مقاله پژوهشی

مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره نهم شماره اول. ص ۲۴-۱۲، ۱۴۰۱

توسعه سیستم ترکیبی مبتنی بر داده‌کاوی و وب معنایی برای تشخیص بیماری اوتیسم

سمن شیشه چی[۱]، سید یاشار بنی‌هاشم[۱]*

• دریافت مقاله: ۱۴۰۰/۱۲/۶ • پذیرش مقاله: ۱۴۰۱/۲/۷

مقدمه: اوتیسم به عنوان یک اختلال سیستم عصبی است و از آنجایی که تشخیص مستقیمی برای آن وجود ندارد، داده‌کاوی می‌تواند به تشخیص این بیماری کمک شایانی کند. آنتولوژی به عنوان ستون وب معنایی، یک پایگاه دانش با قابلیت اشتراک‌پذیری و استفاده قابلیت مجدد می‌تواند تأییدی بر درستی سیستم‌های تشخیص بیماری باشد. هدف این تحقیق ارائه سیستمی جهت تشخیص کودکان اوتیسمی با روشی مرکب از وب معنایی و داده کاوی می‌باشد.

روش: داده‌های موجود، برگرفته از بانک داده‌ای UCI می‌باشد. در مجموع ۲۹۲ رکورد داده‌ای موجود بود که ۸۰ درصد این داده‌ها یعنی ۲۳۴ رکورد جهت مدل‌سازی به وسیله درخت تصمیم استفاده شدند. اطلاعات مربوط به بیماران و بیماری اوتیسم در قالب دانش در آنتولوژی با استفاده از نرم‌افزار Protégé 5 ارائه داده شدند. آنتولوژی دارای ۴ کلاس و ۱۲ خصوصیت جهت برقراری ارتباط بین نمونه‌های موجود در کلاس‌ها بود. قانون‌های استخراج شده از درخت تصمیم، به شکلی قابل فهم (SWRL) برای تفسیر در آنتولوژی توسط یک مبدل، تبدیل شدند.

نتایج: سالم بودن و یا نبودن کودک از روی قوانین به دست آمده در درخت تصمیم قابل تشخیص است. در ضمن، خروجی آنتولوژی با استفاده از تفسیر ۲۵ قانون، تشخیص کودکان بیمار با استفاده از درخت تصمیم را تأیید کرد. همچنین ارزیابی آنتولوژی، درستی آن را تأیید کرد.

نتیجه‌گیری: با توجه به همسان بودن خروجی آنتولوژی و درخت تصمیم در رابطه با تشخیص بیماری، می‌توان به دقت و درستی روش ارائه شده تأکید کرد.

کلیدواژه‌ها: اوتیسم، داده کاوی، درخت تصمیم، آنتولوژی، سیستم مبتنی بر قانون

•ارجاع: شیشه چی سمن، بنی‌هاشم سید یاشار. توسعه سیستم ترکیبی مبتنی بر داده‌کاوی و وب معنایی برای تشخیص بیماری اوتیسم. مجله انفورماتیک سلامت و زیست پزشکی ۱۴۰۱؛ ۹(۱): ۲۴-۱۲.

۱. دکترای مهندسی فناوری اطلاعات، استادیار مرکز آموزش عالی فنی و مهندسی بویین زهرا، بویین زهرا، قزوین، ایران

*** نویسنده مسئول:** سمن شیشه چی

آدرس: مرکز آموزش عالی فنی و مهندسی بویین زهرا

•شماره تماس: ۰۹۱۲۰۷۴۱۲۷۳ • **Email:** shishechi@bzte.ac.ir

[Downloaded from jhbmi.ir on 2026-11-09]

مقدمه

اوتیسم، دسته‌ای از اختلالات تکاملی سیستم عصبی می‌باشد که از جمله نشانه‌های اصلی آن می‌توان به نقص در تعاملات اجتماعی، ارتباطات و نیز وجود رفتارهای تکراری و علائق محدود اشاره نمود [۱]. تاکنون هیچ راهی برای پیشگیری از ابتلای کودک به اوتیسم یافت نشده است، ولی با این حال با تشخیص به موقع، بیماری بهبود یافته و مهارت‌های اجتماعی و کلامی کودک ارتقاء پیدا می‌کند. با توجه به این موضوع، می‌توان از داده‌کاوی برای تشخیص و پیش‌بینی بیماری در تست‌های بالینی استفاده کرد. امروزه با توسعه تکنولوژی، پزشکان قادر هستند اطلاعات دقیق‌تری را در مورد بیماران جمع‌آوری نمایند و برای تحلیل این حجم اطلاعات از الگوریتم‌های داده‌کاوی کمک بگیرند. آنتولوژی [۲] به عنوان ستون وب معنایی، یک تکنولوژی برای به اشتراک گذاشتن دانش به زبان قابل فهم برای ماشین است. تمام دانش مربوط به بیماری و ارتباطات بین آن‌ها را می‌توان به وسیله آنتولوژی به نمایش کشید و همچنین با استفاده از قوانین تعریف شده در آنتولوژی می‌توان دانش جدیدی به آنتولوژی اضافه کرد که در تشخیص بیماری نقش بسزایی دارد [۳]. از آنجایی که آنتولوژی مبتنی بر قانون می‌باشد، بنابراین باید الگوریتم‌های داده‌کاوی انتخابی، قابلیت تولید قوانین را داشته باشند. درخت تصمیم [۴] از مدل‌های موجود در داده کاوی قابلیت تولید قوانین را برای داده‌های مورد نظر دارد که در تشخیص بیماری بسیار استفاده شده است [۸-۵]. خیلی از محققان اعتقاد دارند که درخت تصمیم و آنتولوژی در کنار هم بسیار جذاب و موضوعی قابل بحث می‌باشند [۱۳-۹]، اغلب سیستم‌های تشخیص در حوزه پزشکی از آنتولوژی [۱۵، ۱۴] و یا درخت تصمیم [۱۸-۱۶] استفاده کرده‌اند. اساسی‌ترین قسمت آنتولوژی در بخش تفسیر داده و تولید حقایق، قوانین مبتنی بر شکل (Web Semantic Rule Language)SWRl می‌باشد [۱۹]. از آنجایی که گاهی اوقات یکسری قوانین با توجه به سلیقه طراح آنتولوژی ممکن است در نظر گرفته نشود؛ بنابراین اگر قبل از تولید قوانین توسط آنتولوژی، بتوان آن‌ها را به صورت آماده در دست داشت، آنگاه دقت کار بالا خواهد رفت. برای بالا بردن دقت کار در آنتولوژی در این مورد بهترین راه استفاده از قوانین به دست آمده توسط درخت تصمیم برای مدل مورد نظر می‌باشد.

