DOI: 10.22100/jkh.v15i2.2411

مجله دانش و تندرستی در علوم پایه پزشکی، دوره ۱۵، شماره ۲، تابستان ۱۳۹۹

دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شاهرود

مقاله پژوهشی

ارایه قانون تشخیصی جدید برای بیماری پارکینسون مبتنی‌بر روش استخراج ترکیبی

فاطمه آهوز[۱]، امین گلاب‌پور[۲*]

۱- عضو هیئت علمی- دانشگاه صنعتی خاتم‌الانبیاء بهبهان- بهبهان- ایران.

۲- استادیار- گروه فناوری اطلاعات سلامت- دانشکده پیراپزشکی- دانشگاه علوم پزشکی شاهرود- شاهرود- ایران.

تاریخ دریافت: ۱۳۹۹/۰۴/۰۶، **تاریخ پذیرش:** ۱۳۹۹/۰۵/۱۱

چکیده

مقدمه: *بیماری پارکینسون یک معضل بهداشت جهانی مهم است که با گذشت زمان علایم آن شدیدتر می‌شود. تشخیص و درمان سریع این بیماری منجر به بهبود مهارت‌ها، توانمندی‌ها و عملکرد مبتلایان در کارهای روزمره زندگی می‌شود. به‌منظور تشخیص زودهنگام این بیماری، تولید سیستم‌های کمک تصمیم‌یار بالینی که قادر به کشف قوانین تشخیصی این بیماری باشند، ضروری است.*

مواد و روش‌ها: *در این مقاله یک روش خودکار استخراج قوانین تشخیصی جدید برای بیماری پارکینسون ارایه می‌شود. روش پیشنهادی مبتنی بر رگرسیون منطقی و الگوریتم شبیه‌سازی تبرید است. جهت ارزیابی روش از مجموعه داده پارکینسون آکسفورد استفاده شده که اطلاعات ۲۳ فرد مبتلا به پارکینسون و ۸ فرد سالم را در قالب ۱۹۵ رکورد ذخیره کرده است. برای هر رکورد ۲۳ ویژگی ذخیره شده که شامل ۲۲ اندازه‌گیری صوتی است.*

نتایج: *نتایج به‌دست آمده شامل دو قانون تشخیصی است؛ در سیستم‌های خودکاری که صحت بالا اولویت دارد، یک قانون جدید شامل ۲۱ عبارت منطقی ارایه شده است که دارای صحت ۹۲/۳۱، حساسیت ۸۵/۴۲ و ویژگی ۹۴/۵۶ درصد است. همچنین برای سیستم‌های بلادرنگ و کمک تصمیم‌یار بالینی با قابلیت تفسیرپذیری بالا، قانونی متشکل از سه عبارت منطقی پیشنهاد شده است که دارای صحت ۷۸/۹۷، حساسیت ۷۷/۰۸ و ویژگی ۷۹/۵۹ درصد است.*

نتیجه‌گیری: *نتایج به‌دست آمده نشان‌دهنده قدرت بالای تفسیرپذیری و قابلیت اعتماد قانون به‌دست آمده در تشخیص بیماری پارکینسون است که می‌تواند در پیاده‌سازی سیستم‌های از راه‌دور تشخیصی استفاده شود.*

واژه‌های کلیدی: پارکینسون، رگرسیون منطقی، شبیه‌سازی تبرید، استخراج قانون.

***نویسنده مسئول:** شاهرود- دانشگاه علوم پزشکی شاهرود، **تلفن:** ۳۲۳۹۵۰۵۴، **نمابر:** ۳۲۳۹۴۸۰۰، **Email:** a.golabpour@shmu.ac.ir

ارجاع: آهوز فاطمه، گلاب‌پور امین. ارایه قانون تشخیصی جدید برای بیماری پارکینسون مبتنی‌بر روش استخراج ترکیبی. مجله دانش و تندرستی در علوم پایه پزشکی ۱۳۹۹؛۱۵(۲):۴۲-۵۰.

مقدمه

پارکینسون دومین بیماری شایع نورودژنراتیو و پیشرونده در دنیا و یکی از معضلات سلامت عمومی است (۱ و ۲). این بیماری به‌دلیل تولید ناکافی دوپامین در مغز ایجاد می‌شود و به توانایی‌های حرکتی و گفتاری آسیب می‌زند (۵-۳). از دست دادن دوپامین عامل ایجاد اختلال حرکتی هیپوکینتیک سلول‌های عصبی است (۵). در حال حاضر بیش از ۱۰ میلیون نفر در سراسر دنیا از این بیماری رنج می‌برند (۶). پدیده سالخوردگی در جهان یکی از عوامل افزایش تعداد مبتلایان به این بیماری است (۱ و ۷).

برخی از علایم این بیماری شامل لرزش، کندی حرکات، جمود ماهیچه‌ها، اختلال در وضعیت بدنی و تعادل و تغییر در نوشتار و گفتار است (۸). اختلالات صوتی و گفتاری از جمله از نفس افتادن، کاهش بلندی و لرزش غیرمعمول صدا در بیش از ۹۰ درصد افراد مبتلا به بیماری پارکینسون مشاهده می‌شود (۱، ۳، ۸ و ۹). از آنجا که بروز اختلالات گفتاری در بیماران مبتلا به پارکینسون پنج سال قبل از تشخیص بالینی اتفاق می‌افتد، طراحی سیستم‌های خودکار تشخیص بیماری پارکینسون برای تشخیص زودهنگام این بیماری به کمک بررسی داده‌های صوتی سودمند به نظر می‌رسد (۱). تشخیص زودهنگام این بیماری در حفظ کیفیت زندگی از طریق درمان مناسب ضروری بوده و می‌تواند علایم مرتبط با این بیماری را به میزان قابل توجهی کاهش دهد (۱، ۵ و ۱۰).

