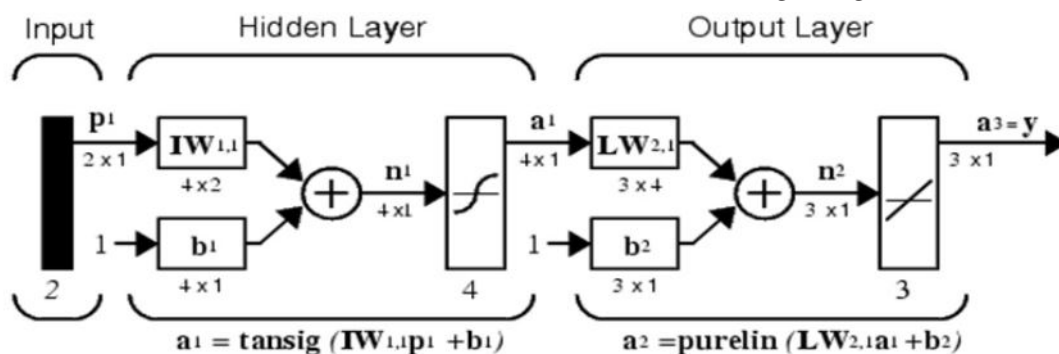
* i بین ۱۰ تا ۱۰۰ نرون

بردار ورودی می‌باشند. هردو لایه رقابتی و خطی دارای یک نرون به‌ازای هر کلاس هستند (۱۱).

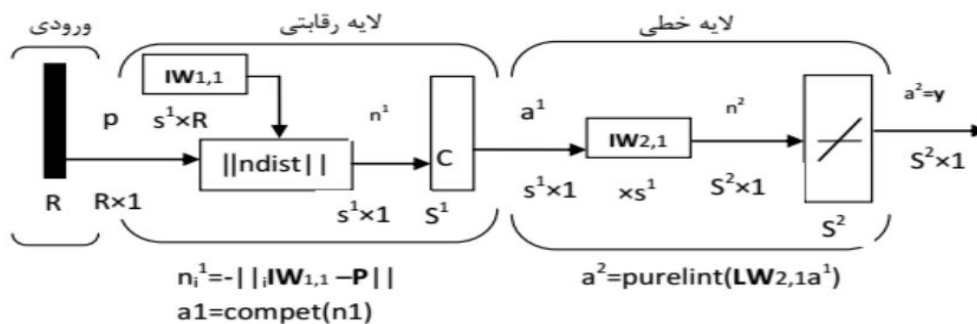
در این مطالعه از ۱۰ شبکه عصبی LVQ استفاده شد تا بهترین معماری شبکه جهت تشخیص بیماری شناسایی شود؛ و میانگین خطای شبکه‌ها ثبت گردید. در تمامی شبکه‌ها از ۸۰٪ داده جهت آموزش شبکه و ۱۰٪ آن‌ها جهت اعتبارسنجی و ۱۰٪ به‌منظور آزمون استفاده گردید. تعداد تکرارها برای تمامی آن‌ها ۳۰ و الگوریتم یادگیری از نوع (Learning Vector Quantization) انتخاب شده است. نتایج نشان می‌دهد که الزاماً با افزایش تعداد نرون‌های لایه رقابتی وضعیت شبکه بهتر نشده است. در نهایت میانگین خطا با مقدار ۰/۲۷۴ به‌دست آمد که در شکل ۱-ج نشان داده شده است.

نتایج جدول ۲ نشان می‌دهد که تعداد نرون‌ها در لایه‌ی مخفی اول متغیر بین ۱۰ تا ۴۰ و در لایه‌ی مخفی دوم متغیر بین ۱ تا ۱۰۰ است، از نظر خطا تغییر محسوسی ندارد؛ اما معماری ۱-۹۲-۲۰-۵ که از نظر زمان محاسباتی نسبت به بقیه عملکرد بهتری دارد، انتخاب شده است. ضمن اینکه میانگین خطای شبکه‌ها ۰/۰۵۵ به‌دست می‌آید.

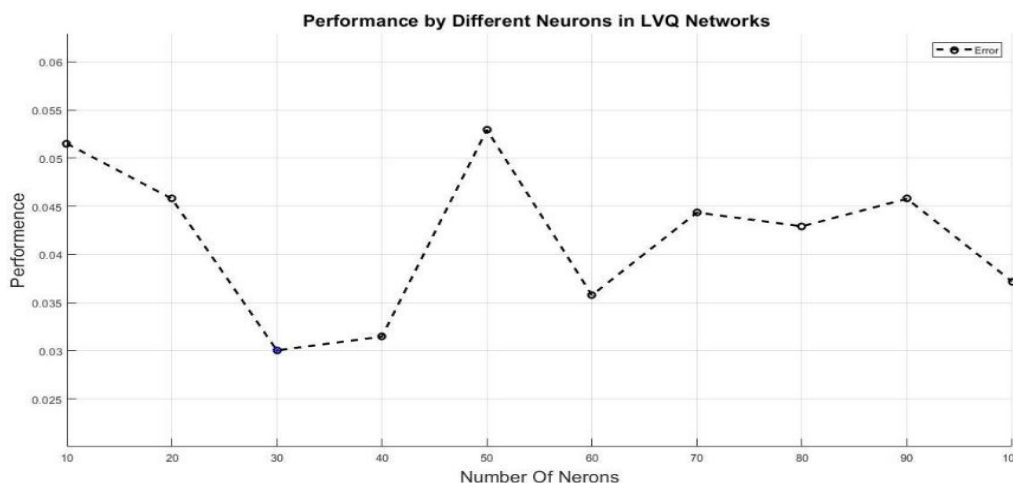
علاوه بر شبکه عصبی پرسپترون (MLP) شبکه عصبی بردار یادگیر (LVQ) نیز جهت تشخیص بیماری بررسی گردید. شبکه عصبی LVQ دارای دو لایه‌ی رقابتی و خطی است. لایه‌ی رقابتی دسته‌بندی کردن بردارهای ورودی را یاد گرفته و در نهایت لایه‌ی خطی کلاس‌های لایه‌ی رقابتی را به دسته‌های هدف که توسط کاربر تعیین شده نگاشت می‌کند. شکل ۱-ب معماری شبکه‌های LVQ را نشان می‌دهد. در این شکل S1 و S2 به ترتیب تعداد نرون‌های لایه رقابتی و خطی و R تعداد عضوهای