Al-Diabat [۲۰] از روش داده‌کاوی فازی برای تشخیص و شناسایی کودکان ۴ تا ۱۱ سال استفاده کرد که بر مبنای الگوریتم فازی پیاده‌سازی شده است. این تحقیق با استفاده از الگوریتم‌های (Fuzzy Unordered Rule Induction Algorithm)، FURIA، JRIP (Ripple Down Rule Learner)، RIDOR و PRISM و مجموعه داده ASDTest انجام شد و در نتیجه به دقت ۹۱/۳۵٪ و حساسیت ۹۱/۴۰٪ برای الگوریتم FURIA رسید. همچنین، Thabtah و Peebles [۲۱] با استفاده از الگوریتم‌های (Repeated Incremental Pruning to Produce Error Reduction)RIPPER، RIDOR (NonNested Generalization)، Nnge (Bootstrap Aggrigating)، Bagging (A Classification and Regression tree)، CART، PRISM، Adaptive Boosting، AdaBoost، C4.5 و الگوریتم پیشنهادی جدیدی از یادگیری ماشین به نام (Recursive Maximum Likelihood) RML بر روی مجموعه داده ASDTest مطالعه انجام داده و به دقت ۹۴/۴٪ با استفاده از الگوریتم RML دست یافتند. در مطالعه Achenie و همکاران [۲۲] یک گروه تحقیقاتی با استفاده از تکنیک (Fast Artificial Neural Network) FaNN و مجموعه داده M-CHAT-R که شامل ۱۴۹۹۵ داده از کودکان ۱۶ تا ۳۰ ماهه بود، به دقت ۹۹/۷۲٪ دست یافتند که دقت قابل‌ملاحظه‌ای است. همچنین Kong و همکاران [۲۳] با استفاده از (Deep Neural Network) DNN به دقت ۹۰/۳۹٪ رسیدند. در تحقیقی دیگر [۲۴] محققان با استفاده از الگوریتم‌های رگرسیون به دقت ۹۸/۹٪ و با استفاده از SVM (Support Vector Machine) به دقت ۹۹/۸۳٪ دست یافتند. Alwidian و همکاران [۲۵] به بررسی تأثیر طبقه‌بندی انجمنی روی پیش‌بینی و تشخیص بیماری اوتیسم با استفاده از الگوریتم‌های (Classification Based on Association Rules)CBA، (Multiple Class- Classification Based on Association Rules)CMAR،MCAR، FCBA، ECBA و WCBA (Weighted Classification Based on Association Rules) پرداختند، که بهترین عملکرد را الگوریتم WCBA با دقت ۹۷٪ دارا بود. در مطالعه Shuvo و همکاران [۲۶] با استفاده از الگوریتم جدول تصادفی دقت ۹۵/۷٪ را به دست آوردند. در تحقیقی دیگر [۲۷] با استفاده از داده‌کاوی فازی و الگوریتم FURIA به تحلیل رفتار بیماران اوتیسمی پرداخته شده و روشی برای غربالگری بیماران ارائه داده است. در این روش نیز از داده‌های ASD Test استفاده شده است که نتیجه این روش دقت ۹۲/۱۴٪ و حساسیت ۹۲/۴۹٪ را به همراه داشته است. در پژوهش دیگری

[Downloaded from jhbmi.ir on 2026-11-09]

[۲۸] با استفاده از دسته‌بندی اطلاعات بر اساس سن افراد و روش‌های مختلف انتخاب مشخصه از جمله: Concatenation، Spearman، Mutual Information و Covariance و الگوریتم SVM به ترتیب به دقت (۸۶٪، ۶۹٪، ۷۸٪، ۸۰٪ و ۹۵٪) و حساسیت (۹۱٪، ۸۰٪، ۸۰٪، ۸۷٪، ۹۱٪) رسیدند. Subbaraju و همکاران [۲۹] به تشخیص بیماری اوتیسم با استفاده از الگوریتم SVM پرداختند. در این تحقیق اطلاعات در ۴ دسته براساس سن افراد طبقه‌بندی شده‌اند که شامل (خانم ≤ ۱۸، آقا ≤ ۱۸، خانم >۱۸ و آقا >۱۸) است. نتیجه این تحقیق به ترتیب دقت (۸۶/۷٪، ۷۸/۶٪، ۹۵٪ و ۸۵/۴٪) رسیده است. در مطالعه Chowdhury و Iraj [۳۰] با استفاده از الگوریتم‌های Naïve Bayes، kNN، Logistic Regression، Gradient Boosting، SVM، Decision Tree و MLP Classifier و بانک اطلاعاتی ASD Test به تشخیص بیماری با استفاده از هوش مصنوعی پرداختند که بهترین نتیجه مربوط به الگوریتم SVM است که دقت ۹۵٪ را به همراه داشته است. با وجود این که الگوریتم MLP(Multilayer Perceptron) دقت بالاتری یعنی ۱۰۰٪ را داشته؛ اما به دلیل AUC، Precision و F1 پایین این الگوریتم نتیجه ارزیابی الگوریتم SVM مطلوب‌تر بوده است. در مطالعه Musa و همکاران [۳۱] با استفاده از دیتاست AQ-10 و الگوریتم‌های Random Forest، Naïve Bayes، J48 Decision tree و SVM به تشخیص بیماری روی افراد در سنین کودکی و نوجوانی پرداخته‌اند که نتایج برای همه الگوریتم‌های ذکر شده بسیار مناسب است. نتیجه ارزیابی این الگوریتم‌ها دقت و صحت ۱۰۰٪ را به همراه داشته است. برخی محققین جهت تحقیق بر روی بیماری اوتیسم از آنتولوژی استفاده کرده‌اند. Galán-Mena و همکاران [۳۲] آنتولوژی جهت تشخیص بیماری اوتیسم برای ۳۲ کودک را توسعه دادند. این سیستم جهت تشخیص بیماری در کنار آنتولوژی از ابزارهای هوشمند دیگری کمک گرفت. محققان دیگری [۳۳] سعی داشتند تا با استفاده از آنتولوژی GEN که دارای اطلاعاتی ساختاریافته و قابل‌پردازش از عملکرد ژن‌ها و محصولات ژنتیکی می‌باشد، به تشخیص بیماری اوتیسم کمک کنند. همچنین آنتولوژی موجود در مطالعه Tu و همکاران [۳۴] قابلیت تشخیص بیماری را ندارد و فقط جنبه تحقیقاتی داشته و قابل استفاده برای تکنسین‌ها و افراد درمان‌گرا و پزشکان نیست. در رابطه با بیماری اوتیسم پژوهش‌های متفاوتی با استفاده از الگوریتم‌های مختلف یادگیری ماشین و داده‌کاوی صورت گرفته است ولی متأسفانه سیستمی جهت تشخیص بیماری اوتیسم که بتواند از قابلیت‌های ترکیبی داده‌کاوی و وب معنایی استفاده کند دیده نمی‌شود. در حالی که با داشتن چنین سیستمی می‌توان تشخیصی درست با دقت بسیار بالا و همچنین قابلیت اشتراک‌گذاری و استفاده مجدد پایگاه دانش در سیستم‌های درمانی دیگری داشت. هدف اصلی در این تحقیق طراحی و پیاده‌سازی مدلی مبتنی بر آنتولوژی می‌باشد که از درخت تصمیم برای تشخیص بیماری اوتیسم استفاده کند تا بتواند از قابلیتهای هر دو تکنیک، در بهبود تشخیص بیماری نسبت به روش‌های قبلی تأثیر بسزایی داشته باشد.

روش

همان‌طور که پیش‌تر اشاره شد، فرآیند تحقیق از دو بخش داده‌کاوی و ایجاد پایگاه دانش توسط آنتولوژی تشکیل شده است. شکل ۱، فرآیند طراحی و پیاده‌سازی در این تحقیق را نشان داده است. در بخش داده‌کاوی، درخت تصمیم به دلیل تولید قوانین برای آنتولوژی و داشتن دقت بالا نسبت به سایر الگوریتم‌های دسته‌بندی در داده‌کاوی انتخاب شده است. برای انجام این تحقیق از متغیرهای موجود در بانک UCI Machine Learning Repository استفاده شده است. بعد از طراحی مدل مبتنی بر درخت تصمیم، از روی برگ‌های –درخت تولید شده و همچنین قوانین ایجاد شده از روی این درخت می‌توان کودکان اوتیسم و سالم را تشخیص داد. این قوانین بر مبنای اگر– آنگاه می‌باشند. فاز دوم تحقیق مربوط به ساخت آنتولوژی و ایجاد پایگاه دانش می‌باشد. از روی اطلاعات مربوط به بیماران و همچنین علائم بیماری برای هر کودک، آنتولوژی مربوطه با استفاده از نرم‌افزار Protege ایجاد می‌شود. کلیه کلاس‌ها و خصوصیات لازم جهت برقراری ارتباط بین اعضای کلاس‌ها در آنتولوژی ایجاد می‌شوند. سپس، قوانین تولید شده از درخت تصمیم به شکل SWRL جهت تفسیر در مفسر آنتولوژی تبدیل می‌شوند. این تبدیل دستی نبوده و با استفاده از مبدلی که توسط اعضای گروه به کمک نرم‌افزار متلب پیاده‌سازی شده است به صورت خودکار انجام می‌گیرد. در نهایت، با استفاده از مفسر Drool، قوانین موجود در آنتولوژی جهت تشخیص کودکان اوتیسم تفسیر می‌شوند.

