دوفصلنامه محاسبات و سامانه‌های توزیع شده
www.jdcs.ir

طراحی یک سیستم‌خبره فازی نوع‌دوم جهت بهبود نقطه‌ضعف کاوش قواعد انجمنی فازی نوع اول به‌منظور تشخیص بیماری دیابت

میثم روشن فکر[1]، گلنوش عبائی[1*]، حسین ابراهیم پور کومله[2]

۱- دانشکده مهندسی برق، کامپیوتر و مهندسی پزشکی، دانشگاه شهاب دانش، قم، ایران
۲- دانشکده مهندسی برق وکامپیوتر، دانشگاه کاشان، کاشان، ایران

چکیده

دیابت، شایع‌ترین بیماری غدد در جهان است که سالانه باعث مرگ ۴میلیون نفر در جهان می‌شود. برای تشخیص دیابت بایدتمام ویژگی‌هایی که منجر به وقوع آن می‌شوند را شناخت و درصورتیکه این ویژگی‌ها هرکدام از محدوده مقداری معین (درعلم پزشکی) تجاوز کنند مشخص می‌شود که فرد مبتلا به دیابت است. در این پژوهش با استفاده از ترکیب علوم داده‌کاوی، سیستم‌خبره و فازی نوع‌دوم، ابزاری برای کمک به پزشک در تشخیص بیماری دیابت ارائه شده است. با یک دید کلی، دو مولفه اصلی در سیستم تشخیص دیابت بهبود داده شده است: مولفه اول، دقت، کارایی و قابلیت اعتماد بالا در تشخیص بیماری می‌باشد که بدین منظور از رویکرد فازی نوع دوم برای تنظیم و فازی‌سازی ورودی‌ها استفاده شده است و مولفه دوم، کاهش حجم قواعد موجود در پایگاه‌دانش و قابلیت تفسیر ساده آن است که بدین منظور از کاوش قواعد انجمنی که یکی از تکنیک‌های داده‌کاوی می‌باشد استفاده شده است. به‌طور دقیق‌تر، در روش پیشنهادی، بعضی از معایب سیستم‌های ارائه شده قبلی بهبود داده شده که عبارتند از: بهبود تاثیرمنفی عدم‌قطعیت در مرزهای توابع عضویت فازی نوع اول، درنظرگرفتن داده‌های ورودی نویزی و اغتشاشات، جایگزین‌کردن یک سیستم‌خبره هوشمند با تصمیم‌گیری واحد در تشخیص بیماری به جای سیستم‌هایی که توسط کارشناسان مختلف به‌دلیل تفاوت در نژاد و وضعیت بدنی افراد مختلف ساکن در قاره‌های مختلف، نتایج متفاوتی را ارائه می‌دهند، بهبود دادن حجم پایگاه‌دانش با انتخاب قواعد موثرتر در تولید خروجی، افزایش سرعت سیستم با کاهش دادن زمان محاسباتی و نیز بهبود دادن محدوده نقاط پارتیشن‌بندی در محل برخورد مرزهای دو تابع‌عضویت. برای تست عملکرد مطلوب روش پیشنهادی، از مجموعه داده بیماران دیابتی سایت UCI که شامل ۷۶۸ بیمار است استفاده شده است و با دست یافتن به میزان دقت ۹۴٫۸۸ درصد، اثبات کردیم که توانسته‌ایم به اهداف خود برای بهبود دقت در سیستم تشخیص دیابت برسیم.

واژگان‌کلیدی: سیستم‌خبره، فازی نوع دوم، کاوش قواعد انجمنی، تشخیص بیماری، دیاب

* نویسنده مسئول: دکتر گلنوش عبائی، مدیر پژوهش دانشگاه شهاب دانش، قم، ایران
ایمیل: abaee@shahabdanesh.ac.ir

تاریخچه مقاله:

تاریخ ارسال:۹۷/۳/۱

تاریخ اصلاحات:۹۷/۵/۱

تاریخ پذیرش:۹۷/۵/۹

تاریخ انتشار:**۹۷/۵/۱۵**

Keywords:

Expert System
Type-2 Fuzzy
Association Rule Mining
Diagnosis of Disease
Diabetes

Designing a Type-2 Fuzzy Expert System to Improve Fuzzy Association Rule Mining Weakness for Diagnosis of Diabetes Disease

Meisam Roshanfekr[1], Dr.Golnoush Abaei[1], Dr.Hossein Ebrahimpour Komleh[2]

1- Faculty of Electrical, Computer and Biomedical Engineering, Shahab Danesh University, Qom, Iran

2- Faculty of Electrical and Computer Engineering, Kashan University, Kashan, Iran