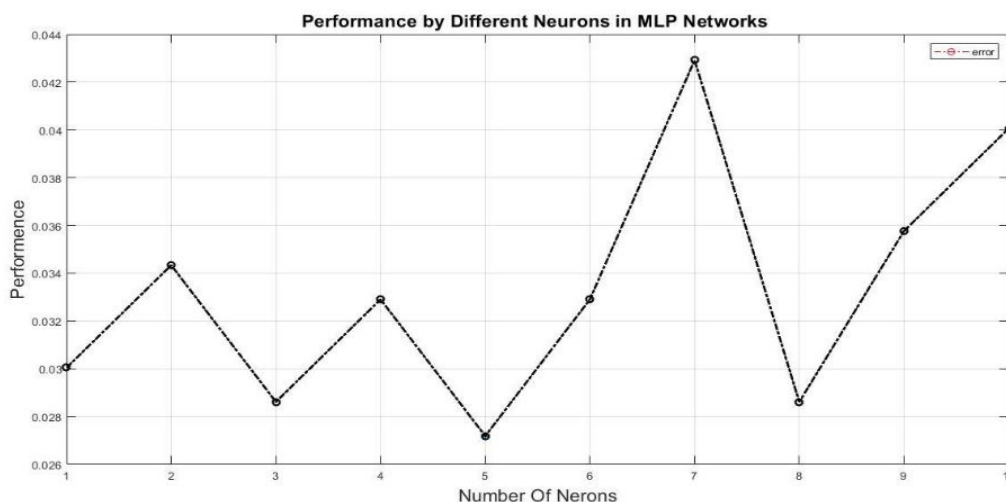
شکل ۱- الف: معماری شبکه عصبی پرسپترون (MLP)



شکل ۱-ب: معماری شبکه عصبی بردار یادگیر (LVQ)



شکل ۱-ج: نمودار فضای شبکه عصبی پرسپترون (LVQ) در تشخیص بیماری تیروئید با تغییر تعداد نرون در لایه پنهان



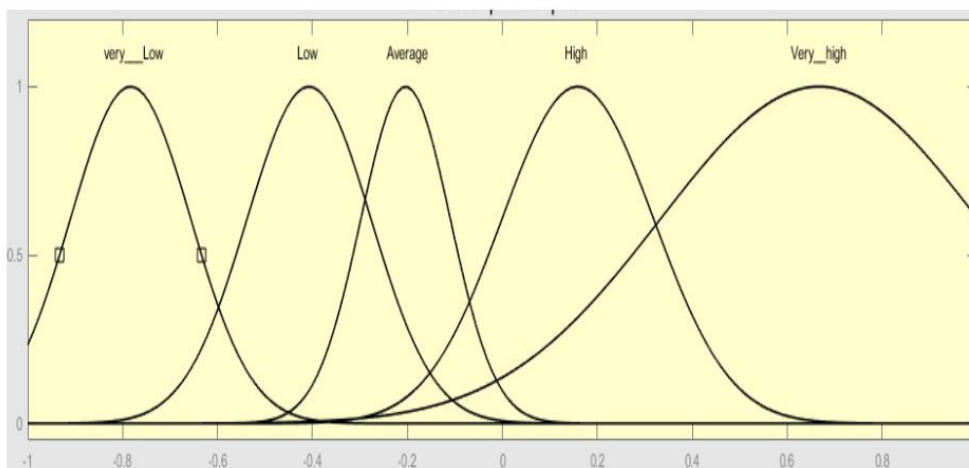
شکل ۱-د: نمودار فضای شبکه عصبی پرسپترون (MLP) در تشخیص بیماری تیروئید با تغییر تعداد نرون در لایه پنهان

در کلاس‌های دقیق و جداگانه مورد استفاده قرار گیرند (۱۲). الگوریتم طبقه‌بندی عصبی فازی مورد استفاده در این پژوهش یک شبکه‌ی ویژه‌ی اشتراک به جلو سه لایه‌ای است که:

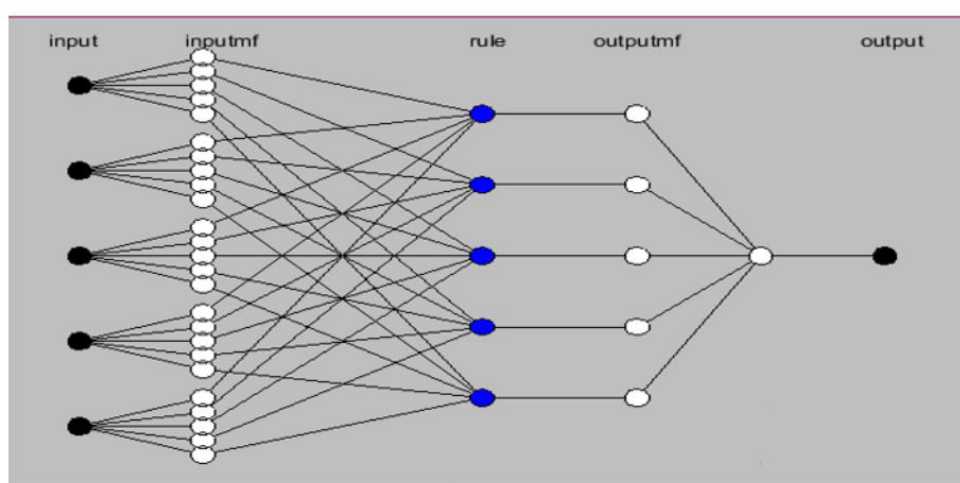
- اولین لایه نشان‌دهنده‌ی متغیرهای ورودی است.
 - لایه‌ی پنهان نشان‌دهنده‌ی قوانین فازی است.
 - لایه‌ی سوم نشان‌دهنده‌ی متغیرهای خروجی هر کلاس است.
- واحدها از T_norm و T_conorm به‌عنوان فعال‌سازی توابع استفاده می‌کنند. شکل ۲ مدل طراحی شده را نشان می‌دهد.