[Downloaded from jhbmi.ir on 2026-11-09]



شکل ۱: مراحل انجام تحقیق

جامعه آماری و آماده‌سازی داده

جامعه آماری این پژوهش از مجموعه داده غربالگری اوتیسم کودکان (Autism Screening Children) که از مخزن UCI که مرجعی برای داده‌های یادگیری ماشین می‌باشد انتخاب گردیده است. تعداد ۲۹۲ نمونه داده در ابتدا در دسترس است که ۲۱ ویژگی برای تشخیص اوتیسم در کودکان مورد استفاده قرار می‌گیرد. ۱۰ مورد از این ویژگی‌ها در واقع پاسخ به سؤالاتی در رابطه با علائم کودک می‌باشد که در جدول ۱ نمایش داده شده‌اند. ۱۲ ویژگی دیگر شامل سن، جنسیت، قومیت، داشتن بیماری زردی در بدو تولد، سابقه اوتیسم در خانواده، کشور محل اقامت، استفاده از اپلیکیشن تشخیص بیماری، مجموع نتیجه، گروه سنی، نسبت فرد پر کننده پاسخ نامه، اوتیسم و پاسخ به علائم کودک می‌باشد. نوع همه ویژگی‌های موجود در جدول ۱، از نوع باینری می‌باشند. در واقع پاسخ به هر سؤال در صورت مثبت بودن مقدار ۱ و در غیر این صورت مقدار ۰ را به خود می‌گیرد.

جدول ۱: مجموعه سؤالات تشخیص اوتیسم

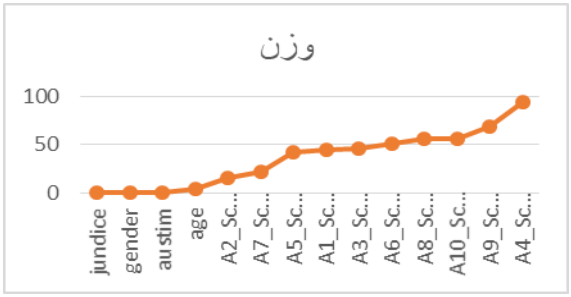
توضیحات
او معمولاً متوجه صداهای کمی می‌شود که دیگران نمی‌شود.
او معمولاً روی کلیات تمرکز می‌کند تا جزئیات کوچک
او می‌تواند در گروه‌های اجتماعی به سادگی مکالمات چندین نفر را دنبال کند.
او به راحتی می‌تواند از فعالیتی به فعالیت دیگر مشغول شود.
او نمی‌داند چگونه به مکالمه با دوستانش ادامه دهد.
او مکالمه‌های اجتماعی را به خوبی انجام می‌دهد.
وقتی که او داستان می‌خواند برایش سخت است که هدف شخصیت‌ها و احساسات آن‌ها را درک کند
وقتی او در پیش دبستانی بود از انجام بازی‌هایی که شامل نقش بازی کردن همراه با دوستانش می‌شد، لذت می‌برد.
او تنها با نگاه کردن به دیگران به راحتی می‌تواند بفهمد که آن‌ها چه فکر می‌کنند و چه حسی دارند.
برای او سخت است که دوستان جدید پیدا کند.

[Downloaded from jhbmi.ir on 2026-11-09]

طراحی و پیاده‌سازی فاز داده‌کاوی

با توجه به ساختار داده و نوع ویژگی‌ها و نیاز به داشتن خروجی به صورت قانون، درخت تصمیم برای این مطالعه انتخاب شد. برای مدل کردن، از نرم‌افزار رپیدماینر نسخه ۹,۹ استفاده شده است [۳۵]. این نرم‌افزار دارای ابزاری مانند "Split Data" می‌باشد که با استفاده از آن داده‌ها را به دو قسمت تقسیم می‌شود. قسمت اول که ۸۰٪ داده‌ها (۲۳۴ رکورد) را شامل می‌شود جهت آموزش مدل به کار گرفته شد و قسمت دوم که ۲۰٪ داده‌ها بود برای تست مدل ساخته شده به کار گرفته شد.

سپس برای بهینه سازی مدل و بالا بردن سرعت محاسبات از ابزار دیگری به نام Chi-Squared استفاده شده است [۳۶]. این ابزار برای هر ویژگی وزنی را محاسبه می‌کند که گویای اهمیت نقش آن ویژگی می‌باشد. پس از محاسبه وزن ویژگی‌ها، تمامی مقادیر از زیاد به کم مرتب شدند (شکل ۲) و از میان آن‌ها چهارده ویژگی نخست که بالاترین وزن را داشتند انتخاب شدند. در نمودار زیر وزن‌های محاسبه شده به صورت درصدی نمایش داده شده‌اند.



شکل ۲: وزن‌های مناسب هر ویژگی

مدل درخت تصمیم

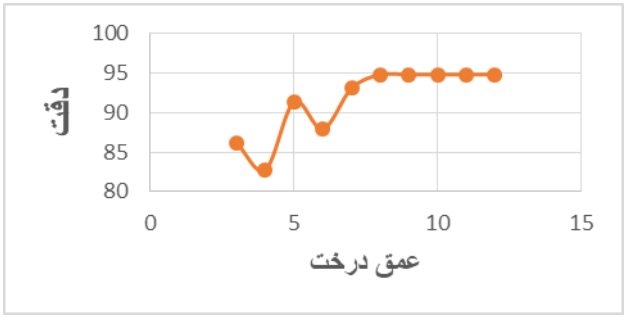
برای مدل‌سازی بر اساس درخت تصمیم باید متغیرهای متعددی از جمله معیار ساخت درخت "Criterion"، حداقل بهره "Minimal Gain"، حداکثر عمق "Maximal Depth"، حداقل سایز تقسیم "Minimal Size of Split" را مقداردهی کرد. در ادامه فرآیند مقدارهای انتخابی برای متغیرهای مذکور توضیح داده خواهد شد. پارامترهای مربوط به درخت تصمیم برای حالت‌های متفاوت ساخته شد و در هر مورد دقت مدل ساخته شده با داده‌های یکسان بررسی شد و در نهایت حالت Gini Index با دقت ۹۴/۸۳ درصد مطابق جدول ۲، انتخاب گردید.