Abstract

World Health Organization has recognized diabetes as the more prevalent endocrine disease in the world, which causes deaths of 4 million people per year. In order to diagnose diabetes, all symptoms and features leading to diabetes should be detected and if each feature exceeds a certain value, the case is known as a diabetic; otherwise, the case is healthy. In this research, combination of data mining, expert system, and type-2 fuzzy are applied to provide a tool as an assistant to physicians in diagnosing diabetes or even to be substituted with the physicians. In general, two main components in diabetes diagnosing system are developed: the first one is accuracy, and high performance and reliability in diagnosing the disease by applying type-2 fuzzy approach, and the second component is decreasing the value of rules available in knowledge base and the capability of simple interpretation of it, by employing data mining techniques named association rule. The proposed method, improves flaws in previously provided systems via combination of association rules mining and type-2 fuzzy sciences in following ways: improving negative efficacy of uncertainty in boundaries of membership functions of type-1 fuzzy; Taking noisy input data into consideration; substitution of single decision making method of diagnosing disease with an intelligent expert system instead of systems which applied by various experts that provided different results due to the differences existing in peoples residing in various continents in terms of their pedigrees and body mass indexes; decreasing the volume of knowledge base by selecting more effective rules in generation of output; increasing the system's runtime speed by decreasing the computational time; and also improvement of area of partitioning points in the two membership functions' intersection of boundaries. The proposed model was evaluated by diabetic patients dataset obtained from UCI repository, which contains 768 samples. The results show 94.88% of accuracy, which proves the capability of the proposed model in better diagnosing the diabetics.

روش ارجاع به مقاله: م.روشن فکر، گ.عبائی، ح.ابراهیم پور کومله، طراحی یک سیستم‌خبره فازی نوع‌دوم جهت بهبود نقطه‌ضعف کاوش قواعد انجمنی فازی نوع اول به‌منظور تشخیص بیماری دیابت، دوفصلنامه محاسبات و سامانه های توزیع شده، سال اول، شماره اول، ص ۱۱۸-۱۳۴، سال انتشار ۱۳۹۷.

۱-مقدمه

دیابت شایع‌ترین بیماری غدد در جهان توسط سازمان بهداشت جهانی معرفی شده که سالانه بیش از ۴ میلیون نفر را به کام مرگ فرو می‌برد. به دلیل عدم آگاهی بعضی از افراد از دیابت و روش کنترل آن، در هر ۱۰ ثانیه یک نفر در جهان جان خود و در هر ۳۰ ثانیه یک نفر در جهان، پای خود را از دست می‌دهد. میزان ابتلا به بیماری دیابت در جهان با سرعت درحال افزایش است و این میزان در کشورهای درحال توسعه بیش از سایر کشورها است، زیرا ساکنان این کشورها به رژیم غذایی مملو از قند و نشاسته روی آورده‌اند که سال‌هاست کشورهای غنی را دچار مشکل کرده است. طبق گزارش انجمن دیابت آمریکا[۲] [۱]، فناوری اندازه‌گیری چربی در بدن نشان می‌دهد که آسیایی‌ها با شاخص توده بدنی مساوی در مقایسه با اروپایی‌ها، چربی انباشته بیشتری در شکم خود دارند و این مسئله، احتمال ابتلا به دیابت را در آن‌ها افزایش می‌دهد. برای تشخیص یک بیماری باید تمام علائم و ویژگی‌هایی که منجر به وقوع آن بیماری می‌شوند را شناخت و درصورتیکه این ویژگی‌ها هرکدام، از محدوده مقدار معینی تجاوز کنند فرد مبتلا به دیابت است.

هنگامیکه تعداد تست‌ها زیاد باشد، ممکن است تشخیص بیماری حتی برای یک پزشک متخصص نیز به سختی انجام شود. همین امر موجب شده است که در طول سال‌های اخیر، سیستم‌های خبره کامپیوتریِ تشخیص بیماری با قصد کمک به پزشک متخصص، توسعه داده شوند تا به‌نحوی بی‌نظمی را از داده‌های موجود خارج کنند. در این پژوهش، با استفاده از ترکیب علوم داده‌کاوی، فازی و سیستم‌خبره، ابزاری برای کمک به پزشک در تشخیص بیماری دیابت ارائه شده است. که در حالت کلی دو مولفه اصلی را در سیستم تشخیص بیماری بهبود داده است: مولفه اول، دقت، کارایی و قابلیت اعتماد بالا در تشخیص بیماری می‌باشد که بدین منظور از رویکرد فازی نوع دوم استفاده کرده‌ایم. مولفه دوم، قابلیت تفسیر ساده آن است؛ یعنی قواعد استخراج شده توسط این سیستم باید به حدی مفید، ساده و موثر درخروجی باشند که به راحتی توسط پزشک قابل تفسیر باشند که بدین منظور از کاوش قواعد انجمنی که یکی از تکنیک‌های داده‌کاوی می‌باشد استفاده کرده‌ایم.