از آنجا که اختلالات صدا می‌تواند اولین نشانه‌ی ابتلا به پارکینسون باشد (۲)، مجموعه داده پارکینسون آکسفورد که شامل تعدادی اندازه‌گیری صوتی برای تمایز بیماران مبتلا به پارکینسون از افراد سالم به کمک ویژگی‌های صوتی مؤثر است، جهت تشخیص و پیگیری پیشرفت علایم پارکینسون مورد توجه پژوهشگران قرار گرفته است. اغلب کارهای صورت گرفته شامل بررسی روند پیشرفت بیماری و پیشگویی مراحل میانی (۲)، تشخیص بیماری بر اساس روش‌های متفاوت داده‌کاوی (۱، ۵، ۸ و ۱۱)، استفاده از الگوریتم‌های تکاملی در پیش‌بینی بیماری (۱، ۹، ۱۲ و ۱۳)، تجزیه و تحلیل الگوهای رفتاری و تعیین ویژگی‌های مؤثر بر آن (۹ و ۱۵-۱۳) و ارایه روش‌های ترکیبی به‌منظور افزایش صحت سیستم پیش‌بینی (۷ و ۱۶) هستند. از میان کارهای صورت گرفته تنها تعداد محدودی به استخراج قوانین از این مجموعه داده پرداخته‌اند.

وانگ و همکاران در سال ۲۰۱۴ (۱۷) یک شبکه عصبی ترکیبی هایپر مستطیل فازی (PFHRCNN) مبتنی بر الگوریتم بهینه‌سازی ذرات (PSO) جهت استخراج قوانین از مجموعه داده پارکینسون آکسفورد پیشنهاد دادند. با تقسیم تصادفی مجموعه داده‌ها به‌صورت ۵۰-۵۰ درصدی به مجموعه‌های آموزش و آزمون و ۱۰ بار اجرای روش، صحت روش پیشنهادی با ۸ قانون، ۸۲/۴ درصد روی مجموعه‌ی آزمون و ۹۲/۶ درصد روی آموزش گزارش شده است. در این مقاله تنها دو قانون از ۸ قانون ذکر شده که هر قانون شامل بررسی محدوده برای تمام ۲۲ ویژگی موجود در مجموعه داده هم برای افراد سالم و هم افراد مبتلا به پارکینسون است. با توجه به اینکه طول قوانین یکی از شاخص‌های اصلی تفسیرپذیری در سیستم‌های کمک تصمیم‌یار بالینی است، تعداد زیاد فرض‌ها در هر قانون تفسیرپذیری آن را دشوار می‌کند (۲۰-۱۸).

آبی‌زاده و آبیف (۲۱) یک مدل ترکیبی از سیستم فازی و شبکه عصبی جهت شناسایی افراد مبتلا به پارکینسون از افراد سالم، پیشنهاد دادند. روش ارایه شده از قوانین اگر-آنگاه فازی استفاده می‌کند؛ جهت تعیین فرض‌های قوانین از روش c-means فازی و برای تعیین تالی قوانین از الگوریتم گرادیان استفاده شده است. صحت مدل با ۱۶ قانون، پس از اجرا به صورت ارزیابی متقاطع ۱۰ فولد ۱۰۰ درصد گزارش شده است. با این حال در این مقاله هیچ اشاره‌ای به قوانین تولید شده، نشده است.

رودریگز و کارونانیدی (۲۲) از سیستم استنتاج فازی ممدانی استفاده کردند. بدون ذکر روش دستیابی، چهار ویژگی Spread1، DFA، FoH و Spread2 برای تحلیل تشخیص و شدت بیماری پارکینسون انتخاب شده است. این چهار ویژگی در سه مجموعه‌ی افراد مبتلا به پارکینسون، مشترک بین افراد مبتلا و سالم و مجموعه افراد سالم طبقه‌بندی شدند. به گفته نویسندگان در تحلیل مقادیر این چهار ویژگی، تفاوت واضحی بین افراد سالم و مبتلا به پارکینسون مشاهده شده اما اشاره‌ای به صحت مدل ایجاد شده از این قوانین نشده است. سیستم نهایی آن‌ها شامل ۸۱ قانون است که تنها ۹ قانون آن در مقاله ذکر شده است.

آوچی و دوگانتیکن (۹) یک سیستم خبره تشخیص بیماری پارکینسون برمبنای الگوریتم ژنتیک (GA)- کرنل ویولت (WK)، یادگیری ماشینی Extreme (به اختصار ELM) ارایه دادند. دسته‌بند شبکه عصبی تک‌لایه (SLNN) است که برای آموزش آن از روش یادگیری ELM استفاده می‌شود. تنظیم سه پارامتر کرنل ویویلت به همراه تعداد نرونهای شبکه عصبی با استفاده از الگوریتم ژنتیک محاسبه می‌شود. با آموزش روش GA-WK-ELM روی ۱۰۰ نمونه آموزشی، ارزیابی مدل روی ۹۵ نمونه آزمون به‌صورت حساسیت ۹۵/۴۵درصد، ویژگی ۹۸/۱۷ و صحت ۹۶/۸۷ درصد گزارش دادند.

یکی از جنبه‌های مهم بررسی این بیماری که تا کنون به‌طور جامع مورد توجه قرار نگرفته است، کشف و استخراج قوانین نهفته در داده‌های صوتی با قابلیت صحت و تفسیرپذیری بالاست. با توجه به

اینکه دانش متخصصان در قالب قوانین تشخیصی به‌طور تجربی و طی سالیان در ذهن پزشک شکل می‌گیرد، تولید سیستم‌های کمک تصمیم‌یار بالینی که قادر به کشف قوانین حاکم بر داده‌ها باشند در کشف زودهنگام بیماری حتی در مناطقی که به فرد متخصص دسترسی وجود ندارد، نقشی حیاتی ایفا می‌کند. از این‌رو، در این مقاله با استفاده از ترکیب روش‌های رگرسیون منطقی و الگوریتم شبیه‌سازی تبرید، با تعیین مهمترین ویژگی‌های موثر، مجموعه قوانین تشخیص بیماری پارکینسون با قابلیت صحت و تفسیرپذیری بالا پیشنهاد می‌شود. هدف به‌دست آوردن قوانین بهینه، جامع و ساده است تا ضمن دستیابی به صحت بالا، توسط انسان قابل درک باشد.

رگرسیون منطقی ابزاری برای کشف اثرات متقابل میان متغیرهای پیشگو است که مرتبط با متغیر پاسخ هستند. قدرت رگرسیون منطقی این است که می‌تواند حتی اثرات متقابل پیچیده میان پیشگوها را پیدا کند که نقش مهمی در کشف عوامل مؤثر در داده‌های پزشکی دارد (۲۳). همچنین این روش با تولید درخت‌های منطقی و بهره‌برداری از قدرت روش‌های رگرسیون خطی و لوجستیک، جهت استخراج قوانین از داده‌ها مؤثر است. بر اساس دانش ما تاکنون از قابلیت استخراج قانون این الگوریتم در تشخیص پارکینسون استفاده نشده و این ظرفیت بدون بهره مانده است. از این‌رو، در این مقاله روشی برحسب رگرسیون منطقی به منظور استخراج قوانین ارایه می‌شود.