جدول ۲: پارامترهای درخت تصمیم

پارامتر	مقدار
Gain Ratio	۹۱/۳۸
Accuracy	۸۲/۷۶
Gini Index	۹۴/۸۳
Information Gain	۸۹/۶۶

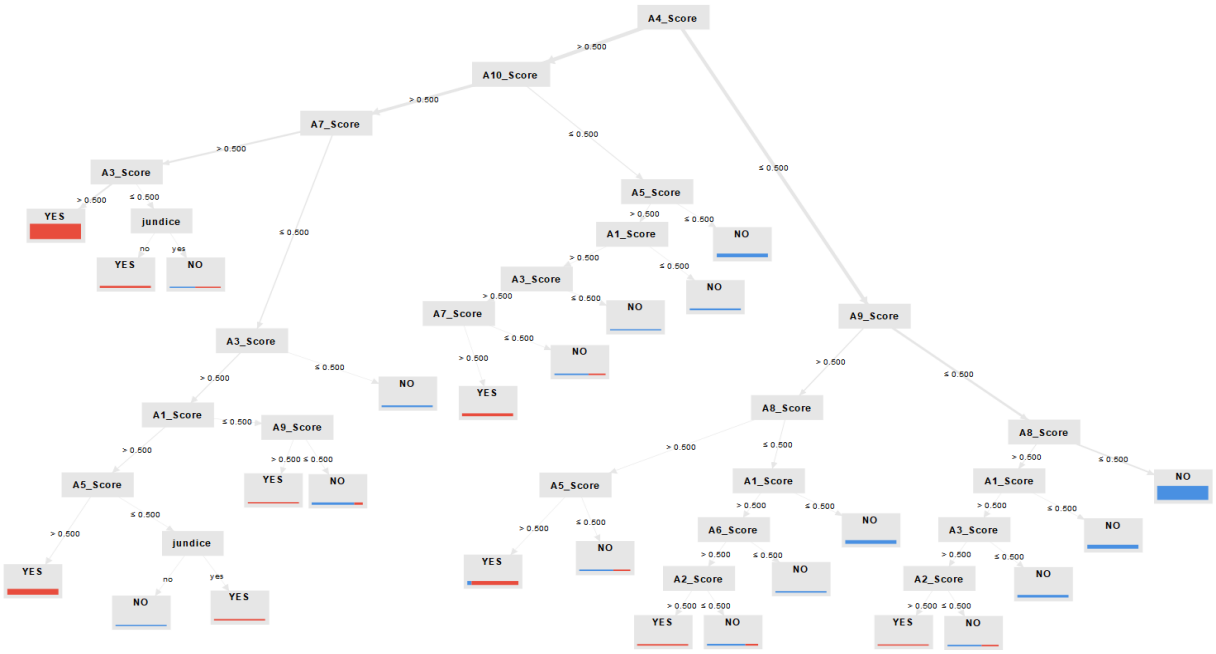
مورد بعدی تعیین حداکثر عمق درخت می‌باشد. باید به این نکته توجه داشت که تعین مقدار بهینه برای عمق می‌تواند تعداد محاسبات را برای مدل‌سازی کاهش دهد و از طرفی از پیچیدگی بی‌دلیل مدل جلوگیری کند. طی ده مرحله مقدارهای سه تا دوازده برای مدل امتحان شد. در هر مرحله دقت مدل بررسی شد. تغییر رفتار مدل را می‌توان به وضوح در شکل ۳ دید. مدل ایجاد شده، بهترین دقت را در عمق هشت از خود نشان داد و بعد از آن افزایش عمق دیگر تأثیری در دقت مدل نداشت؛ بنابراین عدد هشت برای عمق مدل انتخاب گردید.

[Downloaded from jhbmi.ir on 2026-11-09]



شکل ۳: تغییر عمق درخت

در نهایت پس از ساخت مدل درخت تصمیم که در شکل ۴ نمایش داده شده، مرحله آزمایش مدل جهت اطمینان از کارایی به دو روش آزمایش شد. در روش اول داده‌ها به نسبت هشتاد و بیست درصد به ترتیب برای آموزش و آزمایش مدل استفاده گردید. در این حالت بهترین مقدار دقت ۹۴/۸۳ درصد بود. در روش دوم از روش 10 fold Cross Validation استفاده شد. در این روش داده‌ها به ده قسمت تقسیم می‌شوند و هر دفعه یک قسمت به عنوان داده آزمایش به مدل خوراند می‌شود. سپس در نهایت میانگین دقت اعلام می‌گردد که در مورد مدل مذکور مقدار دقت ۹۰/۴۱ +/- ۵/۰۶ درصد شد. همچنین متغیرهای درستی "Precision" ۹۳/۴۶ درصد و فراخوان مجدد "Recall" ۸۶/۵۷ درصد را نشان داد که مقادیر قابل قبولی هستند.



شکل ٤: درخت تصمیم ایجاد شده

استخراج قوانین

تعداد ۲۵ برگ موجود در درخت تصمیم نشان دهنده شروط تشخیص کودکان سالم از اوتیسم می‌باشد. نرم افزار رپیدماینر، ماژولی جهت تبدیل درخت تصمیم به قوانین دارد که به کمک آن قوانین مربوط به این مدل استخراج شد. تعدادی از این قوانین در شکل ۵ نمایش داده شده‌اند. در این قوانین، خروجی صفر (۰) برای هر قانون به منزله سالم بودن کودک و نمایش عدد یک (۱) در خروجی نشان دهنده مبتلا بودن کودک به اوتیسم می‌باشد.

[Downloaded from jhbmi.ir on 2026-11-09]

1. if A4_Score > 0.500 and A10_Score > 0.500 and A7_Score > 0.500 and A3_Score ≤ 0.500 and jaundice] = yes then NO (1 / 1)
2. if A4_Score > 0.500 and A10_Score > 0.500 and A7_Score ≤ 0.500 and A3_Score > 0.500 and A1_Score > 0.500 and A5_Score > 0.500 then YES (0 / 19)
3. if A4_Score > 0.500 and A10_Score > 0.500 and A7_Score ≤ 0.500 and A3_Score > 0.500 and A1_Score > 0.500 and A5_Score ≤ 0.500 and jaundice = no then NO (2 / 0)
4. if A4_Score > 0.500 and A10_Score > 0.500 and A7_Score ≤ 0.500 and A3_Score > 0.500 and A1_Score > 0.500 and A5_Score ≤ 0.500 and jaundice = yes then YES (0 / 3)
5. if A4_Score > 0.500 and A10_Score > 0.500 and A7_Score ≤ 0.500 and A3_Score > 0.500 and A1_Score ≤ 0.500 and A9_Score ≤ 0.500 then NO (5 / 1)
6. if A4_Score > 0.500 and A10_Score ≤ 0.500 and A5_Score > 0.500 and A1_Score > 0.500 and A3_Score > 0.500 and A7_Score > 0.500 then YES (0 / 8)
7. if A4_Score > 0.500 and A10_Score ≤ 0.500 and A5_Score > 0.500 and A1_Score > 0.500 and A3_Score > 0.500 and A7_Score ≤ 0.500 then NO (2 / 1)
8. if A4_Score > 0.500 and A10_Score ≤ 0.500 and A5_Score > 0.500 and A1_Score ≤ 0.500 then NO (4 / 0)
9. if A4_Score ≤ 0.500 and A9_Score > 0.500 and A8_Score ≤ 0.500 and A1_Score > 0.500 and A6_Score > 0.500 and A2_Score > 0.500 then YES (0 / 3)
10. if A4_Score ≤ 0.500 and A9_Score > 0.500 and A8_Score ≤ 0.500 and A1_Score > 0.500 and A6_Score ≤ 0.500 then NO (3 / 0)
11. if A4_Score ≤ 0.500 and A9_Score ≤ 0.500 and A8_Score > 0.500 and A1_Score > 0.500 and A3_Score > 0.500 and A2_Score > 0.500 then YES (0 / 2)
12. if A4_Score ≤ 0.500 and A9_Score ≤ 0.500 and A8_Score > 0.500 and A1_Score > 0.500 and A3_Score ≤ 0.500 then NO (7 / 0)