ج- روش با نظارت عصبی فازی (ANFIS)

مدل شبکه عصبی-فازی، یک مدل ترکیبی است که قابلیت‌های منحصر به فردی در طبقه‌بندی دارد. در واقع ترکیب دو مدل شبکه عصبی و سیستم فازی سبب افزایش دقت در کلاس‌بندی مجموعه‌هایی با داده‌های مبهم و کم‌دقت می‌شود. تفاوت اصلی شبکه فازی-عصبی با شبکه‌های عصبی مصنوعی در آن است که وزن‌های شبکه عصبی-فازی به‌صورت غیرقطعی تعریف می‌شوند نه قطعی. الگوریتم طبقه‌بندی عصبی فازی از مجموعه‌ای از داده‌ها، قواعد فازی را استخراج می‌کند که می‌توانند



شکل ۲-الف: توابع عضویت مربوط به متغیر Resin

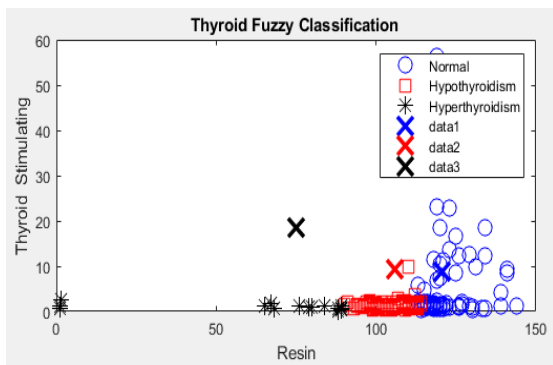


شکل ۲-ب: تعداد نرون‌ها و تعداد لایه‌های شبکه عصبی فازی به منظور تشخیص بیماری تیروئید. دایره‌های مشکی به معنی ورودی و خروجی، دایره‌های آبی نشان دهنده قوانین، دایره‌های سفید بیانگر توابع انتقال

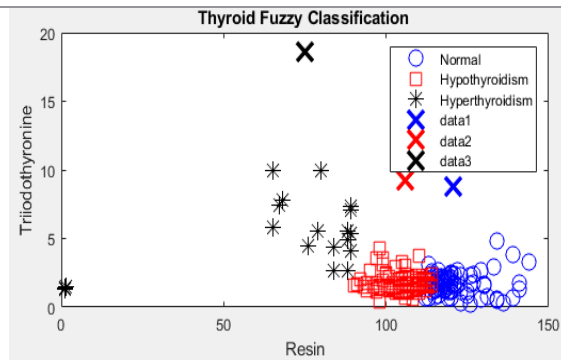
ممکن است علائم یک بیماری مانند تیروئید به‌طور قطع متعلق به یکی از ۳ کلاس کم‌کار و پرکار و نرمال نباشد. بلکه می‌تواند به کلاس کم‌کار به میزان ۰/۰۳ و پرکار به میزان ۰/۰۶ و به نرمال به میزان ۰/۱ متعلق باشد. در خوشه‌بندی کلاسیک هر نمونه‌ی ورودی متعلق به یک و فقط یک خوشه است و نمی‌تواند عضو دو خوشه و یا بیش‌تر باشد؛ اما در خوشه‌بندی فازی یک نمونه می‌تواند متعلق به بیش از یک خوشه باشد. در این پژوهش در آخرین مرتبه‌ی مدل‌سازی از روش غیرنظارت‌شده‌ی خوشه‌بندی فازی استفاده شده است. جهت مشاهده پراکندگی نمونه‌ها شکل ۳ ترسیم شده است که وضعیت متغیر اول را نسبت به سایر متغیرها نشان می‌دهد. نقاطی که با \times مشخص شده‌اند مرکز هر خوشه را نشان می‌دهند.

طبق شکل ۲ تعداد ۵ متغیر زبانی با توجه به ویژگی‌های مندرج در جدول ۱ در طراحی سیستم در نظر گرفته شده است. تعداد توابع عضویت برای هر یک از متغیرهای زبانی بین ۲ تا ۶ متغیر می‌باشد. به‌عنوان نمونه شکل ۲-الف وضعیت توابع عضویت متغیر Resin شامل مقادیر: «خیلی کم»، «کم»، «متوسط»، «زیاد» و «خیلی زیاد» را نشان می‌دهد. برای سایر متغیرها نیز به‌همین ترتیب توابع عضویت ساخته شده است؛ در این پژوهش از روش غیرفازی سازی، میانگین مراکز (Centroid) و ساختار ممدانی (Mamdani) استفاده شده است. شکل ۲-ب لایه‌های مورد استفاده در سیستم فازی در این مطالعه را نشان می‌دهد. در تشخیص این بیماری توسط سیستم عصبی فازی با تعداد تکرار ۶۰، خطای ۰/۲۱۱۷ حاصل گردید.

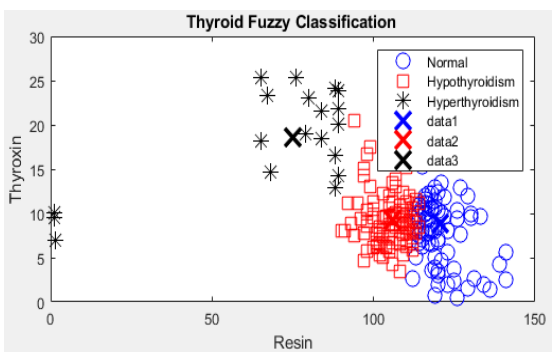
ه- روش بدون ناظر خوشه‌بندی فازی (Fuzzy Clustering)



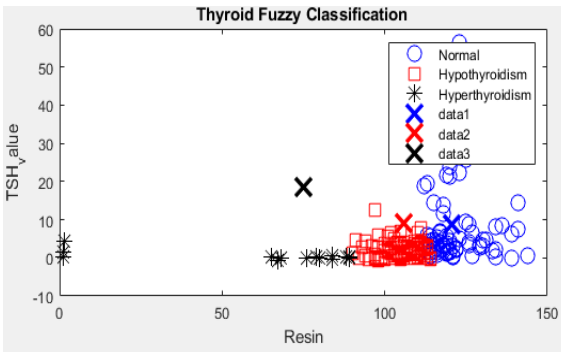
شکل ۳- ب: وضعیت پراکندگی نمونه‌ها بر اساس دو ویژگی Thyroid Stimulating و Resin