به‌منظور تشخیص بیماری، باید ویژگی‌های بیماری را در طراحی سیستم، به‌عنوان ورودی تنظیم کرد. اگر آن‌ها با مجموعه کلاسیک و باینری (دو مقداری) تعریف شوند، جواب مطلوبی به دست نمی‌آید. بنابراین هر یک از ویژگی‌ها فازی شده و از سه متغیرکلامی (کم، متوسط، زیاد) و یا بیشتر برای آن‌ها استفاده می‌شود که درنتیجه یک سیستم خبره فازی برای تشخیص بیماری دیابت بوجود می‌آید. اما از معایب سیستم طراحی‌شده با فازی نوع اول این است که نتیجه به‌دست آمده از آن، در شرایط و یا مکان‌های مختلف، باهم فرق می‌کند و این به‌دلیل تفاوت نظر کارشناسان خبره و یا تفاوت در شرایط محیطی و انسانی و نیز عدم قطعیت در داده‌های ورودی می‌باشد. در این سیستم نیز به‌دلیل اینکه وضعیت بدن انسان‌ها در شرایط مختلف و درکشورهای مختلف، یکسان نیست (مثلا کسی که در آفریقا زندگی می‌کند با کسی که در روسیه هست شرایط بدنی و عکس‌العمل‌های بدنی متفاوتی دارد)، یک نویز و عدم قطعیت در این مجموعه فازی به‌وجود می‌آید که فازی نوع اول نمی‌تواند به‌طور دقیق، درجه عضویت را برای هریک از ویژگی‌ها تعیین کند. بنابراین در این پژوهش، یک سیستم خبره با استفاده از فازی نوع دوم طراحی شده است که در آن، مرزهای این مجموعه (درجات عضویت فازی نوع اول) مجددا فازی شده است تا مشکل عدم تعیین دقیق در مجموعه فازی نوع اول را برای ویژگی‌های بیماری حل شود و نویزهای موجود درسیستم که در خروجی تاثیر گذارند نیز در نظر گرفته شود. با این کار علاوه بر اینکه برای موارد عدم قطعیت و ابهام میشود جواب داشت ، اختلاف نظر کارشناسان خبره نیز پوشش داده خواهد شد و از نتایج متفاوت حاصل از اجرای یک سیستم‌خبره فازی یکسان، درکشورهای مختلف جلوگیری میشود. درنتیجه دقت سیستم نسبت به روش‌های پیشین تا حد قابل قبولی افزایش می‌یابد. ضمن اینکه ما با یک مجموعه بزرگ قواعد (پایگاه‌دانش بزرگ در FIS[۳]) روبرو هستیم. باید بتوانیم، سطح بالاتری از دانش را با توجه به پایگاه‌دانش مذکور به‌دست بیاوریم؛ درنتیجه برای سرعت بخشیدن به زمان اجرا و نیز ساده شدن فهم سیستم، با استفاده از کاوش قواعد انجمنی که یکی از تکنیک‌های داده‌کاوی است، بهترین و قوی‌ترین قواعد[۴] که تاثیر بیشتری را در ایجاد نتیجه خروجی دارند را استخراج می‌کنیم.

[۲] American Diabetes Association

[3] Fuzzy Inference Engine

[۴] Rule

برای تست عملکرد روش پیشنهادی، از یک مجموعه داده به‌نام PDD [۲] برگرفته از سایت UCI شامل اطلاعات گرفته شده از ۷۶۸ بیماردیابتی برای آموزش و تست سیستم استفاده شده است.

در بخش دوم، کارهای مرتبط انجام شده در گذشته با ذکر معایب تشریح می‌شوند و در بخش سوم، مبانی مورد نیاز در زمینه دیابت، سیستم‌خبره، فازی و فازی نوع دوم آورده شده است، دربخش چهارم براساس جمع‌بندی‌های بدست‌آمده از پژوهش‌های قبلی یک سیستم نمونه اولیه بر پایه فازی نوع اول ارائه شده و پس از نمایش نتیجه خروجی، معایب آن بیان می‌شود؛ سپس در بخش پنجم روش پیشنهادی با استفاده از رویکرد فازی نوع دوم و تکنیک داده‌کاوی کاوش قواعد انجمنی، تلاش در رفع عیب‌ها کرده و با بهبود دادن این معایب، یک سیستم جامع ارائه می‌شود. در بخش ششم، بعد از انجام آزمایشات، میزان دقت سیستم پیشنهادی نمایش داده شده و در ادامه مقایسه با کارهای قبلی انجام شده است و در آخر نیز نتیجه‌گیری ارائه شده است.

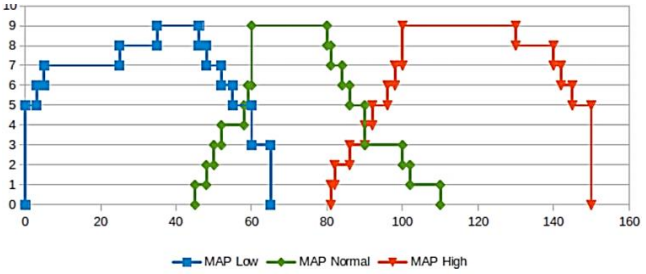
۲- کارهای مرتبط در گذشته

در سال ۲۰۱۱، کارگودا و همکارانش [۳] با استفاده از یک شبکه عصبی- ژنتیکی، راه‌حل جدیدی برای شناسایی بیماران دیابتی ارائه دادند که در آن برای آموزش شبکه و تست، از مجموعه داده PIDD[۵] استفاده شده‌بود. ۸ ویژگی به عنوان معیارهای اصلی تشخیص دیابت درنظرگرفته شده بود. روش پیشنهادی در این مقاله برمبنای شناسایی الگو بود، وظیفه الگوریتم ژنتیک نیز بهینه‌سازی مسائل آموزش و یادگیری بود که با کدکردن پارامترها درون ژن‌ها و تعریف یک تابع هدف مناسب، به مقصود آموزش شبکه و میانگین نرخ دقت ٪ ۸۴,۷ دست یافتند.