شکل ۵: قوانین استخراج شده از درخت تصمیم

طراحی و پیاده‌سازی پایگاه دانش مبتنی بر آنتولوژی

آنتولوژی یک ویژگی صریح و واضح از مفهوم‌سازی می‌باشد که ارتباطی بین یادگیری ماشین و هوش مصنوعی برقرار می‌کند [۳۷]. یکی از خصوصیات بارز و اصلی آنتولوژی اشتراک‌گذاری دانش و استفاده مجدد از دانش می‌باشد که در سیستم‌هایی با داده‌های زیاد، استفاده از آنتولوژی راه‌حل مناسبی می‌باشد [۳۸]. کلاس‌ها، صفات، خصوصیات و نمونه‌های موجود در هر کلاس از جمله مواردی هستند که باید در آنتولوژی تعریف شوند. در این تحقیق از Protégé 5 برای ساخت آنتولوژی استفاده شده است [۳۹]. یک کلاس مجموعه‌ای از چندین نمونه با خصوصیات مشترک می‌باشد که نمونه‌های موجود در کلاس در واقع یکسری مفاهیم را نمایش می‌دهند. Tbox، در واقع شامل کلیه بدیهیات (axioms) موجود در آنتولوژی می‌باشد که در حین لود شدن آنتولوژی مورد استفاده قرار می‌گیرند. Abox، شامل همه نمونه‌های موجود در آنتولوژی و ارتباطات بین آن‌ها می‌باشد. کلاس‌های موجود در آنتولوژی مربوط به این تحقیق شامل HasOtism, NoOtism,Patients و Feature می‌باشند. کلاس Patients شامل همه کودکانی می‌باشد که ممکن است سالم باشند و یا بیمار. کلاس features در واقع کلیه متغیرها و کل ۲۱ ویژگی را شامل می‌شود که برای هر کودک متفاوت است. بعد از تفسیر قوانین، کودکانی که سالم تشخیص داده شدند در کلاس NoOtism و آن‌هایی که بیمار تشخیص داد در کلاس HasOtism قرار خواهند گرفت.

خصوصیات در واقع نمونه‌ها را از کلاس‌های متفاوت را به هم ربط می‌دهد و یا این که به نمونه‌ها، مقادیری را انتساب می‌دهد. به‌عنوان‌مثال، به ازای هر کدام از ۱۰ سؤالی که در رابطه با هر کودک مطرح می‌شود، پاسخ مربوطه به کمک یک نوع ارتباط تعریف شده به کودک مورد نظر اختصاص داده می‌شود. کلیه خصوصیات تعریف شده در این آنتولوژی در جدول ۳، نشان داده شده است. این خصوصیات بر اساس نیاز مدل تعریف شده‌اند که در واقع رابطه مستقیمی با متغیرهای انتخابی در درخت تصمیم دارد. متغیرهای انتخابی در واقع همان‌طور که اشاره شد بر اساس وزن و اولویت استفاده انتخاب شده‌اند.

[Downloaded from jhbmi.ir on 2026-11-09]

جدول ۳: خصوصیات تعریف شده در آنتولوژی

خصوصیات تعریف شده در آنتولوژی	توضیحات
HasJaundice	نشان می‌دهد کودک زردی داشته یا خیر. (۱/۰)
IsScore1 IsScore10 تا	مقدار عددی مربوط به پاسخ سؤال اول تا سؤال دهم (جدول ۱)
Otism	بعد از تفسیر قوانین، اگر کودک دارای اوتیسم بود ۱ وگرنه ۰

روش‌های متعددی برای ارزیابی و تأیید درستی آنتولوژی وجود دارد که یکی از پرکاربردترین آن‌ها ارزیابی به کمک افراد خبره می‌باشد [۴۰]. در این تحقیق از دو فرد خبره برای ارزیابی آنتولوژی استفاده شد. یکی روانشناس و متخصص بیماری اوتیسم و دیگری خبره در مهندسی آنتولوژی. فرد خبره در زمینه اوتیسم تأکید بر درستی مفاهیم پزشکی به کار برده شده داشت و فرد متخصص در زمینه آنتولوژی، بردرستی ساختار Abox و Tbox و همچنین نحوه درست برقراری ارتباطات تأکید لازم را داشت. در نتیجه تأییدات هر دو فرد خبره می‌توان به درستی مفهومی، ساختاری و معنایی آنتولوژی اشاره کرد.

تولید و تفسیر قوانین در آنتولوژی

همان‌طور که اشاره شد فرمت قوانین در آنتولوژی بر مبنای OWL (Web Ontology Language) و به شکل SWRL می‌باشند [۴۱]. مفسر قانون استفاده شده در Protege5 به نام Drool می‌باشد که قوانین را تفسیر کرده و دانش و حقایق جدید به پایگاه دانش اضافه می‌کند. در این تحقیق، قوانین استخراج شده از مدل درخت تصمیم به شکل SWRL توسط مبدلی، تبدیل و در آنتولوژی ذخیره می‌شوند. این مبدل با متلب پیاده‌سازی شده است که خط به خط قوانین درخت تصمیم را خوانده و به شکل SWRL تبدیل می‌کند. به عنوان نمونه، ۵ قانون از ۲۵ قانون متناظر با درخت تصمیم و آنتولوژی در شکل ۶ نمایش داده شده است.