شکل ۳- الف: وضعیت پراکندگی نمونه‌ها بر اساس دو ویژگی Triiodothyronine و Resin



شکل ۳- د: وضعیت پراکندگی نمونه‌ها بر اساس دو ویژگی Thyroxin و Resin



شکل ۳- ج: وضعیت پراکندگی نمونه‌ها بر اساس دو ویژگی TSH_value و Resin

شکل ۳: پراکندگی متغیر نمونه‌ها نسبت به ویژگی‌های مختلف

است. در این پژوهش، سعی شده است برای بهبود تشخیص بیماری تیروئید به بررسی روش‌های با ناظر و بدون ناظر پرداخته شود. در اولین مرتبه، مدل سازی از یک شبکه عصبی پرسپترون چندلایه با الگوریتم پس انتشار خطا به منظور کلاس بندی داده‌ها به سه دسته تیروئید نرمال و کم‌کار و تیروئید پرکار استفاده گردید و ۵ فاکتور مربوط به بیماری تیروئید به شبکه عصبی اعمال شد. آموزش شبکه با ۷۰٪ داده‌ها و تست آن با ۳۰٪ از داده‌ها انجام گردید. هم‌چنین معماری‌های مختلف شبکه عصبی برای حصول بهترین کلاس بندی بررسی گردید؛ در مرتبه دوم مدل سازی، از شبکه عصبی LVQ استفاده گردید. علاوه بر بررسی شبکه‌های عصبی معماری‌های مختلف، تعداد ورودی‌های بیماری در ورودی شبکه عصبی نیز بررسی شد و میانگین خطای شبکه‌ها محاسبه گردید. در مرحله سوم از شبکه‌های عصبی فازی استفاده شده است. الگوریتم آموزش برای همگی از نوع سوگنو (یکی از معروف ترین الگوریتم‌های آموزشی سیستم عصبی فازی) انتخاب گردید. کمترین خطای به دست آمده مربوط به زمانی بود که از ۷ تابع عضویت استفاده شده بود. جدول ۳ انواع معماری‌های مورد استفاده در شبکه عصبی فازی و دقت آن‌ها را نشان می‌دهد.

یافته‌ها

در این پژوهش جهت تشخیص بیماری تیروئید از دو روش اصلی داده‌کاوی یعنی روش با ناظر و بدون ناظر استفاده شده است. داده‌های این پژوهش از پایگاه داده UCI شامل ۲۱۵ نمونه، هر یک شامل ۵ ویژگی، تأمین گردید. پارامترها پیوسته بوده و به ترتیب میانگین و انحراف معیار: TSH برابر $۸/۰۸۴۷ \pm ۴/۲۱۲$ ، T3 برابر $۱۰۸/۱۷۸ \pm ۱۸/۳۳۲۲$ ، TT4 (تترایدوتیرونین) برابر $۲/۰۵۰ \pm ۴/۲۲۸$ ، T4U برابر $۹/۸۰۳ \pm ۴/۷۰۸۳$ و Thyroid Stimulating برابر با $۲/۹۳۱ \pm ۶/۱۴۸۷$ دارند. ضریب همبستگی پیروسون برای TSH برابر $۰/۰۵۶$ ، میانگین T3 برابر $۰/۰۴۶$ ، میانگین TT4 برابر $۰/۱۰$ میانگین T4U برابر $۰/۰۲۷$ و Thyroid Stimulating برابر با $۰/۰۸۹$ می‌باشد.

هدف نهایی یک سیستم شناسایی الگو رسیدن به بالاترین نرخ طبقه‌بندی ممکن برای مسئله مورد نظر است. هیچ الگوریتم طبقه‌بندی وجود ندارد که به‌تنهایی و به‌طور کامل برای تمام مسایل مناسب باشد. در مطالعات مختلف با عنوان داده‌کاوی پزشکی، راهکارهای متعددی جهت کشف روابط بین عوامل بیماری تیروئید ارایه شده

جدول ۳: معماری‌های شبکه عصبی فازی مورد استفاده در این مطالعه

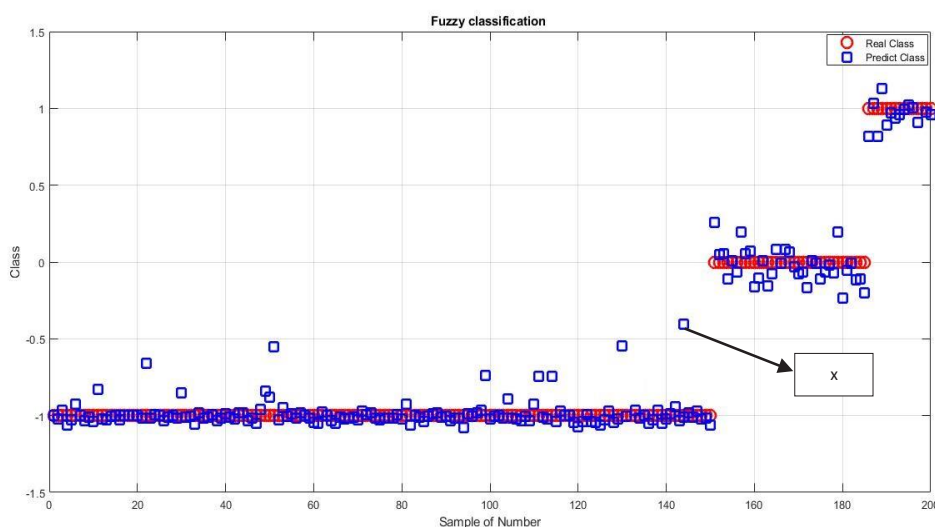
ردیف	تعداد توابع عضویت	الگوریتم مورد استفاده	مقدار خطا RMSE
۱	۴	سوگنو	۰/۲۳
۲	۵	سوگنو	۰/۲۱
۳	۱۰	سوگنو	۰/۱۳
۴	۲۰	سوگنو	۰/۰۵
۵	۳۰	سوگنو	۰/۰۰۹

* در تمام موارد نوع تابع عضویت گوسی و تعداد تکرار ۱۰۰ می‌باشد.