درسال ۲۰۱۵ توسط جین و راهجا [۴] پژوهشی با عنوان بهبود نرخ پیش‌بینی دیابت با استفاده از سیستم‌خبره فازی ارائه شد. در این پژوهش ادعا این بود که منطق فازی در تشخیص بیماری دیابت بسیار مفید است و دقت سیستم تشخیص را بالا می‌برد، زیرا روش‌های دیگر هرکدام سیستم‌های متفاوتی را ارائه می‌دهند که دقت پیش‌بینی آن‌ها دقیق نیست. در این روش، از میان تمام ویژگی‌های بیماری، ویژگی‌های مفید در تشخیص دیابت انتخاب شده و با اضافه‌کردن بعضی از علائم (مانند فیلدUrine) بعنوان پارامترهای ورودی، دقت سیستم به ٪۸۷,۲ افزایش یافته است. در این پژوهش نیز از مجموعه‌داده PIDD استفاده شده بود و بوسیله سیستم استنتاج ممدانی و با استفاده از توابع مثلثی، عمل فازی‌سازی متغیرهای ورودی (علائم بیماری) انجام می‌شد. سپس پایگاه‌دانش ایجاد و براساس ورودی‌ها و قوانین موجود در پایگاه‌دانش، خروجی ایجاد شده و در آخر عمل نافازی‌سازی انجام می‌شد. عیب اول این سیستم این است که ممکن است داده‌های ما دارای نویز باشند، بنابراین خود مرزهای توابع عضویت نیز با یک عدم‌قطعیت مواجه می‌شوند و نمی‌توان برای آن‌ها یک مقدار عددی قطعی تعیین کرد، در نتیجه دقت سیستم کاهش می‌یابد. عیب دوم آن، بالابودن تعداد قوانین فازی در پایگاه‌دانش آن بود که باعث بالارفتن زمان محاسبات سیستم و درنتیجه، پایین‌آمدن سرعت می‌شد. همچنین برای ارزیابی کارایی و دقت سیستم طراحی شده با رویکرد فازی نوع اول، با استفاده از تعیین شکل تابع مثلثی برای توابع عضویت، آزمایشی انجام شد که درنتیجه به عیب سوم این سیستم پی‌بردیم و آن میزان دقت کم سیستم است، زیرا مقادیر پیش‌بینی‌شده توسط سیستم با مقدار واقعی فاصله قابل قبولی ندارند. ممکن است خطای جواب‌های به‌دست‌آمده ناشی از نادرست بودن پارامترهای توابع عضویت باشد.

اسپرونک و همکارانش [۵]، طرحی با عنوان کنترلرهای فازی نوع دوم برای تنظیم فشار خون ارائه دادند. اولین دلیل اثباتی آن‌ها این بود که در تشخیص‌های پزشکی، فازی نوع اول نمی‌تواند به‌طور کامل و دقیق عمل تشخیص را انجام دهد و نظرات از یک خبره یا گروهی از خبره‌ها گرفته شده است و آزمایشات نیز وابسته به علم آن خبره‌هاست و دلیل نتایج به‌دست‌آمده متفاوت از خروجی سیستم‌های مختلف در واقع نظرات مختلف خبره‌ها می‌باشد. در این پژوهش، طی یک آزمایش پیاده‌سازی سیستم با فازی نوع اول، از نظرات و تشخیص پزشکی ۹ نفر از پزشکان استفاده شده بود. نتایج نموداری نشان می‌داد که با درنظرگرفتن مرزهای قطعی برای توابع عضویت «کم»، «متوسط» و «زیاد» در فازی نوع اول، نمی‌توان به جواب دلخواه و دقیقی رسید. شکل(۱) نظرات مختلف به‌دست‌آمده برای میانگین فشار شریان را نشان می‌دهد.

[5]Pima Indians Diabetes Database



شکل۱. نظرات وتشخیص پزشکی ۹ نفربرای توابع عضویت «کم»، «متوسط » و « زیاد »

در این پژوهش طبق شکل(۲)، دلیل اول استفاده فازی نوع دوم در علم پزشکی، متحدکردن نظرات تشخیصی پزشکان به یک نظر واحد بوده است. دلیل دوم این است که کنترلرهای فازی خیلی استاتیک هستند، بدین معنی که جواب خروجی وابسته به ورودی است و دقیقا باید ثابت، یکسان و بدون تغییر باشند. بنابراین همان سیستم را با فازی نوع دوم پیاده‌سازی کردند.