مواد و روش‌ها

در این مقاله از مجموعه داده پارکینسون آکسفود برای آموزش و ارزیابی مدل پیشنهادی استفاده می‌شود. این مجموعه‌داده از مخزن یادگیری ماشین UCI که توسط مکس لیتل از دانشگاه آکسفورد در ارتباط با مرکز ملی صدا و گفتار دنور، کالیفرنیا آمریکا شکل گرفته به‌دست آمده است (۲۴). این مجموعه‌داده شامل اطلاعات ۳۱ فرد است که ۲۳ نفر از آن‌ها مبتلا به بیماری پارکینسون هستند و مجموعاً دارای ۲۴ ویژگی شامل نام بیماران، وضعیت افراد (سالم یا مبتلا به پارکینسون) و ۲۲ اندازه‌گیری صوتی است که از برنامه صوتی چندبعدی (MDVP) گرفته شده است (۳). این مجموعه داده مجموعاً دارای ۱۹۵ صدای ضبط شده است که از ضبط ۵ الی ۶ صدا از هر شخص برای ادا کردن حروف صدادار به مدت ۳۶ ثانیه تشکیل شده است که این جزئیات در ادامه بیان می‌گردد. تشخیص افراد طی یک بازه‌ی ۲۸ ساله (از صفر تا ۲۸) صورت گرفته و سن افراد بین ۴۶ تا ۸۵ سال با میانگین ۶۵/۸ است (۱۳). هدف اصلی این مجموعه‌داده تمایز افراد دارای پارکینسون از افراد سالم از طریق ویژگی‌های صوتی تاثیرگذار است. جدول ۱ اطلاعات آماری میانگین، واریانس و محدوده‌ی مقادیر هر ویژگی این مجموعه داده را به ازای اطلاعات افراد سالم و مبتلا به پارکینسون نشان می‌دهد که در نرم‌افزار SPSS محاسبه شده است.

این مجموعه‌داده دارای مقادیر گمشده نیست و چون داده استاندارد است در این مقاله تجزیه و تحلیل حجم نمونه برای آن انجام نشده است. پس از حذف ویژگی نام بیماران، مجموعه‌داده در بازه [۱ و ۰] طبیعی شده و سپس تمام مجموعه‌داده وارد مرحله تحلیل و ساخت مدل شد.

رگرسیون منطقی یک متدولوژی رگرسیون تطبیقی است که توسط روزینسکی و همکاران در (۲۳) ارایه شده است. این روش تلاش می‌کند متغیرهای پیشگو را به‌صورت ترکیب بولین از همبستگی‌های دودویی ایجاد کند. در این روش با ارایه‌ی مجموعه‌ای از متغیرهای پیشگوی دودویی، پیشگوهای جدیدی به‌صورت قوانین تصمیم‌گیری ایجاد می‌شود. این قوانین به‌صورت عبارت‌های منطقی بوده و تنها از عملگرهای and و or منطقی استفاده می‌کنند. مدل رگرسیونی ایجاد شده به فرم زیر است:

(فرمول ۱):

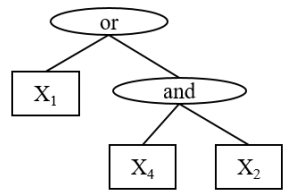
$$g(E[Y]) = \beta_0 + \sum_{j=1}^{t} \beta_j L_j$$

که در آن Ljها عبارت‌های منطقی از متغیرهای پیشگو، t حداکثر تعداد متغیرهای پیشگو جدید و βjها ضرایب هستند. در این مدل هدف یافتن ترکیبی از متغیرهای دودویی است که قدرت پیشگویی بالایی برای متغیر پاسخ را داشته باشند. از آنجا که این متغیرهای پیشگو، دودویی هستند، هر ترکیبی از آن‌ها نیز دودویی است. با توجه به تفسیر عبارت‌های منطقی می‌توان آن‌ها را به فرم درخت‌های دودویی ارایه داد. ارزیابی چنین درختی، به‌عنوان دستوری منطقی برای حالتی خاص، به‌صورت بازگشتی و جایگزینی آن به‌صورت پایین به بالا انجام می‌شود. از آنجا که بررسی تمام ترکیب‌های ممکن جهت انتخاب بهترین مجموعه از پیشگوهای جدید امکان‌پذیر نیست، از روش‌های بهینه‌سازی جهت تخمین پارامترهای مدل استفاده می‌شود. در این مقاله الگوریتم شبیه‌سازی تبرید برای تخمین همزمان Ljها و βjها استفاده می‌شود. شکل ۱ مثالی از چنین عبارت‌های منطقی و درخت‌های متناظر را نشان می‌دهد. در این شکل Xiها متغیرهای پیشگو مجموعه داده هستند.

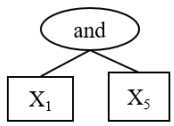
در این مقاله، سه سطح تصادفی‌سازی صورت گرفته است. در سطح اول، جهت تبدیل متغیرهای پیشگوی پیوسته به متغیرهای پیشگوی دودویی، مقادیری تصادفی در بازه صفر تا یک با طول گام ۰/۰۱ برای هر متغیر پیشگو انتخاب می‌شود؛ سپس از بین عملگرهای رابطه‌ای کوچکتری (>)، بزرگتری (<) یا بازه‌ای (a<x<b) برای هر متغیر به‌طور تصادفی یک رابطه انتخاب می‌شود. به این ترتیب اگر رابطه برقرار باشد مقدار این متغیر یک و در غیر این صورت مقدار آن صفر خواهد بود.