کلاس تعلق دارند.

شکل ۴ نمودار خطای کلاس‌بندی شبکه عصبی فازی را نشان

می‌دهد. در واقع هر نمونه در کلاس‌بندی فازی به اندازه‌ای به یک



شکل ۴: میزان خطای کلاس‌بندی در شبکه عصبی نرو فازی

مدل‌سازی از روش کلاس‌بندی بدون ناظر، استفاده گردید؛ که نتیجه برای

به‌عنوان مثال نمونه X، تقریباً به همان اندازه‌ای که به کلاس نرمال

۳۰ نمونه در جدول ۴ ثبت گردید.

تعلق دارد به کلاس تیروئید کم‌کار نیز تعلق دارد. در مرحله‌ی پایانی

جدول ۴: کلاس‌بندی فازی در تشخیص بیماری تیروئید برای ۳۰ نمونه مختلف

شماره نمونه	میزان تعلق به کلاس پرکار	میزان تعلق به کلاس کم‌کار	میزان تعلق به نرمال
۱	۰/۰۲۸	۰	۰/۹۸
۲	۰/۵	۰/۰۲	۰/۴۸
۳	۰/۸۲	۰/۰۳	۰/۱۵
۴	۰/۱۴	۰/۰۲	۰/۸۴
۵	۰/۰۴	۰/۰۱	۰/۹۵
۶	۰/۱	۰/۰۲	۰/۸۷
۷	۰/۱۳	۰/۰۱	۰/۸۵
۸	۰/۶۱	۰/۰۲	۰/۳۸
۹	۰/۰۳	۰/۰۱	۰/۹۷
۱۰	۰/۰۷	۰/۰۱	۰/۹۱

۰/۸۳	۰/۰۲	۰/۱۵	۱۶۰
۰/۲۱	۰/۱۰	۰/۷۸	۱۶۱
۰/۶۴	۰/۰۲	۰/۳۴	۱۶۲
۰/۰۷	۰/۰۱	۰/۹۲	۱۶۳
۰/۹۱	۰/۰۱	۰/۰۸	۱۶۴
۰/۷۷	۰/۰۲	۰/۲۱	۱۶۵
۰/۴۲	۰/۰۲	۰/۵۶	۱۶۶
۰/۱۴	۰/۰۱	۰/۸۵	۱۶۷
۰/۳۵	۰/۰۲	۰/۶۳	۱۶۸
۰/۹۱	۰/۰۳	۰/۰۶	۱۶۹
۰/۲۷	۰/۰۵۳	۰/۲	۱۷۰
۰/۹۴	۰/۰۲	۰/۰۴	۲۰۰
۰/۱۳	۰/۰۱	۰/۸۶	۲۰۱
۰/۸۶	۰/۰۵	۰/۰۹	۲۰۲
۰/۹۹	۰	۰/۰۱	۲۰۳
۰/۳	۰/۰۲	۰/۶۸	۲۰۴
۰/۹۷	۰/۰۱	۰/۰۳	۲۰۵
۰/۶۳	۰/۰۴	۰/۳۳	۲۰۶
۰/۰۹	۰/۰۱	۰/۹	۲۰۷
۰/۲۷	۰/۰۲	۰/۲۱	۲۰۸
۰/۷۵	۰/۰۲	۰/۲۳	۲۰۹
۰/۵۱	۰/۳۶	۰/۱۳	۱۲۰

آنچه در جدول ۴ مشهود است، این است که به عنوان مثال نمونه اول که در خوشه بندی کلاسیک کاملاً به خوشه نرمال تعلق دارد، در خوشه بندی فازی به میزان ۰/۹۸ به خوشه نرمال و به میزان ۰/۰۲ به مجموعه پرکار و به میزان صفر به مجموعه کم کار تعلق دارد. هم چنین نمونه آخر به میزان ۰/۵۱ به نرمال و ۰/۳۶ به کم کار و به میزان ۰/۱۳ به پرکار تعلق دارد. لذا پزشک در تشخیص نوع بیماری این نمونه می تواند با دقت نظر بیش تری

نسبت به روند درمان آن اقدام کند. نگاه فازی به بسیاری از بیماری هایی که در آن هدف پزشک کلاس بندی بیمار به یکی از کلاس های نرمال یا غیرنرمال است، مانند تشخیص ابتلا و یا عدم ابتلا به بیماری کرونر قلب و یا تشخیص سرطان خوش خیم یا بدخیم در پستان، ممکن است موجب بهبود روند تشخیص شود. جدول ۵ مقایسه بین روش های مختلف در این پژوهش را نشان می دهد:

جدول ۵: مقایسه عملکرد میانگین روش های مختلف در این مطالعه بر مبنای مربعات فضا (RMSE)

میانگین خطا در شبکه عصبی پرسپترون چندلایه MLP	میانگین خطا در شبکه عصبی LVQ	میانگین خطا در شبکه نرو فازی ANFIS	خوشه بندی فازی Cmean
۰/۰۵۵	۰/۲۷۴	۰/۰۱۲	۱/۰۳

بحث

همان طور که مشاهده می شود روش نروفازی توانسته است با دقت بالاتری پیش بینی بیماری تیروئید را نسبت به سایر روش ها انجام دهد.

نتیجه ی مطالعه، حاکی از آن است که استفاده از کلاس بندی های

مبتنی بر یادگیری بدون ناظر، مانند خوشه‌بندی فازی و کلاس‌بندی‌های نظارت‌شده، مانند شبکه عصبی و نروفازی عملکرد قابل قبولی دارند. لیکن در مسئله تشخیص بیماری تیروئید، استفاده از متدهای مبتنی بر یادگیری نظارت‌شده به‌خصوص یادگیری‌های نظارت‌شده‌ای که با درجه‌ای از عدم قطعیت به حل مسئله می‌پردازند، مانند شبکه‌های نروفازی در مقایسه با کلاس‌بندی‌های کلاسیک، عملکرد بهتری دارند (۱۲). در پژوهشی که توسط میراحمدی‌زاده و همکاران جهت تشخیص بیماری تیروئید انجام شده است از شبکه عصبی خودسازمانده و منطق فازی استفاده شده است. در آن مطالعه، یک شبکه عصبی فازی که طبقه‌بندی الگوها و انتخاب ویژگی‌ها را ممکن می‌سازد معرفی شده است و شامل یادگیری نظارت‌شده است. مطالعه ایشان توانایی شبکه بردار ویژگی وزن داده شده (FWD) برای طبقه‌بندی الگو و انتخاب ویژگی را نشان می‌دهد. با نشان دادن دقت خروجی $94/2\%$ در PLVQ و دقت $91/3\%$ در LVQ برتری PLVQ نسبت به LVQ را نشان داده شده است (۱۳).