شکل ۲. تعیین یک محدوده فازی برای هریک از توابع عضویت

در ادامه با بررسی معایب این پژوهش‌ها، برای بهبود دادن این معایب، یک سیستم‌خبره با ترکیب رویکرد فازی نوع دوم و کاوش قواعد انجمنی ارائه می‌شود تا دقت سیستم افزایش یابد.

۳- اصطلاحات و مبانی موردنیاز

۳-۱- گلوکز، انسولین و کارکرد آن‌ها در بدن

غذایی که می‌خوریم شامل ۴ دسته از مولکول‌های سازنده بدن به نام کربوهیدارت‌ها[۶] (مواد قندی)، پروتئین‌ها، لیپیدها (مواد چربی) و اسیدهای نوکلئیک[۷] (مواد ژنتیکی) است. گلوکز (قندخون) منبع اصلی انرژی سلولی است که بویژه برای سلول‌های عضلانی ونیز مغز ضروری می‌باشد. برخی از ارگان‌های بدن مثل مغز به گلوکز به‌عنوان تنها منبع انرژی وابسته هستند.

[6]Carbohydrate
[7]Nucleic acid

در نتیجه برای بدن حفظ مقدار گلوکز در محدوده نرمال اهمیت دارد و مقادیر بالا و پایین آن در خون عواقب جدی به‌دنبال دارد. انسولین یک هورمون است که سبب انتقال گلوکز حاصل از هضم و جذب غذا در جریان خون می‌شود و این کار را توسط تغییر سوخت سلولی از چربی به گلوکز انجام می‌دهد. لازم به ذکر است که هورمون انسولین از پانکراس یا لوزالمعده (غده‌ای که درپشت معده قرارگرفته) به درون جریان خون ترشح می‌شود و در سطح سلولی به گیرنده‌های مخصوص متصل می‌گردد و ورود گلوکز به سلول‌ها و نیز سوخت و ساز طبیعی آنرا تسهیل مینماید.

۳-۲- مرض قند یا دیابت[۸]

مرض قند یا دیابت یک اختلال سوخت و ساز (متابولیک)[۹] در بدن است. در این بیماری توانایی تولید هورمون انسولین در بدن از بین می‌رود و یا بدن در برابر انسولین مقاوم شده، بنابراین انسولین تولیدی نمی‌تواند عملکرد طبیعی خود را انجام دهد. نقش اصلی انسولین انتقال، تنظیم و پایین آوردن قندخون است. دیابت دو نوع اصلی دارد; در دیابت نوع یک، تخریب سلول‌های بتا در پانکراس منجر به نقص تولید انسولین می‌شود و در دیابت نوع دوم، مقاومت پیش‌رونده بدن به انسولین وجود دارد که در نهایت ممکن است به تخریب سلول‌های بتای پانکراس و نقص کامل تولید انسولین منجر شود. در دیابت نوع دو مشخص است که عوامل ژنتیکی، چاقی و کم‌تحرکی نقش مهمی در ابتلای فرد دارند [۱].

۳-۲-۱-انواع دیابت

انجمن دیابت آمریکا، دیابت را به چهار گروه اصلی تقسیم می‌کند [۱]:

- دیابت نوع یک[۱۰] (وابسته به انسولین)
- دیابت نوع دو[۱۱] (غیر وابسته به انسولین)
- دیابت بارداری یا حاملگی[۱۲]
- انواع دیگر دیابت

[8]Diabetes (Diabetes Mellitus)
[9]Metabolic
[10]diabetes1type
[11]diabetes ۲ type
[12]Gestational Diabetes Mellitus

۳-۲-۲- تشخیص فرد مبتلا به دیابت

بر اساس آخرین معیارهای انجمن دیابت آمریکا [۱]، تشخیص دیابت با ثبت یکی از موارد زیر قطعی است:

۱. ثبت حداقل یک میزان قند خون تصادفی بالای ۲۰۰ میلی‌گرم بر دسی‌لیتر، به همراه علائم کلاسیک دیابت (پرنوشی، پرخوری، کاهش وزن بدون دلیل ،تشنگی فراوان، عفونت‌های مکرر)

۲. قند ناشتای پلاسمای بالاتر یا مساوی۱۲۶میلی‌گرم بر دسی‌لیتر (ناشتا یعنی۸ ساعت عدم‌مصرف هرنوع کالری قبل از انجام آزمایش)