جدول ۱- اطلاعات آماری مجموعه داده پارکینسون آکسفورد

نام ویژگی	توضیحات	افراد سالم (status=0)		افراد مبتلابه پارکینسون (status=1)	
		میانگین±واریانس	[کمترین-بیشترین]	میانگین±واریانس	[کمترین-بیشترین]
MDVPFo(Hz)	میانگین فرکانس اساسی صدا	۱۸۱/۹۴ ± ۵۲/۷۳	[۱۱۰/۷۴ – ۲۶۰/۱۱]	۱۴۵/۱۸ ± ۳۲/۳۵	[۸۸/۳۳ – ۲۲۳/۳۶]
MDVPFhi(Hz)	ماکزیمم فرکانس اساسی صدا	۲۲۳/۶۴ ± ۹۶/۷۳	[۱۱۳/۶۰ – ۵۹۲/۰۳]	۱۸۸/۴۴ ± ۸۸/۳۴	[۱۰۲/۱۵ – ۵۸۸/۵۲]
MDVPFlo(Hz)	مینیمم فرکانس اساسی صدا	۱۴۵/۲۱ ± ۵۸/۷۶	[۷۴/۲۹ – ۲۳۹/۱۷]	۱۰۶/۸۹ ± ۳۲/۲۷	[۶۵/۴۸ – ۱۹۹/۰۲]
MDVPJitter	برخی از اندازه‌گیری‌های تغییرات در فرکانس اساسی	۰/۰۰ ± ۰/۰۰	[۰/۰۰ – ۰/۰۱]	۰/۰۱ ± ۰/۰۱	[۰/۰۰ – ۰/۰۳]
MDVPJitterAbs		۰/۰۰ ± ۰/۰۰	[۰/۰۰ – ۰/۰۰]	۰/۰۰ ± ۰/۰۰	[۰/۰۰ – ۰/۰۰]
MDVPRAP		۰/۰۰ ± ۰/۰۰	[۰/۰۰ – ۰/۰۱]	۰/۰۰ ± ۰/۰۰	[۰/۰۰ – ۰/۰۲]
MDVPPPQ		۰/۰۰ ± ۰/۰۰	[۰/۰۰ – ۰/۰۱]	۰/۰۰ ± ۰/۰۰	[۰/۰۰ – ۰/۰۲]
JitterDDP		۰/۰۱ ± ۰/۰۰	[۰/۰۰ – ۰/۰۲]	۰/۰۱ ± ۰/۰۱	[۰/۰۰ – ۰/۰۶]
MDVPShimmer		۰/۰۲ ± ۰/۰۱	[۰/۰۱ – ۰/۰۴]	۰/۰۳ ± ۰/۰۲	[۰/۰۱ – ۰/۱۲]
MDVPShimmerdB		۰/۱۶ ± ۰/۰۶	[۰/۰۹ – ۰/۴۱]	۰/۳۲ ± ۰/۲۱	[۰/۰۹ – ۱/۳۰]
ShimmerAPQ3	برخی از اندازه‌گیری‌های تغییرات در دامنه	۰/۰۱ ± ۰/۰۰	[۰/۰۰ – ۰/۰۲]	۰/۰۲ ± ۰/۰۱	[۰/۰۰ – ۰/۰۶]
ShimmerAPQ5		۰/۰۱ ± ۰/۰۰	[۰/۰۱ – ۰/۰۲]	۰/۰۲ ± ۰/۰۱	[۰/۰۱ – ۰/۰۸]
MDVPAPQ		۰/۰۱ ± ۰/۰۰	[۰/۰۱ – ۰/۰۳]	۰/۰۳ ± ۰/۰۲	[۰/۰۱ – ۰/۱۴]
ShimmerDDA		۰/۰۳ ± ۰/۰۱	[۰/۰۱ – ۰/۰۷]	۰/۰۵ ± ۰/۰۳	[۰/۰۱ – ۰/۱۷]
NHR	دو اندازه‌گیری نسبت نویز	۰/۰۱ ± ۰/۰۲	[۰/۰۰ – ۰/۱۱]	۰/۰۳ ± ۰/۰۴	[۰/۰۰ – ۰/۳۱]
HNR	به اجزای تونی در وضعیت صدا	۲۴/۶۸ ± ۳/۴۳	[۱۷/۸۸ – ۳۳/۰۵]	۲۰/۹۷ ± ۴/۳۴	[۸/۴۴ – ۲۹/۹۳]
RPDE	دو اندازه‌گیری پیچیدگی	۰/۴۴ ± ۰/۰۹	[۰/۲۶ – ۰/۶۶]	۰/۵۲ ± ۰/۱۰	[۰/۲۶ – ۰/۶۹]
D2	پویایی غیرخطی	۲/۱۵ ± ۰/۳۱	[۱/۴۲ – ۲/۸۸]	۲/۴۶ ± ۰/۳۸	[۱/۷۷ – ۳/۶۷]
DFA	نماینده مقیاس‌پذیری فراکتال سیگنال	۰/۷۰ ± ۰/۰۵	[۰/۶۳ – ۰/۷۹]	۰/۷۳ ± ۰/۰۵	[۰/۵۷ – ۰/۸۳]
spread1	سه اندازه‌گیری غیرخطی از تغییرات فرکانسی اساسی	−۶/۷۶ ± ۰/۶۴	[(−۷/۹۶) – (−۵/۲۰)]	−۵/۳۳ ± ۰/۹۷	[(−۷/۱۲) – (−۲/۴۳)]
spread2		۰/۱۶ ± ۰/۰۶	[۰/۰۱ – ۰/۲۹]	۰/۲۵ ± ۰/۰۸	[۰/۰۶ – ۰/۴۵]
PPE		۰/۱۲ ± ۰/۰۴	[۰/۰۴ – ۰/۲۵]	۰/۲۳ ± ۰/۰۸	[۰/۰۹ – ۰/۵۳]



الف) عبارت منطقی $L_1 = X_1 or\ (X_4 and\ X_2)$



ب) عبارت منطقی $L_2 = X_1 and\ X_5$

شکل ۱- دو عبارت پیشگو جدید که از متغیرهای پیشگو اصلی ایجاد شده است.

سطح دوم تصادفی‌سازی مربوطبه الگوریتم شبیه‌سازی تبرید است. در این مقاله، الگوریتم شبیه‌سازی تبرید مسئول انتخاب تصادفی زیرمجموعه‌ای از متغیرهای پیشگو برای ساخت متغیرهای پیشگو جدید، یعنی Lj است که توسط الگوریتم رگرسیون منطقی ایجاد می‌شود. صحت درخت ایجاد شده توسط رگرسیون منطقی به‌عنوان درجه‌ی بهینگی راه‌حل‌ها در شبیه‌سازی تبرید در نظر گرفته می‌شود.