هم‌چنین در پژوهش عسگری و خسروی‌دانش، تشخیص بیماری تیروئید با سیستم خبره انجام شده است که ESTDD نام‌گذاری شده است. ایشان با این روش توانسته‌اند، بیماری تیروئید را با دقت $95/44\%$ تشخیص دهند (۱۴).

فرقانی و همکاران با استفاده از روش شبکه فازی عصبی بر پایه‌ی تحلیل تفکیک‌کننده‌ی خطی برای دسته‌بندی نمونه‌های تیروئید استفاده کرده‌اند. به این ترتیب که از روش تحلیل تفکیک‌کننده‌ی خطی برای نگاشت فضای ورودی به فضای دیگر به‌کار برده‌اند. در فضای جدید ویژگی‌هایی انتخاب می‌شوند که دارای قدرت تفکیک‌کنندگی بسیار بالایی هستند. پس از استخراج ویژگی، شبکه را با نمونه‌های آموزشی نگاشت یافته آموزش داده، سپس با نمونه‌های تست، شبکه تست می‌شود. در این سیستم از هر دو نوع ساختار ممدانی و سوگنو و هر دو نوع داده پیوسته و گسسته نیز استفاده شده است (۱۵). ادیبی و همکاران با استفاده از ترکیب دو روش فازی و عصبی که از قابلیت یادگیری و پردازش موازی شبکه‌های عصبی و استنتاج تقریبی فازی بر پایه تحلیل تفکیک‌کننده خطی استفاده می‌کند یک سیستم هوشمند جهت تشخیص بیماری تیروئید به‌وجود آورده‌اند (۱۶).

Senol و Yildirim یک ساختار ترکیبی از شبکه‌های عصبی مصنوعی و منطق فازی ایجاد کردند. در الگوریتم پیشنهادی آن‌ها عملکرد سه ساختار ترکیبی شبکه عصبی پرسپترون (MLP)، منطق فازی (Fuzzy) و شبکه عصبی مقطع مخروطی فازی (CSFNN-Fuzzy) و شبکه عصبی تابع شعاعی پایه (RBF)، منطق فازی (Fuzzy) باهم مقایسه شده است (۱۷). در پژوهشی دیگر یک سیستم خبره مبتنی بر قوانین فازی برای بیماری تشخیص تیروئید ارائه شد. این سیستم پیشنهادی شامل سه مرحله است: (انتخاب ویژگی) قبل از پردازش، طبقه‌بندی فازی عصبی و سیستم ارزیابی. در مرحله طبقه‌بندی فازی-عصبی، تولید قوانین فازی اولیه توسط الگوریتم نزدیک‌ترین همسایه (k-means algorithm) و سپس الگوریتم گرادینت مزدوج (scaled conjugate gradient algorithm) برای تعیین مقدار بهینه پارامترها مورد استفاده قرار گرفت. در مرحله آخر، از قوانین فازی تولیدشده برای مدل کردن و ارزیابی سیستم استفاده کردند (۱۸). از جمله پژوهش‌های دیگری که در این زمینه صورت گرفته است، می‌توان به ترکیب شبکه عصبی فازی و روش تحلیل ممیزی خطی (Linear Discriminant Analysis) (۱۹) و ترکیب الگوریتم ژنتیک با شبکه عصبی احتمالی و ماشین بردار پشتیبان اشاره کرد (۲۰). به طور کلی توجه به متدهای مختلف داده کاوی در تشخیص بیماری تیروئید مورد توجه محققان بوده است (۲۱) و مطالعات متعدد نشان می‌دهد که ابزارهای یادگیری ماشین می‌توانند کمک قابل توجهی به کشف و طبقه‌بندی بیماری‌های مختلف از جمله تیروئید نمایند (۲۲).

آنچه پژوهش حاضر نسبت به پژوهش‌های مشابه قابل توجه است، استفاده از چند مدل یادگیر نظارت شده و غیرنظارت شده است. اگرچه در برخی از مطالعات همچون Amrollahi Biyouki و همکاران (۲۳) استفاده از مدل‌های مختلف مانند شبکه عصبی و بیزین و نزدیک‌ترین همسایه بررسی شده است؛ لیکن در آن مطالعه و اکثر مطالعات مشابه توجه به مدل‌های با ناظر و بدون ناظر صورت گرفته است که در آن‌ها قطعیت وجود دارد. جدول ۶ عملکرد استفاده از روش نروفازی در مقایسه با الگوریتم‌های برخی دیگر از پژوهش‌ها را نشان می‌دهد.

جدول ۶: مقایسه‌ی عملکرد روش استفاده شده در این پژوهش با سایر مطالعات

الگوریتم مورد استفاده	مرجع	نوع داده	درصد دقت آموزش
CSFNN	(۲۴)	UCI دسته اول	۹۱/۴
RNF	(۲۴)	UCI دسته اول	۷۹/۰۸

MLP	(۲۴)	UCI دسته اول	۸۹/۸۰
Bayesian	(۲۵)	UCI دسته دوم	۹۷/۷۰
شبکه عصبی احتمالی	(۲۶)	داده‌های بیماری تیروئید جمع آوری شده از بیمارستان امام خمینی	۹۶/۶۷
MLP	(۲۷)	UCI دسته دوم	۹۴/۴
عصبی فازی	پژوهش حاضر	UCI دسته اول	۹۸/۲