۳. هموگلوبین A1C بالای ۶٫۵٪

۴. نتیجه آزمایش تحمل قند ۷۵ گرمی خوراکی دوساعته بالای ۲۰۰ میلی‌گرم بر دسی‌لیتر

برای تعیین اینکه بیمار، سالم، دیابتی یا پیش‌دیابتی (درمعرض دیابت) است، دو آزمایش قند خون ناشتا (FBG) و تست تحمل گلوکز خوراکی (OGTT) انجام می‌شود. براساس شاخص‌های انجمن دیابت آمریکا و همان‌طور که در شکل (۳) مشاهده می‌کنید، چنانچه میزان گلوکز خون درحالت ناشتا ۱۰۰ تا ۱۲۵ میلی‌گرم در دسی‌لیتر خون باشد، نشان‌دهنده پیش‌دیابتی و اگر گلوکز خون ناشتا در فرد ۱۲۶ یا بیشتر باشد، فرد دیابتی است. درصورتی فرد مبتلا به دیابت اعلام می‌شود که دوبار تست خون ناشتا در او بالاتر از حد مذکور (یعنی ۱۲۶) باشد. در تست تحمل گلوکز خوراکی، سطح گلوکز خون درحالت ناشتا و ۲ساعت بعد از نوشیدن شربتی غنی ازگلوکز، اندازه‌گیری می‌گردد. اگرمیزان گلوکز خون بعد از ۲ ساعت، ۱۴۰ تا ۱۹۹ میلیگرم در دسی‌لیتر خون باشد، فرد پیش‌دیابتی استواگر ۲۰۰ یا بیشترباشد، فرد دیابتی است.

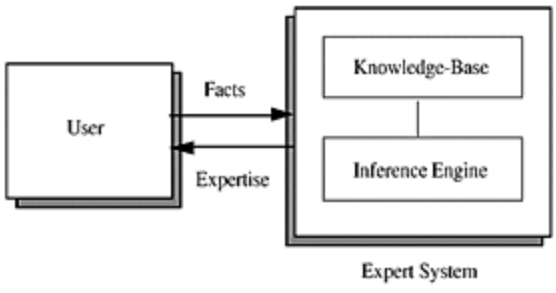
شکل ۳. تشخیص دیابت باتوجه به آزمایش میزان قندخون در حالت ناشتا (FBG) و همچنین نتیجهٔ آزمایش گلوکز پلاسما دو ساعت بعد ازغذا خوردن

۳-۳- سیستم‌خبره[13]

سیستم خبره یک نرم‌افزار کامپیوتری است که برای شبیه‌سازی اعمال و فعالیتهای یک فرد خبره، بعنوان نتیجه‌ای از توانایی سیستم‌ها برای پیداکردن حقایق جدید از حقایق موجود طراحی شده است و به شما اجازه می‌دهد تا کارهای هوشمند را به آن آموزش داده و به اجرا درآورید [۶]. اولین سیستم‌های مبتنی بر پایگاه‌دانش[14]، در دهه‌های ۱۹۷۰ و ۱۹۸۰ برای محیط‌های مستقل طراحی شدند که برمبنای کامپیوترهای بزرگ[15]یا کامپیوترهای شخصی و یا برمبنای کاربرهای توزیع‌شده شبکه محلی بودند. سیستم‌های‌خبره کلاسیک عبارتند از: Dendral برای تعیین ساختارمولکولی ذرات (۱۹۶۵)، MACSYMA (۱۹۶۸) برای کمک به ریاضی‌دانان، Prospector برای اکتشافات معدنی (۱۹۷۴)، MYCIN برای تشخیص اختلالات خونی (۱۹۷۶) و سیستم‌خبرهXCON (۱۹۸۰) برای پیکربندی سیستم‌های کامپیوتری[۷].

۳-۳-۱- ساختار کلی یک سیستم‌خبره

سیستم‌خبره، دو بخش کلی دارد: بخش اول، پایگاه‌دانش[16] است که به اختصار به آن KB می‌گویند. پایگاه‌دانش، شامل مجموعه‌ای از قواعد است که از یک فرد خبره یا گروهی از کارشناسان خبره در زمینه‌ای خاص گرفته شده‌است. بعنوان مثال، سیستم‌خبره دکترآنلاین، برای تشخیص بیماری و ارائه راه درمان آن. بخش دوم، موتور استنتاج[17] است که قواعد موجود در پایگاه‌دانش را با منطق خود مورد استفاده قرار می‌دهد. شکل(۴) ساختار کلی یک سیستم‌خبره را نشان می‌دهد[۸].