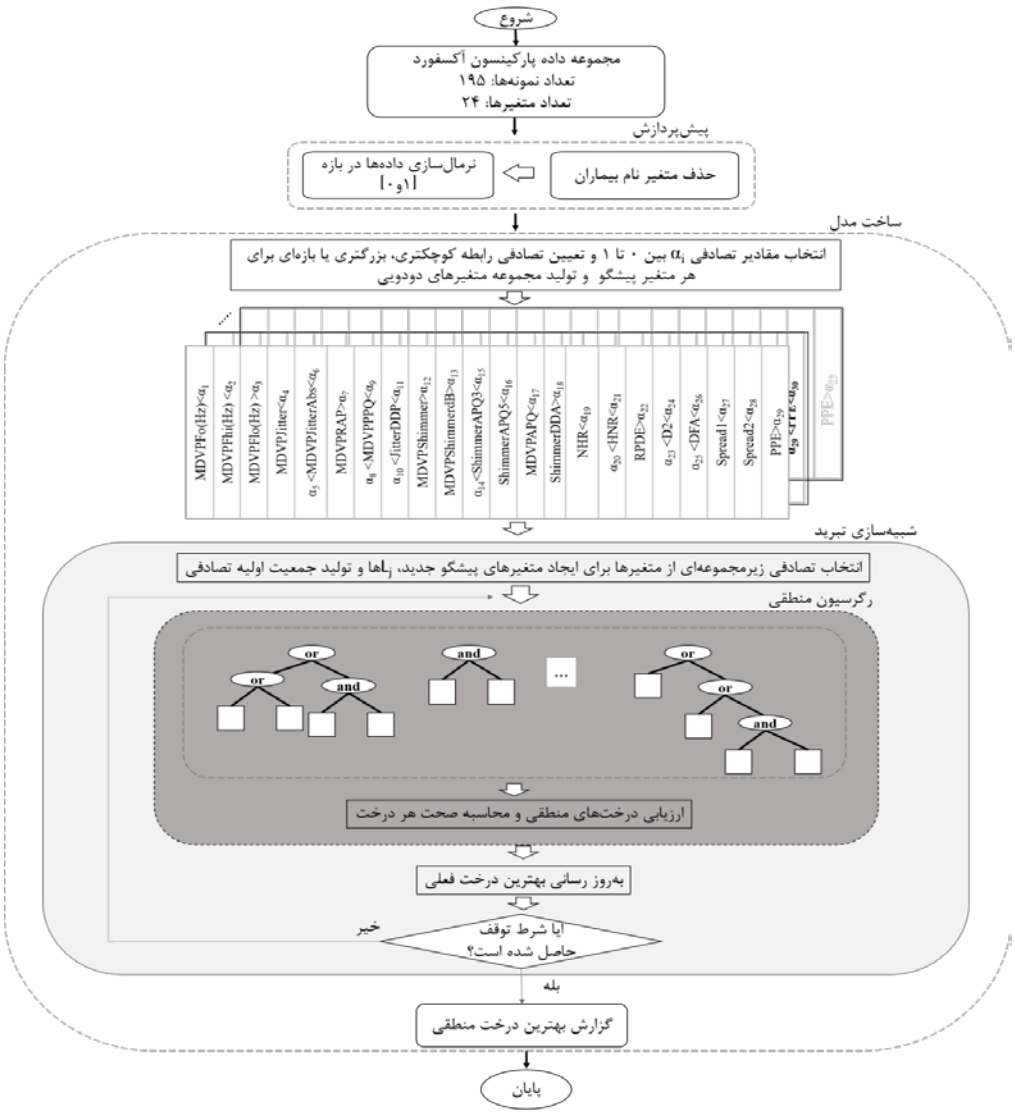
آخرین سطح تصادفی‌سازی مربوط به ایجاد عبارت‌های منطقی است که از متغیرهای انتخاب شده از مرحله شبیه‌سازی تبرید و عملگرهای منطقی and و or تشکیل می‌شود. این عبارت‌ها معرف قانون پیشگویی بیماری پارکینسون هستند. سپس درخت منطقی جهت ارایه‌ی قانون تولید شده، ایجاد و با توجه به مقادیر مجموعه داده ارزیابی می‌شود. به ازای هر نمونه از مجموعه داده، مقدار به‌دست آمده در ریشه به‌عنوان پاسخ الگوریتم در نظر گرفته می‌شود. مجموعه

پاسخ‌های تولید شده میزان صحت عبارت منطقی ایجاد شده را مشخص می‌کند.

میزان صحت به‌دست آمده از درخت تولید شده توسط الگوریتم رگرسیون منطقی، به‌عنوان صحت راه‌حل انتخابی الگوریتم شبیه‌سازی تبرید استفاده می‌شود. این الگوریتم پس از طی تعداد تکرارهای مشخص و رسیدن به شرط توقف، نتیجه نهایی را به‌عنوان بهترین راه‌حل مسأله گزارش می‌دهد. در این مقاله هر راه‌حل یک قانون تشخیصی برای بیماری پارکینسون است. شکل ۲ فلوچارت مدل پیشنهادی را نشان می‌دهد.

جهت ارزیابی مدل از مجموعه‌داده پارکینسون آکسفورد با ۱۹۵ رکورد و ۲۴ ویژگی (شامل ویژگی برچسب کلاسی) استفاده شد. جهت حذف تاثیر مقیاس ویژگی‌ها، این مجموعه‌داده با استفاده از روش min-max در بازه [۱و۰] طبیعی شد. از آنجا که برای مدلسازی داده‌های استاندارد نیاز به تقسیم مجموعه داده به مجموعه‌های آموزشی و آزمون نیست، مدل روی تمام مجموعه ایجاد و ارزیابی می‌شود (۲۵). جهت تولید قوانین کوچک شاخص t در الگوریتم رگرسیون منطقی، یک انتخاب شده است. شاخص‌های الگوریتم شبیه‌سازی تبرید در جدول ۲ آمده است.

پیاده‌سازی روش پیشنهادی در نرم‌افزار متلب و روی سیستمی با مشخصات پردازشی Intel ® Core ™ i7-8550U CPU @ 1.80GHz 1.99 GHz و ۱۲.۰ GB حافظه RAM و سیستم عامل ۶۴ بیتی اجرا شده است.