نتیجه گیری

به‌رغم تمام روش‌های استانداردسازی در تشخیص پزشکی، یکی تشخیص درست هنوز یک هنر به‌شمار می‌آید. قسمت اعظم این وضعیت به این دلیل است که تشخیص پزشکی نیازمند تخصص و تجربه در مواجهه با عدم قطعیت است. به این معناکه پزشک نمی‌تواند با یقین کامل نسبت به ابتلا و یا عدم‌ابتلا به برخی از بیماری‌ها اظهارنظر کند. در جهان ماشینی امروز که مرزهای دانش پزشکی بسیار گسترده شده، به‌آسانی نمی‌توان بر این عدم قطعیت چیره شد. ارایه یک چارچوب قدرتمند برای مدل‌سازی سیستم‌های فعلی باعث می‌شود نظریه فازی به عاملی ارزشمند برای پیشرفت تشخیص پزشکی تبدیل گردد. استفاده از این روش‌ها می‌تواند، خطاهای احتمالی ناشی از خستگی یا بی‌تجربگی را در امر تشخیص و درمان بیماری‌ها کاهش دهد. هم‌چنین با استفاده از این سیستم‌ها، می‌توان پایگاه داده‌های پزشکی را در زمان بسیار کمتر و با جزئیات بیش‌تر تحلیل نمود.

مدل ارایه شده در این مطالعه می‌تواند در برنامه‌های غربالگری جهت شناسایی افراد در معرض خطر استفاده شود. از آنجایی‌که در تحقیقات پزشکی مبتنی بر کاوش داده اهمیت ثبت داده‌های صحیح بسیار مشهود است، پیشنهاد می‌شود این مدل در یکی از مراکز درمانی به‌صورت آزمایشی و به‌عنوان دستیار پزشک مورد استفاده قرارگیرد تا ضمن ثبت داده‌های بومی بیماران مبتلا به تیروئید، بتواند در کنار پزشک متخصص به‌عنوان یک دستیار هوشمند استفاده گردد.

به‌طور کلی در مرجع UCI دو نوع داده وجود دارد. دسته اول که شامل ۲۱۵ نمونه با ۵ ویژگی است که در این مطالعه استفاده شده است و دسته دوم شامل ۷۲۰۰ نمونه با ۲۱ ویژگی است که مورد استفاده در برخی دیگر از مطالعات بوده است. ناگفته نماند که کاوش و مدل‌سازی داده‌های دسته دوم UCI به‌دلیل اینکه تعداد نمونه و تعداد ویژگی بیشتری دارد امکان‌پذیرتر بوده و همان‌طورکه جدول ۶ نشان می‌دهد، مطالعات مختلف نشان از دقت بالای یادگیرها در این دسته از داده‌ها، نسبت به داده‌های دسته اول دارد. در مطالعه‌ی حاضر علت بهبود نتیجه را می‌توان این‌گونه توصیف کرد که: اگرچه هر یک از شبکه‌های عصبی به‌عنوان سیستم‌های نظارت‌شده می‌توانند به‌تنهایی به پیش‌بینی بیماری تیروئید بپردازند، اما درجه یقین آن‌ها قطعی است. به این ترتیب برای هر طبقه‌بندی پایه، توزیع نمونه‌های ورودی برای آموزش در جهتی پیش می‌رود که نهایتاً به یکی از کلاس‌های قطعی خاتمه یابد. لیکن در طبقه‌بندی فازی نظارت‌شده یک نمونه با درجه‌ای از عدم قطعیت به یک کلاس تعلق می‌یابد و به پزشک اجازه می‌دهد که تجویز دارو و فرایند درمان را با درجه‌ای از عدم قطعیت به‌پیش برد.

یکی از محدودیت‌هایی که در این مطالعه با آن مواجه هستیم، مربوط به ماهیت داده تیروئید است. وجود داده‌های بیماری تیروئید، که بر اساس سبک زندگی ایرانی یعنی متکی بر داده‌های بومی جمع‌آوری شده باشد، می‌تواند به‌طور عملیاتی‌تری استفاده شده و در پژوهش‌های بعدی بدان توجه شود.

منابع

1. Geetha K & Baboo SS. An empirical model for Thyroid disease classification using evolutionary multivariate bayesian prediction method. Global Journal of Computer Science and Technology 2016; 16(1): 1-10.
2. Abdollahpour N, Sharifi A & Aghini O. Using multiple batches in diagnosing Thyroid disorders, Tehran: First International Conference on New Research Achievements in Electrical and Computer Engineering, 2014.