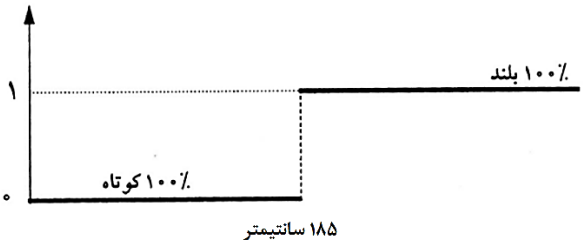


شکل ۴.ساختارکلی یک سیستم‌خبره

[13]Expert System
[14]KBS
[15]mainframe
[16]Knowledge-Base
[17]inference engine

۳-۴- منطق باینری، کلاسیک یا ارسطویی

در بسیاری از علوم نظیر ریاضیات و منطق، فرض براین است که مرزها و محدوده‌های دقیقا تعریف شده‌ای وجود دارد و یک موضوع خاص، یا درمحدوده آن مرز می‌گنجد یا نمی‌گنجد، یک متغیر یا متعلق به یک مجموعه است یا نیست. در مواردی از قبیل سفید یا سیاه، صفر یا یک، نیز این منطق صدق می‌کند. به این منطق که یک منطق دو ارزشی و باینری است، «منطق ارسطویی یا کلاسیک» می‌گویند. منطق دو ارزشی در بعضی موارد، ابزار توصیفی خوبی برای مدل‌سازی نمی‌باشد. مثلا برای تعیین قد افراد طبق این منطق، بلندی یک مفهوم دو ارزشی است. انسان‌ها یا بلندقد هستند یا نیستند و هیچ انسانی نمی‌تواند هم در مجموعه "بلند بودن"و هم در مجموعه "بلندنبودن" وجود داشته باشد. در چنین مجموعه‌ای شما در یک ارتفاع معین (مثلاً ۱۸۵ سانتیمتر) به‌طور ناگهانی از "بلند قد نبودن" به "بلند قد بودن" جهش می‌کنید. شکل(۵) این مطلب را نشانمی‌دهد.

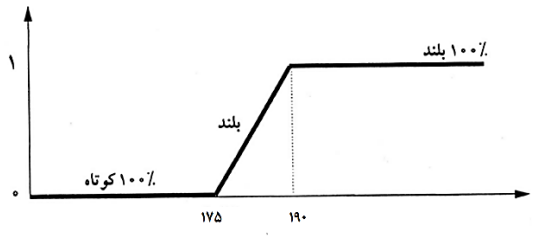


شکل۵.دیدگاه ارسطویی در ارتباط با بلندقد بودن

۳-۵- منطق فازی

کلمات و عباراتی وجود دارند که مرزهای مشخصی ندارند، کلماتی مثل"جوان"، "پیر"، "گرم"، "سرد". اکثر اتفاقات و رویدادهایی که در زندگی روزمره برای ما اتفاق می‌افتد، دارای ابهام می‌باشند. پروفسور لطفی عسگرزاده (معروف به پروفسور زاده) مجموعه‌های فازی را برای اولین بار در سال ۱۹۶۵معرفی‌کرد [۹] و ایده آن، با این عبارت توسط ایشان ایجاد شد: "ما نیاز به یک نوع مختلف از ریاضی داریم تا بتوانیم ابهامات و عدم دقت رویدادها را مدل‌سازی کنیم. مدلی که متفاوت از نظریه احتمالات می‌باشد". این مجموعه‌ها پایه‌گذار روشی موفق برای مدل‌کردن عدم‌قطعیت و ابهام بودند. از آن پس، استفاده از «مجموعه‌های فازی» در سیستم‌های کامپیوتری مخصوصاً درکاربردهای کنترلی، گسترش یافت [۹].

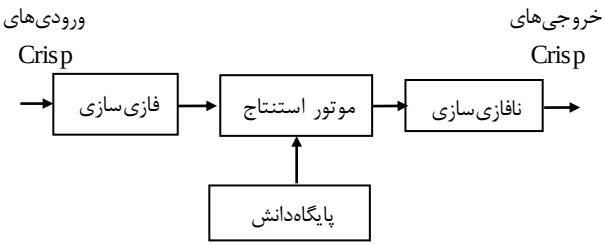
"بلند قد بودن"نیز مانند تمام خصوصیات دیگر در جهان، موضوع و مطلبی نسبی است. یک منحنی می‌تواند این تغییر یکنواخت را نشان دهد اما یک خط هرگز نمی‌تواند آن را نشان دهد (رجوع شود به شکل(۶)).



شکل ۶. دیدگاه فازی در ارتباط با بلندقد بودن

۳-۵-۱- ساختار سیستم فازی نوع اول

در شکل(۷) ساختار یک سیستم فازی نوع اول نشان داده شده است [۹]. ورودی‌های عددی و قطعی ۱۸ وارد سیستم می‌شوند و سپس در مرحله اول، عمل فازی‌سازی آن‌ها و ایجاد توابع عضویت برای آن‌ها انجام می‌شود. بعد از آن، با ترکیبی از ورودی‌های فازی شده، قواعدی در پایگاه‌دانش ایجاد می‌شود که این پایگاه‌دانش توسط موتور استنتاج استفاده می‌شود. در آخر عمل نافازی‌سازی انجام می‌شود تا داده‌های فازی، مجددا به داده‌های عددی تبدیل شوند. سیستم‌های فازی بعلت دارابودن توابع عضویت با درجات تعلق دقیق، توانایی محدودی در کاهش اثر عدم‌قطعیت در قوانین فازی دارند.