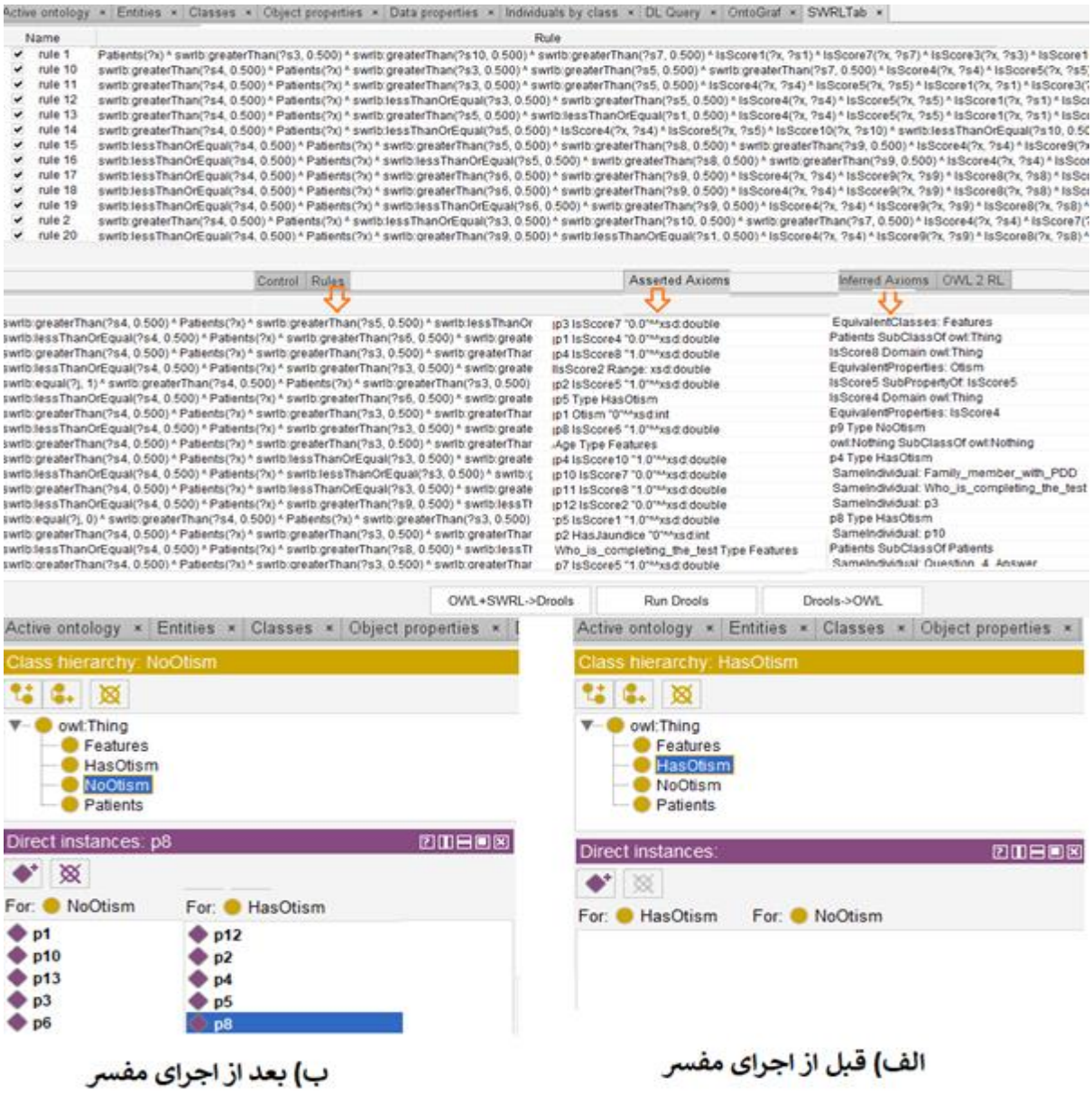
قوانین درخت تصمیم	قوانین SWRL در آنتولوژی
if A4_Score > 0.500 and A10_Score > 0.500 and A7_Score > 0.500 and A3_Score > 0.500 then YES (0 / 54)	Patients(?x) ^ swrlb:greaterThan(?s3, 0.500) ^ swrlb:greaterThan(?s10, 0.500) ^ swrlb:greaterThan(?s7, 0.500) ^ IsScore1(?x, ?s1) ^ IsScore7(?x, ?s7) ^ IsScore3(?x, ?s3) ^ IsScore10(?x, ?s10) ^ swrlb:greaterThan(?s1, 0.500) -> HasOtism(?x)
if A4_Score > 0.500 and A10_Score > 0.500 and A7_Score > 0.500 and A3_Score ≤ 0.500 and jundice = yes then NO (1 / 1)	swrlb:greaterThan(?s4, 0.500) ^ Patients(?x) ^ swrlb:lessThanOrEqual(?s3, 0.500) ^ swrlb:greaterThan(?s10, 0.500) ^ swrlb:greaterThan(?s7, 0.500) ^ IsScore4(?x, ?s4) ^ IsScore7(?x, ?s7) ^ IsScore3(?x, ?s3) ^ IsScore10(?x, ?s10) ^ HasJaundice(?x, ?j) ^ swrlb:equal(?j, 0) -> NoOtism(?x)
if A4_Score ≤ 0.500 and A9_Score ≤ 0.500 and A8_Score ≤ 0.500 then NO (49 / 0)	swrlb:lessThanOrEqual(?s4, 0.500) ^ Patients(?x) ^ IsScore4(?x, ?s4) ^ IsScore9(?x, ?s9) ^ IsScore8(?x, ?s8) ^ swrlb:lessThanOrEqual(?s8, 0.500) ^ swrlb:lessThanOrEqual(?s9, 0.500) -> NoOtism(?x)
if A4_Score ≤ 0.500 and A9_Score > 0.500 and A8_Score ≤ 0.500 and A1_Score > 0.500 and A6_Score > 0.500 and A2_Score > 0.500 then YES (0 / 3)	swrlb:lessThanOrEqual(?s4, 0.500) ^ Patients(?x) ^ swrlb:greaterThan(?s6, 0.500) ^ swrlb:greaterThan(?s9, 0.500) ^ IsScore4(?x, ?s4) ^ IsScore9(?x, ?s9) ^ IsScore8(?x, ?s8) ^ IsScore1(?x, ?s1) ^ IsScore6(?x, ?s6) ^ IsScore2(?x, ?s2) ^ swrlb:lessThanOrEqual(?s8, 0.500) ^ swrlb:greaterThan(?s2, 0.500) ^ swrlb:greaterThan(?s1, 0.500) -> HasOtism(?x)

شکل ٦: تبدیل قوانین از درخت تصمیم به SWRL

[Downloaded from jhbmi.ir on 2026-11-09]

نتایج

Drool [۴۲] به عنوان یک مفسر در پایگاه دانش شروع به تفسیر قوانین کرده و یکسری بدیهیات و دانش جدید را به پایگاه دانش اضافه می‌کند. خروجی این تفسیر نشان دهنده افراد بیمار و سالم است که شکل ۷ گویای بخشی از این تفسیر می‌باشد.



شکل ۷: قسمتی از نتیجه خروجی آنتولوژی

همان‌طور که اشاره شد دو کلاس NoOtism و HasOtism در برگیرنده کودکان سالم و بیمار بعد از تفسیر آنتولوژی به ترتیب می‌باشند. در شکل ۷ قسمت "الف" هنوز Drool به عنوان مفسر قانون اجرا نشده است و دو کلاس NoOtism و HasOtism خالی از کودک هستند درحالی که در قسمت "ب" مفسر اجرا شده و هر دو کلاس مورد نظر شامل یکسری کودک می‌باشند. بر اساس قوانین اجرا شده منطبق بر علائم کودک، کلاس NoOtism شامل کودکان با شناسه P1,P10,P13,P3 و P6 و کلاس HasOtism شامل شناسه های P12, P2,P4,P5 و P8 می‌باشد. حال باید دید که آیا این خروجی معادل با خروجی درخت تصمیم می‌باشد و مدل طراحی شده را تأیید می‌کند؟

[Downloaded from jhbmi.ir on 2026-11-09]

برای تأیید نتایج به دست آمده در آنتولوژی و همچنین درستی مدل درخت تصمیم، مقایسه خروجی‌ها برای تعدادی کودک به عنوان نمونه در جدول ۴ نشان داده شده است. این جدول شامل ۳ ستون می‌باشد که ستون اول از سمت راست تشخیص اوتیسم توسط آنتولوژی، ستون دوم تشخیص اوتیسم به وسیله درخت تصمیم (عدد ۰ نشان‌دهنده سالم بودن و عدد ۱نشان‌دهنده بیمار بودن) و ستون سوم شناسه کودک را نمایش می‌دهد.

جدول ٤: مقایسه خروجی آنتولوژی و درخت تصمیم

تشخیص اوتیسم در آنتولوژی	تشخیص اوتیسم در درخت تصمیم	شناسه کودک
اوتیسم ندارد (موجود در کلاس NoOtism)	۰	P1
اوتیسم دارد (موجود در کلاس hasOtism)	۱	P2
اوتیسم ندارد (موجود در کلاس NoOtism)	۰	P3
اوتیسم دارد (موجود در کلاس hasOtism)	۱	P4
اوتیسم دارد (موجود در کلاس hasOtism)	۱	P5
اوتیسم ندارد (موجود در کلاس NoOtism)	۰	P6
اوتیسم دارد (موجود در کلاس hasOtism)	۱	P8
اوتیسم ندارد (موجود در کلاس NoOtism)	۰	P10
اوتیسم دارد (موجود در کلاس hasOtism)	۱	P12
اوتیسم ندارد (موجود در کلاس NoOtism)	۰	P13

به عنوان نمونه برای کودکی با شناسه P1، تشخیص اوتیسم توسط آنتولوژی رد شده و کودک را در کلاس NoOtism قرار داده است. همچنین عدد ۰ در ستون دوم، نشان دهنده رد بیماری اوتیسم برای این کودک توسط درخت تصمیم می‌باشد. همان‌طور که دیده می‌شود، نتایج کاملاً متناظر با هم بوده و خروجی آنتولوژی خود تأییدی بر درستی مدل درخت تصمیم و بالعکس می‌باشد.