شکل ۲ - فلوچارت مدل پیشنهادی جهت ساخت درخت منطقی و تولید قانون تشخیصی مجموعه داده پارکینسون آکسفورد

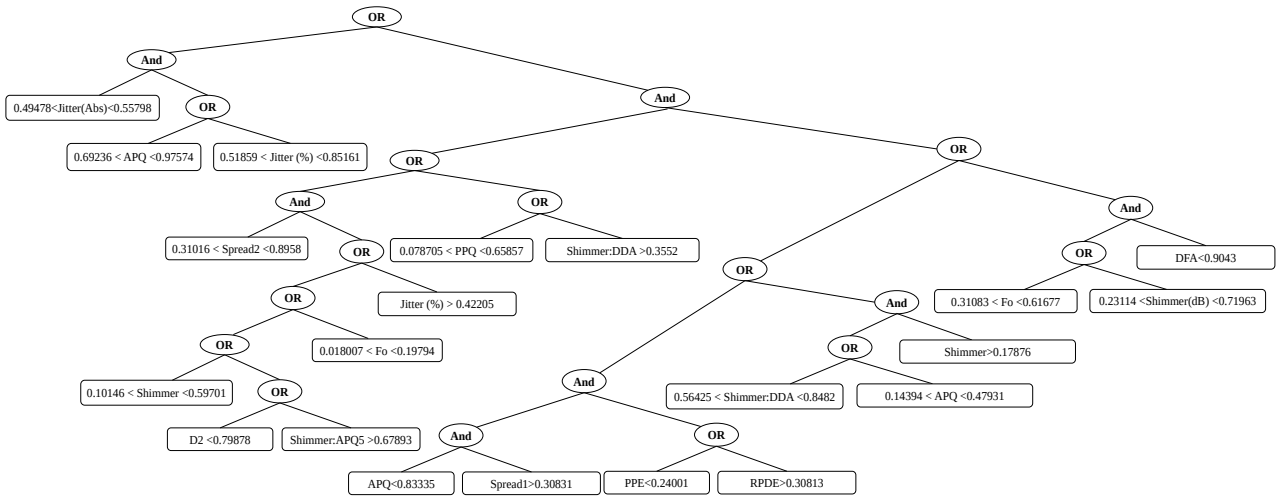
جدول ۲- شاخص‌های الگوریتم شبیه‌سازی تبرید در مدل پیشنهادی

نام شاخص	مقدار شاخص
مجموعه توابع	توابع منطقی and و or
مجموعه متغیرهای پیشگو	ویژگی‌های مجموعه داده پارکینسون آکسفورد
تابع هدف	صحت مدل (تعداد عناصری که به‌درستی دسته‌بندی شدند).
اندازه جمعیت اولیه	۱۰۰۰ فرد
نحوه تولید جمعیت اولیه	تصادفی یکنواخت
شرط توقف	عدم تغییر بهترین تناسب در ۱۰۰ نسل متوالی

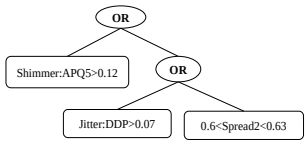
نتایج

از میان قوانین تولید شده، دو قانون یکی با تأکید بر حداکثر میزان صحت و دیگری با تأکید بر مجموعه قوانین فشرده با قابلیت تفسیرپذیری بالا انتخاب شده است. شکل ۲ قوانین به‌دست آمده را نشان می‌دهد. مقادیر نوشته شده برحسب مقدار طبیعی شده هستند.

قانون اول متشکل از ۲۱ عبارت منطقی است که عبارت‌های ایجاد شده روی متغیرهای پیشگو Fo، Jitter (%)، Jitter(Abs)، PPQ، Shimmer، Shimmer (dB)، Shimmer:APQ5، APQ، DDA، RPDE، DFA، Spread 1، Spread2، D2 و PPE تشکیل شده است. در این میان متغیرهای APQ و Shimmer:DDA با سه بار انتخاب و Fo و Shimmer هریک با دو بار انتخاب بیشترین تعداد دفعات انتخاب را در این عبارت داشتند. از میان این ۲۱ مؤلفه، ۱۰ مؤلفه از اندازه‌گیری‌های تغییرات در دامنه، سه مؤلفه از اندازه‌گیری‌های تغییرات در فرکانس، سه مؤلفه از اندازه‌گیری‌های غیرخطی تغییرات فرکانسی، دو مؤلفه از اندازه‌گیری‌های پیچیدگی پویایی غیرخطی، دو مؤلفه میانگین فرکانس اساسی صدا و یک مؤلفه از مقیاس‌پذیری



الف) قانون منتخب برحسب معیارهای کارایی: صحت ۹۲.۳۱ درصد، حساسیت ۸۵.۴۲ درصد و ویژگی ۹۴.۵۶ درصد



ب) قانون منتخب با تاکید بر تفسیرپذیری: صحت ۷۸.۹۷ درصد، حساسیت ۷۷.۰۸ درصد و ویژگی ۷۹.۵۹ درصد

شکل ۳- قوانین به‌دست آمده توسط مدل پیشنهادی جهت تشخیص بیماری پارکینسون روی مجموعه داده پارکینسون آکسفورد

فرکانسی انتخاب شده است. صحت این قانون روی مجموعه داده ۹۲/۳۱ درصد، حساسیت ۸۵/۴۲ درصد و ویژگی ۹۴/۵۶ درصد به‌دست آمده است.

قانون دوم متشکل از سه عبارت منطقی است که از ترکیب متغیرهای پیشگو Shimmer:APQ5، Jitter:DDP و Spread2 بوده و به صحت ۷۸/۹۷ درصد، حساسیت ۷۷/۰۸ درصد و ویژگی ۷۹/۵۹ درصد رسید. این قانون با تأکید بر فشرده‌ترین قانون با تفسیرپذیری بالا و صحت قابل قبول به‌دست آمده است.

بحث

استخراج قوانین مؤثر موجود در داده‌های پزشکی با دو شاخص صحت و تفسیرپذیری بالا در افزایش دقت و سرعت تشخیص بیماری توسط متخصصان بسیار ضروری است. در این مقاله یک مدل ترکیبی از الگوریتم شبیه‌سازی تبرید و رگرسیون منطقی جهت تعیین عوامل موثر بر مجموعه داده پارکینسون و تولید قانون تشخیصی با قابلیت تفسیرپذیری بالا ارایه شده است (۲۶). با توجه به قابلیت درخت

منطقی ایجاد شده در تعیین شاخص‌های بیماری پارکینسون، قوانین به‌دست آمده علاوه بر قابلیت تشخیص این بیماری، عوامل مؤثر را نیز شناسایی می‌کنند (۲۷).