3. Wlodzislaw D, Adamczak R & Grabczewski K. A new methodology of extraction, optimization and application of crisp and fuzzy logical rules. (IEEE) Transactions on Neural Networks 2001; 12(2): 277-306.
4. Baghbani-Oskouei A, Akbarpour S, Tohidi M, Amouzegar A, Mehran L, Hadaegh F, et al . Subclinical Thyroid dysfunction and incident Cardiovascular disease: Tehran Thyroid study. Iranian Journal of Endocrinology and Metabolism 2018; 20(1) :1-9[Article in Persian].
5. Chandel K, Kunwar V, Sabitha S, Choudhury T & Mukherjee S. A comparative study on Thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques. In CSI Transactions on ICT 2016; 4(2): 313-9.
6. Zabbah I, Eskandari A, Sardari Z & Noghandi A. Diagnosis of diabetes using artificial neural network and neuro-fuzzy approach. Journal of Torbat Heydariyeh University of Medical Sciences 2018; 6(2): 10-20[Article in Persian].
7. Mitra S & Hayashi Y. Neuro-fuzzy rule generation: Survey in soft computing framework. (IEEE) Transactions on Neural Networks 2000; 11(3): 748-68.
8. Hassanzadeh M, Zabbah I & Layeghi K. Diagnosis of Coronary Heart disease using mixture of experts method. Journal of Health and Biomedical Informatics 2018; 5(2): 274-85[Article in Persian].
9. Khan E & Venkatapuram P. Neufuz: Neural network based fuzzy logic design algorithms. (IEEE) Institute of Electrical and Electronics Engineers 1993; 3(1): 647-54.
10. Rastogi A & Bhalla M. Study of neural network in diagnosis of Thyroid. International Journal of Computer Technology and Electronics Engineering (IJCTEE) 2014; 4(3): 13-6.
11. Dayhoff JE & Deleo JM. Artificial neural networks: Opening the Black box. Cancer 2001; 91(S8): 1615-35.
12. Haykin S. Neural Networks. A comprehensive foundation. 2nd ed. India: Prentice Hall; 2004: 41.
13. Mirahmadizadeh A, Kavooosi E, Vakili M, Shenavar R & Moghadami M. Iodine deficiency disorders in the south of Iran during 1989-2012: A surveillance system report. Journal of Health Sciences & Surveillance System 2013; 1(2): 57-66.
14. Asgari N & Khosravi Danesh F. Thyroid disease detection using fuzzy-neural classification algorithm. Tehran: Second National Conference Computer Engineer Information Technology, 2014.
15. Forghani N, Teshne Lab M & Abrishami Moghadam H & Nakhjavani M. Diagnosis of Thyroid gland diseases using a hybrid-fuzzy neural network based on linear separation analysis, Tehran: 14th Iranian Conference on Medical Engineering, 2007.
16. Adibi A, Haghighi M, Hosseini SR, Hashemipour M, Amini M & Hovsepian S. Thyroid abnormalities among first-degree relatives of children with congenital hypothyroidism: An ultrasound survey. Hormone Research in Paediatrics 2008; 70(2): 100-4.
17. Senol C & Yildirim T. Thyroid and Breast cancer disease diagnosis using fuzzy-neural networks. Turkey: (IEEE) International Conference on Electrical and Electronics Engineering-ELECO, 2009.
18. Jorian N & Ashoori M. Predicting the effectiveness of Preeclampsia medications based on dose and method of drug consumption using data mining. The Iranian Journal of Obstetrics, Gynecology and Infertility 2014; 17(123): 13-22[Article in Persian].
19. Al-Daoud E. Cancer diagnosis using modified fuzzy network. Universal Journal of Computer Science and Engineering Technology 2010; 1(2): 73-8.
20. Geetha K & Baboo SS. Efficient Thyroid disease classification using differential evolution with SVM. Journal of Theoretical & Applied Information Technology 2016; 88(3): 410-20.
21. Umar Sidiq D, Aaqib SM & Khan RA. Diagnosis of various Thyroid ailments using data mining classification techniques. International Journal of Scientific Research in Computer Science, Engineering and Information Technology 2019; 5(1): 131-6.
22. Razia S, Kumar PS & Rao AS. Machine learning techniques for Thyroid disease diagnosis: A systematic review. USA: InModern Approaches in Machine Learning and Cognitive Science: A Walkthrough; 2020: 203-12.
23. Amrollahi Biyouki S, Turksen IB & Fazel Zarandi MH. Fuzzy rule-based expert system for diagnosis of Thyroid disease, Canada: IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2015.

24. Raghavendra U, Acharya UR, Gudigar A, Tan JH, Fujita H, Hagiwara Y, et al. Fusion of spatial gray level dependency and fractal texture features for the characterization of Thyroid lesions. *Ultrasonics* 2017; 77(1): 110-20.
25. Raymer ML, Kuhn L & Punch WF. Knowledge discovery in biological datasets using a hybrid bayes classifier/evolutionary algorithm, USA: (IEEE) International Symposium on Bioinformatics and Bioengineering, 2001.
26. Zargari F, Rafie A, Jamali JS & Sedaghat Pische S. Thyroid ultrasound images using neural network lassification, Mashhad: National Conference on Electrical Engineering and Sustainable Development Focusing on New Achievements in Electrical Engineering, 2014.
27. Mojahedi ME & Rouhani M. Optimization of the MLP neural network to detect Thyroid disease by clinical data, Gonabad: 5th Iranian Conference on Electrical and Electronics Engineering, 2013.

Comparison of Data Mining Algorithms' Efficiency in Thyroid Disease Diagnosis

Mohsen Rezaei¹ (M.S.) - Nazanin Zahra Jafari¹ (M.S.) - Hossein Ghaffarian²
(Ph.D.) - Masoud Khosravi Farmad³ (Ph.D.) - Iman Zabbah⁴ (M.S.) - Parvaneh
Dehghan⁵ (M.D.)

1 Master of Sciences Student in Information Technology, Department of Computer, Higher Education Technical Institute Atrak Ghoochan, Ghoochan, Iran

2 Assistant Professor, Department of Computer Engineering, Faculty of Engineering, Arak University, Arak, Iran

3 Assistant Professor, Department of Computer, Higher Education Technical Institute Atrak Ghoochan, Ghoochan, Iran

4 Lecturer, Department of Computer, Torbat-e Heydariyeh Branch, Islamic Azad University, Torbat-e Heydariyeh, Iran

5 Assistant Professor, Department of Oncology Radiotherapy, Torbat-e Heydarieh University of Medical Sciences, Torbat-e Heydarieh, Iran

Abstract

Received: Apr 2019

Accepted: Aug 2019

Background and Aim: Timely diagnosis and treatment of abnormal thyroid function can reduce the mortality associated with this disease. However, lack of timely diagnosis will have irreversible complications for the patient. Using data mining techniques, the aim of this study is to determine the status of the thyroid gland in terms of normality, hyperthyroidism or hypothyroidism.

Materials and Methods: Using supervised and unsupervised methods after data preprocessing, predictive modeling was performed to classify thyroid disease. This is an analytical study and its dataset contains 215 independent records based on 5 continuous features retrieved from the UCI machine learning data reference.

Results: In supervised method, multilayer perception(MLP), learning vector quantization(LVQ), and fuzzy neural network(FNN) were used; and in unsupervised method, fuzzy clustering was employed. Besides, these precision figures(0.055, 0.274, 0.012 and 1.031) were obtained by root mean square error(RMSE) method, respectively.

Conclusion: Reducing the diagnosis error of thyroid disease was one of the goals of researchers. Using data mining techniques can help reduce this error. In this study, thyroid disease was diagnosed by different pattern recognition methods. The results show that the fuzzy neural network(FNN) has the least error rate and the highest accuracy.

Keywords: Thyroid Disease, Fuzzy Clustering, Artificial Neural Networks, Neural Fuzzy, Supervised Learning Networks

* Corresponding Authors:
Zabbah I
Email :
Iman.zabbah@iautorbat.ac.ir