بحث و نتیجه‌گیری

با توجه به اهمیت بیماری اوتیسم در کودکان باید راه‌های تشخیص این بیماری را وسعت داد تا بتوان با تشخیص زود هنگام، پیشرفت آن را کندتر کرد. در عصر حاضر، داده‌کاوی توانسته با الگوریتم‌های تشخیص بیماری کمک شایانی را در علم پزشکی داشته باشد. داده‌کاوی در کنار پایگاه داده‌های جدید، علیرغم حجم عظیمی از داده‌ها و رشد سریع آن‌ها، بستر لازم برای مدیریت و دسترسی مؤثر به این حجم از داده را آسان کرده است. اگرچه محدودیت‌هایی هم وجود دارد که داده‌کاوی قادر به انجام آن‌ها نیست به عنوان نمونه، همه الگوهای یافت شده از این طریق درست و دقیق نیستند و برای استفاده از یک الگو و اطمینان به آن، الگوی موردنظر باید منطقی و در عمل قابل اجرا باشد؛ بنابراین داده‌کاوی نیازمند مداخلات انسانی برای به‌کارگیری دانش استخراج شده است. برای مثال داده‌کاوی می‌تواند به ایجاد تشخیص یا تجویز درمان کمک کند، ولی هنوز نتوانسته است جایگزین مهارت‌های تجربی پزشکان گردد. درخت تصمیم به عنوان یکی از الگوریتم‌های داده‌کاوی دارای خروجی به صورت قوانین می‌باشد که می‌تواند در کنار آنتولوژی به عنوان یک ساختار معنایی به خوبی ایفای نقش کند. تعداد رکودهای به کار برده شده در این تحقیق ۲۹۲ رکورد بودند که تعداد ۲۳۴ رکورد به عنوان ۸۰٪ رکوردها در مدل درخت تصمیم استفاده شدند. آنتولوژی ایجاد شده در این تحقیق تأیید می‌کند که آیا فرد موردنظر دارای بیماری اوتیسم است و یا خیر. همان‌طور که پیش‌تر اشاره شد تعدادی از محققین با استفاده از آنتولوژی و الگوریتم‌های داده‌کاوی به صورت مجزا بر روی بیماری اوتیسم کار کرده‌اند، ولی بارزترین تفاوت که در واقع نوآوری تحقیق هم به حساب می‌آید ارائه ساختاری ترکیبی از وب معنایی و داده‌کاوی می‌باشد تا در کنار قدرت داده‌کاوی جهت تشخیص بیماری بتوان از قابلیت‌های آنتولوژی جهت اشتراک‌گذاری و قابلیت استفاده مجدد سیستم استفاده کرد. بیشتر سیستم‌های موجود از روش داده‌کاوی استفاده کرده‌اند و به ندرت مبتنی بر آنتولوژی هستند. یکی از سیستم‌های مبتنی بر آنتولوژی [۳۴] قابلیت تشخیص بیماری را ندارد و فقط جنبه تحقیقاتی داشته و قابل استفاده برای تکنسین‌ها و افراد درمان‌گرا و پزشکان نیست. سیستم دیگری که قابلیت تشخیص بیماری را دارد

[Downloaded from jhbmi.ir on 2026-11-09]

[۳۳] از آنتولوژی Gen استفاده کرده که شامل کلیه اطلاعات ژنتیکی بدن انسان است و مختص بیماری اوتیسم نمی‌باشد. سیستم ارائه شده در مطالعه حاضر نشان داد که آنتولوژی ایجاد شده برای بیماران اوتیسم به درستی کار می‌کند و همچنین صحت تشخیص بیماری را در درخت تصمیم با استفاده از تفسیر قوانین، تأیید کرد.

به عنوان کارهای آتی در نظر است تا با استفاده از جمع‌آوری داده‌هایی از مراکز درمانی ایران این سیستم را ارزیابی کرده و برای بیماری‌های دیگر توسعه داده شود. در واقع در هدف این است تا این سیستم با کسب مجوزهای قانونی در چندین مرکز اوتیسم به کار گرفته شود تا بتوان کمک شایانی در تشخیص این بیماری و کاهش رشد این بیماری در کودکان بیمار داشته باشیم. از جمله کاستی‌های این تحقیق می‌توان به عدم ارائه روش‌های درمانی برای بیماران اشاره کرد که در نظر است در آینده روی این موضوع کار کرده و آنتولوژی سیستم را برای انجام این فاز توسعه داده شود.

تعارض منافع

نویسندگان هیچ‌گونه تضاد منافعی را گزارش نمی‌کنند.

References

1. Mokhtari B, Karimzadeh F. A review on the Authism with the most approach on the critical biomarkers. Razi Journal of Medical Sciences 2018;24(165):35-46. [In Persian]

2. Schriml LM, Arze C, Nadendla S, Chang YW, Mazaitis M, Felix V, et al. Disease Ontology: a backbone for disease semantic integration. Nucleic Acids Research 2012;40(D1):D940-6.

3. Chen L, Lu D, Zhu M, Muzammal M, Samuel OW, Huang G, et al. OMDP: An ontology-based model for diagnosis and treatment of diabetes patients in remote healthcare systems. International Journal of Distributed Sensor Networks 2019;15(5). doi:10.1177/1550147719847112

4. Patel HH, Prajapati P. Study and analysis of decision tree based classification algorithms. International Journal of Computer Sciences and Engineering 2018;6(10):74-8.

5. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13-17; New York, NY, United States: SIGMOD, SIGKDD; 2016. p. 785-94. https://doi.org/10.1145/2939672.2939785

6. Cortez P, Silva AM. Using data mining to predict secondary school student performance; 2008.

7. Ahmed AM, Rizaner A, Ulusoy AH. Using data mining to predict instructor performance. Procedia Computer Science 2016;102:137-42.

8. Kapur B, Ahluwalia N, Sathyaraj R. Comparative Study on Marks Prediction using Data Mining and Classification Algorithms. International Journal of Advanced Research in Computer Science 2017;8(3).

9. Jules GD, Saadat M, Saeidlou S. Ontological extension of PROSA for manufacturing network formation. InIndustrial Applications of Holonic and Multi-Agent Systems 2013 Berlin, Heidelberg: Springer; 2013. p. 140-51.

10. Mehta T, Kshirsagar T, Merchant A, Nair S. Graduate prediction using ontology approach. International Journal of Computer Science and Information Technologies 2015;6(5):4782-4.

11. Ravishankar TN, Shriram R. Ontology based clustering algorithm for information retrieval. Fourth International Conference on Computing, Communications and Networking Technologies; 2013 Jul 4; Tiruchengode, India: IEEE; 2013. p. 1-4.

12. Zhong Z, Saeidlou S, Saadat M, Abukar A. Ontology and rule-based reasoning for intelligent predictive manufacturing. Advances in Manufacturing Technology XXXII; 2018 Aug 29; IOS Press; 2018. p. 355-60.

13. Chang HW, Tai YC, Chen HW, Hsu JY, Kuo CP. iTaxi: Context-Aware Taxi Demand Hotspots Prediction Using Ontology and Data Mining Approaches. International Journal of Business Intelligence and Data Mining 2009; 5(1):1-8.

14. Kulmanov M, Hoehndorf R. DeepPheno: Predicting single gene loss-of-function phenotypes using an ontology-aware hierarchical classifier. PLoS Computational Biology 2020;16(11):e1008453. https://doi.org/10.1371/journal.pcbi.1008453

15. Cheng L, Li J, Ju P, Peng J, Wang Y. SemFunSim: a new method for measuring disease similarity by integrating semantic and gene functional association. PloS one 2014;9(6):e99415.

16. Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. IEEE/ACS International Conference on Computer Systems and Applications; 2008 Mar 31; Doha, Qatar: IEEE; 2008 p. 108-15. doi: 10.1109/AICCSA.2008.4493524

17. Shuja M, Mittal S, Zaman M. Effective prediction of type ii diabetes mellitus using data mining classifiers and SMOTE. In Advances in computing and intelligent systems. Singapore: Springer; 2020. p. 195-211.