با توجه به قوانین به‌دست آمده، ویژگی‌های Shimmer: APQ5 و Spread2 در هر دو قانون به‌عنوان شاخص‌های مؤثر انتخاب شده‌اند. با توجه به قانون دوم اگر مقدار Jitter:DDP از ۰/۰۱ بیشتر باشد یا spread2 بین ۰/۲۷ و ۰/۲۹ بوده و همزمان Shimmer:APQ5 از ۰/۰۲ بیشتر باشد با احتمال ۷۹ درصد فرد به بیماری پارکینسون مبتلا است. این سه مؤلفه به ترتیب اندازه‌گیری‌های مربوطبه تغییرات در اندازه‌گیری فرکانسی، تغییرات فرکانسی غیرخطی و تغییرات در دامنه صداست. ویژگی Spread2 توسط (۷، ۱۳ و ۲۲) به‌عنوان یکی از چهار شاخص اصلی تشخیص پارکینسون روی این مجموعه داده انتخاب شده است. همچنین ویژگی‌های Spread2 و Jitter:DDP توسط رحمان و همکاران (۱۲) به‌عنوان دو شاخص از چهار شاخص انتخابی تأیید شدند.

در میان کارهای مورد بررسی روی این مجموعه داده، سه مقاله (۱۷، ۲۱ و ۲۲) مدلی جهت استخراج قانون ارایه داده‌اند؛ به‌دلیل عدم ارایه قوانین توسط آبی‌زاده و آبیف (۲۱) و عدم ذکر صحت مجموعه قوانین توسط رودریگز و کارونانیدی (۲۲) امکان مقایسه جامع مدل پیشنهادی با کارهای ذکر شده برحسب صحت و تفسیرپذیری وجود ندارد.

با این حال با توجه به گزارش‌های ذکر شده در این مقالات، تعداد قوانین بدست آمده توسط توسط آبی‌زاده و آبیف (۲۱) و رودریگز و کارونانیدی (۲۲) به‌ترتیب ۱۶ و ۸۱ قانون ذکر شده است که نسبت به تک قانون ارایه شده در این مطالعه نشان‌دهنده پیچیدگی بالای مدل‌ها و کاهش تفسیرپذیری است. لذا قانون به‌دست آمده توسط مدل پیشنهادی نسبت به دو روش ذکر شده، فشرده‌ترین قانون بوده و در نتیجه به لحاظ قابلیت تفسیرپذیری برتری دارد.

وانگ و همکاران در (۱۷) تعداد قوانین به‌دست آمده را ۸ قانون ذکر کردند که در مقاله تنها دو قانون آن ارایه شده است. هریک از قوانین ذکر شده شامل بررسی رابطه‌ی بازه‌ای برای ۴۴ م فرکانسی مؤلفه (۲۲ متغیر برای افراد سالم و ۲۲ متغیر برای افراد مبتلا به پارکینسون) بوده و صحت مدل ۸۲/۴ درصد گزارش شده است. در مقایسه با قانون به‌دست آمده توسط مدل پیشنهادی شامل ۲۱ مؤلفه و صحت ۹۲/۳۱ درصد، قانون پیشنهادی با کمتر از نصف تعداد قوانین، میزان صحت آن ۹/۹۱ درصد بیشتر از روش وانگ و همکاران در (۱۷) است.

با توجه به اینکه در قوانین بالینی، سادگی قوانین یک شاخص مهم در تفسیرپذیری و به استفاده آسان توسط پزشک است، قانون به‌دست آمده در این مقاله از یک سو با داشتن تنها سه ویژگی کوچک بوده و به خاطر سپردن آن توسط پزشک آسان است؛ در نتیجه نیازی به اجرای برنامه کامپیوتری در هنگام مراجعه بیمار وجود ندارد. از سوی دیگر به‌دلیل بالابودن دقت، حساسیت و ویژگی قادر است هم افراد دارای پارکینسون و همچنین افراد سالم را به‌درستی تشخیص دهد.

با توجه به کوچک بودن قانون و قابلیت پیاده‌سازی آسان آن، می‌توان برنامه‌ی کاربردی برای گوشی‌های هوشمند طراحی کرد که به‌طور خودکار از روی سیگنال صوتی فرد احتمال داشتن بیماری پارکینسون او را بررسی نمود. این امر نیاز به بستری دارد که آلارم آن به‌طور خودکار به پزشک یا یکی از اعضای خانواده فرستاده شود تا اقدامات به موقع و مراجعه به پزشک صورت گیرد.

تشکر و قدردانی

بدین‌وسیله نویسندگان از دانشگاه صنعتی خاتم‌الانبیاء به جهت حمایت‌های خود در امر پژوهش صمیمانه تشکر می‌نماید.

References

1. Cai Z, Gu J, Wen C, Zhao D, Huang C, Huang H, et al. An Intelligent Parkinson's Disease Diagnostic System Based on a Chaotic Bacterial Foraging Optimization Enhanced Fuzzy KNN Approach. Computational and Mathematical Methods in Medicine 2018;2018:2396952. doi: 10.1155/2018/2396952
2. Li Y, Swift S, Tucker A. Modelling and analysing the dynamics of disease progression from cross-sectional studies. Journal of Biomedical Informatics 2013;46:266-74. doi: 10.1016/j.jbi.2012.11.003
3. Devarajan M, Ravi L. Intelligent cyber-physical system for an efficient detection of Parkinson disease using fog computing. Multimedia Tools Appl 2019;78:32695-719. doi: 10.1007/s11042-018-6898-0
4. Olivares R, Munoz R, Soto R, Crawford B, Cárdenas D, Ponce A, et al. An optimized brain-based algorithm for classifying parkinson's disease. Appl Sci 2020;10. doi: 10.3390/app10051827
5. Pahuja G, Nagabhushan TN. A comparative study of existing machine learning approaches for parkinson's disease detection. IETE J Res 2018. doi: 10.1080/03772063.2018.1531730
6. Understanding Parkinson's-Statistics: Parkinson's Foundation; [2020-05-05 19:26:14]. Available from: https://www.parkinson.org/Understanding-Parkinsons/Statistics.
7. Hariharan M, Polat K, Sindhu R. A new hybrid intelligent system for accurate detection of Parkinson's disease. Comput Methods Programs Biomed 2014;113:904-13. doi: 10.1016/j.cmpb.2014.01.004
8. Marar S, Swain D, Hiwarkar V, Motwani N, Awari A, editors. Predicting the occurrence of parkinson's disease using various Classification Models 2018: Institute of Electrical and Electronics Engineers Inc.
9. Avci D, Dogantekin A. An expert diagnosis system for parkinson disease based on genetic algorithm-wavelet kernel-extreme learning machine. Parkinson's Dis 2016;2016. doi: 10.1155/2016/5264743
10. Xiong Y, Lu Y. Deep Feature Extraction from the Vocal Vectors Using Sparse Autoencoders for Parkinson's Classification. IEEE Access 2020;8:27821-30. doi: 10.1109/ACCESS.2020.2968177
11. Kaninika, Tayal A, editors. Determination of Parkinson's disease utilizing Machine Learning Methods2018: Institute of Electrical and Electronics Engineers Inc.
12. Dash S, Abraham A, Atta ur R. Kernel Based Chaotic Firefly Algorithm for Diagnosing Parkinson's Disease. In: Gandhi N, Madureira AM, Abraham A, Varela ML, editors.: Springer Verlag; 2020.p.176-88.