18. Islam MM, Ferdousi R, Rahman S, Bushra HY. Likelihood prediction of diabetes at early stage using data mining techniques. InComputer Vision and Machine Intelligence in Medical Image Analysis 2020 Singapore: Springer; 2020. p. 113-25.

19. Horrocks I, Patel-Schneider PF, Boley H, Tabet S, Grosof B, Dean M. SWRL: A semantic web rule language combining OWL and RuleML. W3C Member Submission 2004;21(79):1-31.

[Downloaded from jhbmi.ir on 2026-11-09]

20. Al-Diabat M. Fuzzy data mining for autism classification of children. International Journal of Advanced Computer Science and Applications 2018;9(7):11-7. doi:10.14569/IJACSA.2018.090702
21. Thabtah F, Peebles D. A new machine learning model based on induction of rules for autism detection. Health Informatics Journal 2020;26(1):264-86.
22. Achenie LE, Scarpa A, Factor RS, Wang T, Robins DL, McCrickard DS. A machine learning strategy for autism screening in toddlers. J Dev Behav Pediatr. 2019; 40(5): 369–76. doi: 10.1097/DBP.0000000000000668
23. Kong Y, Gao J, Xu Y, Pan Y, Wang J, Liu J. Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier. Neurocomputing 2019;324:63-8.
24. Kosmicki JA, Sochat V, Duda M, Wall DP. Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. Transl Psychiatry 2015;5(2):e514. doi: 10.1038/tp.2015.7.
25. Alwidian J, Elhassan A, Ghnemat R. Predicting autism spectrum disorder using machine learning technique. International Journal of Recent Technology and Engineering 2020;8:4139-43.
26. Shuvo SB, Ghosh J, Oyshi AS. A data mining based approach to predict autism spectrum disorder considering behavioral attributes. 10th International Conference on Computing, Communication and Networking Technologies; 2019 Jul 6; Dhaka, Bangladesh: IEEE; 2019. p. 1-5.
27. Khan S, Alshara M. Fuzzy data mining utilization to classify kids with autism. International Journal of Computer Science and Network Security 2019;19(2):147-54.
2**8.** Kazeminejad A, Sotero RC. Topological properties of resting-state fMRI functional networks improve machine learning-based autism classification. Frontiers in Neuroscience 2019;12:1018.
29. Subbaraju V, Suresh MB, Sundaram S, Narasimhan S. Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging: A spatial filtering approach. Med Image Anal 2017;35: 75-389. doi: 10.1016/j.media.2016.08.003.
30. Chowdhury K, Iraj MA. Predicting autism spectrum disorder using machine learning classifiers. International Conference on Recent Trends on Electronics, Information, Communication & Technology; 2020 Nov 12; IEEE. p. 324-27.
31. Musa RA, Manaa ME, Abdul-Majeed G. Predicting Autism Spectrum Disorder (ASD) for Toddlers and Children Using Data Mining Techniques. Journal of Physics: Conference Series 2021; 1804(1): 1-8.
32. Galán-Mena J, Ávila G, Pauta-Pintado J, Lima-Juma D, Robles-Bykbaev V, Quisi-Peralta D. An intelligent system based on ontologies and ICT tools to support the diagnosis and intervention of children with autism. Biennial Congress of Argentina; 2016 Jun 15; Buenos Aires, Argentina: IEEE; 2016. p. 1-5. doi: 10.1109/ARGENCON.2016.7585361
33. Jemima DD, Selvarani AG, Lovenia JD. A study on Diagnosis of Autism Spectrum Disorder for Children. 3rd International Conference on Signal Processing and Communication; 2021 May 13; IEEE; 2021. p. 547-51.
34. Tu SW, Tennakoon L, O'Connor M, Shankar R, Das A. Using an integrated ontology and information model for querying and reasoning about phenotypes: the case of autism. AMIA Annu Symp Proc 2008; 2008: 727–31.
35. Chisholm A. Exploring Data with Rapidminer. Packt Publishing; 2013.
36. Greenwood PE, Nikulin MS. A guide to chi-squared testing. John Wiley & Sons Inc: United States; 1996.
37. Gruber, T. (2009). Ontology. In: LIU L, ÖZSU, MT. (eds) Encyclopedia of Database Systems. Boston, MA: Springer; 2009. https://doi.org/10.1007/978-0-387-39940-9_1318
38. Chavan PL, Karyakarte MS. Ontology Based System for Prediction of Diseases. International Journal of Scientific Research in Science, Engineering and Technology 2020;7(3): 275-85. doi: https://doi.org/10.32628/IJSRSET207365
39. Musen MA. The protégé project: a look back and a look forward. AI Matters 2015;1(4):4-12.
40. Brank J, Grobelnik M, Mladenic D. A survey of ontology evaluation techniques. In Proceedings of the Conference on Data Mining and Data Warehouses; 2005 Oct 17; Slovenia: Citeseer Ljubljana. 2005. p. 166-70.
41. Sicilia MÁ, Lytras MD, Sánchez-Alonso S, García-Barriocanal E, Zapata-Ros M. Modeling instructional-design theories with ontologies: Using methods to check, generate and search learning designs. Computers in Human Behavior 2011;27(4):1389-98. doi:10.1016/j.chb.2010.07.040
42. O'Connor MJ, Das AK. A Pair of OWL 2 RL Reasoners; 2012. Available from: http://webont.org/owled/2012/papers/paper_31.pdf

[Downloaded from jhbmi.ir on 2026-11-09]

Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2022; 9(1): 12-24

Development of a Combined System Based on Data Mining and Semantic Web for the Diagnosis of Autism

Shishehchi Saman[1], Banihahsem Seyed Yashar[1]*

• **Received:** 25 Feb 2022 • **Accepted:** 27 Apr 2022

Introduction: Autism is a nervous system disorder, and since there is no direct diagnosis for it, data mining can help diagnose the disease. Ontology as a backbone of the semantic web, a knowledge database with shareability and reusability, can be a confirmation of the correctness of disease diagnosis systems. This study aimed to provide a system for diagnosing autistic children with a combination of semantic web and data mining.

Method: Data is taken from the UCI database. There were a total of 292 data records available of which 80% (234 records) were used for modeling through the decision tree. Knowledge about patients and autism disease was presented via ontology using the Protégé 5 software. The ontology has four classes and 12 properties to communicate between the individuals in the classes. The rules extracted from the decision tree were transformed into a comprehensible form (SWRL) for interpretation in the ontology using a converter.

Results: Whether the child is healthy or not can be determined by the rules obtained in the decision tree. In addition, the output of the ontology using the interpretation of 25 rules confirmed the diagnosis of an Autistic child using the decision tree. The evaluation of the ontology also confirmed its correctness.

Conclusion: According to the similarity between the result of the ontology and the decision tree regarding the diagnosis of the disease, the accuracy of the proposed method can be emphasized.

Keywords: Autism, Data Mining, Decision Tree, Ontology, Rule-based System

• **Citation:** Shishehchi S, Banihahsem SY. Development of a Combined System Based on Data Mining and Semantic Web for the Diagnosis of Autism. Journal of Health and Biomedical Informatics 2022; 9(1): 12-24. [In Persian]

1. PhD in Information Technology, Assistant Professor, Buein Zahra Technical University, Buein Zahra, Qazvin, Iran
***Correspondence:** Saman Shishehchi
Address: Buein Zahra Technical University, Buein Zahra, Qazvin, Iran
• **Tel:** 09120741273 • **Email:** shishechi@bzte.ac.ir

[Downloaded from jhbmi.ir on 2026-11-09]