13. Dash S, Thulasiram R, Thulasiraman P, editors. An Enhanced Chaos-Based Firefly Model for Parkinson's Disease Diagnosis and Classification 2018: Institute of Electrical and Electronics Engineers Inc.

14. Kadam VJ, Jadhav SM. Feature ensemble learning based on sparse autoencoders for diagnosis of Parkinson's disease. Advances in Intelligent Systems and Computing: Springer Verlag;2018.p.567-82.

15. Lee SH. Feature selection based on the center of gravity of BSWFMs using NEWFM. Engineering Applications of Artificial Intelligence 2015;45:482-7. doi: 10.1016/j.engappai.2015.08.003

16. Wibawa MS, Nugroho HA, Setiawan NA, editors. Performance evaluation of combined feature selection and classification methods in diagnosing Parkinson disease based on voice feature2016: Institute of Electrical and Electronics Engineers Inc.

17. Hsieh YZ, Su MC, Wang PC. A PSO-based rule extractor for medical diagnosis. Journal of Biomedical Informatics 2014;49:53-60. doi: 10.1016/j.jbi.2014.05.001

18. Seera M, Lim CP. A hybrid intelligent system for medical data classification. Expert Systems with Applications 2014;41:2239-49. doi: 10.1016/j.eswa.2013.09.022

19. Shortliffe E, Cimino JJ. Biomedical Informatics: Computer Applications in Health Care and Biomedicine (Health Informatics) 4th ed. Springer;2013 December 4.

20. Sujatha R, Ephzibah EP, Dharinya S, Uma Maheswari G, Mareeswari V, Pamidimarri V. Comparative study on dimensionality reduction for disease diagnosis using fuzzy classifier. International Journal of Engineering and Technology(UAE) 2018;7:79-84.

21. Abiyev RH, Abizade S. Diagnosing parkinson's diseases using fuzzy neural system. Computational and Mathematical Methods in Medicine 2016;2016. doi: 10.1155/2016/1267919

22. Karunanithi D, Rodrigues P. A fuzzy rule-based diagnosis of parkinson's disease. Lecture Notes in Computational Vision and Biomechanics: Springer Netherlands;2019.p.1229-38.

23. Ruczinski I, Kooperberg C, LeBlanc M. Logic Regression. Journal of Computational and Graphical Statistics 2003;12:475-511. doi: 10.1198/1061860032238

24. UCI Machine Learning Repository: Parkinsons Data Set [cited 2020 2020-05-12 14:40:16]. Available from: http://archive.ics.uci.edu/ml/datasets/Parkinsons/.

25. Ahouz F, Sadehvand M, Golabpour A. Extracting rules for diagnosis of diabetes using genetic programming. International Journal of Health Studies 2019;5:23-32. doi: 10.22100/ijhs.v5i3.691

26. Little M, McSharry P, Roberts S, Costello D, Moroz I. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Biomedical Engineering Online 2007;6:23.

27. Wang Y-Y, Li J. Feature-selection ability of the decision-tree algorithm and the impact of feature-selection/extraction on decision-tree results based on hyperspectral data. International Journal of Remote Sensing - Int J Remote Sens 2008;29:2993-3010. doi: 10.1080/01431160701442070

Journal of Knowledge & Health
in Basic Medical Sciences 2020; Vol 15, No 2

DOI: 10.22100/jkh.v15i2.2411

Original Article

A Novel Diagnostic Rule for Parkinson's Disease Based on a Hybrid Extraction Method

Fatemeh Ahouz (M.Sc.)[1], Amin Golabpour (Ph.D.)[2*]

1- Instructor, Dept. of Computer Engineering, School of Engineering, Behbahan Khatam Alanbia University of Technology, Behbahan, Iran.

2- Assistant Professor, School of Paramedical, Shahroud University of Medical Sciences, Shahroud, Iran.

Received: 23 June 2020**, Accepted:** 1 August 2020

Abstract:

Introduction: *Parkinson's disease has become an increasing public health issue that its symptoms become more severe over time. Early diagnosis and treatment of this disease leads to improving the skills, abilities and performance of patients in daily life. In order to diagnose the disease early, it is necessary to produce clinical decision-making assistance systems that are able to detect the diagnostic rules of the disease.*

Methods: *This study provides an automatic way to extract novel diagnostic rules for Parkinson's disease. The proposed method is based on logic regression and simulated annealing algorithm. To evaluate the method, the Oxford Parkinson's data set was used, which contains 22 biomedical voice measurements from 31 people, 23 with Parkinson's disease. The dataset has 195 voice recording from these individuals.*

Results: *The results include two diagnostic rules; If high accuracy was the main concern, a new rule has been proposed that includes 21 logical statements that have an accuracy of 92.31%, a sensitivity of 85.42%, and a specificity of 94.56%. However, for real-time systems and clinical decision-making assistance with high interpretability, a rule consisting of 3 logical statements has been proposed, which has an accuracy of 78.97%, a sensitivity of 77.08% and a feature of 79.59%.*

Conclusion: *The results show the high power of interpretability and reliability of the proposed rules in the diagnosis of Parkinson's disease, which can be used in the implementation of remote diagnostic systems.*

Keywords: Parkinson, Logic regression, Simulated annealing, Rule extraction.

Conflict of Interest: No

***Corresponding author:** A. Golabpour**, Email:** a.golabpour@shmu.ac.ir*

Citation: Ahouz F, Golabpour A. A novel diagnostic rule for parkinson's disease based on a hybrid extraction method. Journal of Knowledge & Health in Basic Medical Sciences 2020;15(2):42-50.