شکل ۷.بلاک دیاگرام کلی یک سیستم فازی نوع اول

۳-۶- مجموعه فازی نوع دوم[19]

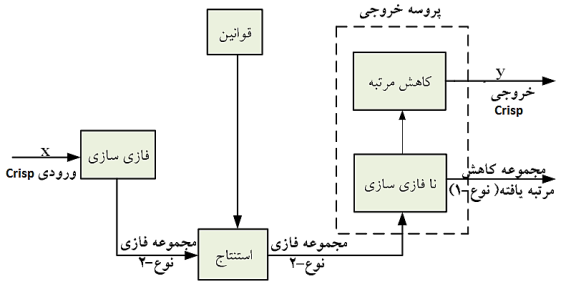
پروفسورزاده درسال ۱۹۷۵، مجموعه‌های فازی نوع-۲را به‌عنوان توسعه‌ای از مجموعه‌های فازی معرفی کرد [۱۰]. مجموعه فازی نوع-۲، یک مجموعه فازی است که دارای درجه عضویت‌های فازی است؛ به همین دلیل به آن، مجموعه **فازی- فازی** نیز

[18]Crisp
[19]Type-2 Fuzzy set

می‌گویند. چنین مجموعه‌ای درجایی که تعیین دقیق درجه عضویت برای یک مجموعه فازی مشکل است، مفید واقع می‌شود. سیستم فازی نوع-۲در برابر عدم قطعیت‌هایی که در قوانین فازی یا پارامترهای سیستم بوجود می‌آیند، مقاوم است و توانایی کاهش اثر و مدل‌کردن آن‌ها را دارد[۱۱،۱۲].

۳-۶-۱- ساختارسیستم فازی نوع دوم

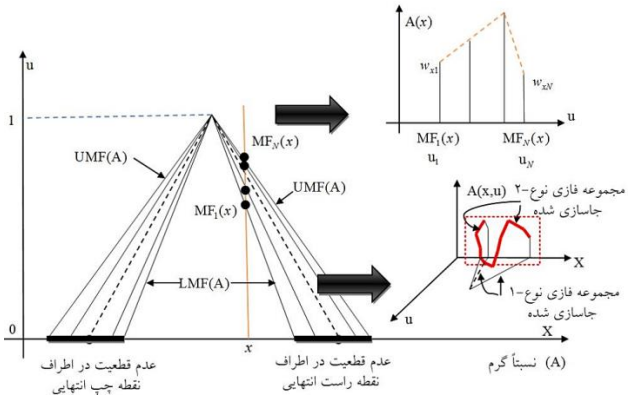
یک سیستم فازی نوع-۲، از چهار قسمت فازی‌سازی، قوانین، استنتاج و پروسه خروجی تشکیل شده است. درواقع یک سیستم فازی نگاشت بین یک ورودی غیرفازی و یک خروجی غیرفازی می‌باشد.در یک سیستم فازی نوع-۲، پروسه خروجی شامل دو مرحله می‌باشد. ابتدا نگاشت یک مجموعه فازی نوع-۲ به یک مجموعه فازی نوع-۱ که به این مرحله کاهش نوع یا کاهش مرتبه می‌گویند. سپس مرحله نافازی‌سازی مجموعه کاهش‌مرتبه یافته می‌باشد. شکل(۸) ساختار یک سیستم فازی نوع-۲ را نشان می‌دهد.



شکل ۸.ساختار یک سیستم فازی نوع-۲

۳-۶-۲- توابع عضویت در فازی نوع دوم

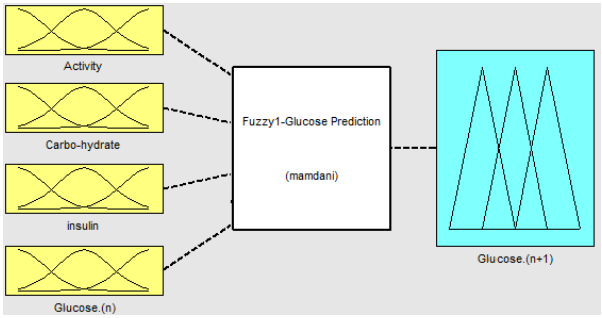
همان‌طورکه در شکل(۹) مشاهده می‌کنید، تابع عضویت یک مجموعه فازی نوع دوم، سه بعدی است [۱۳].



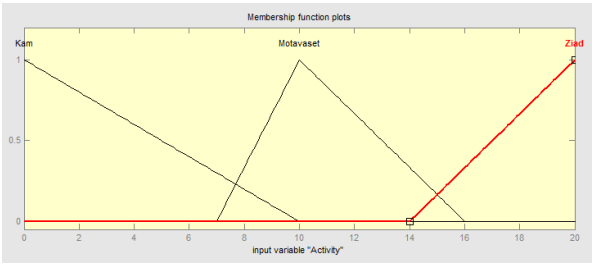
شکل۱۰.تابع عضویت یک مجموعه فازی نوع دوم

۴- بررسی سیستم فازی نوع اول برای تشخیص دیابت، در پژوهش‌های قبلی

در این بخش، روش‌های ارائه شده در پژوهش‌های قبلی و [۱۴] را از همه لحاظ ارزیابی می‌کنیم و معایب آن‌ها را بیان می‌کنیم و در روش پیشنهادی خودمان، با برطرف‌کردن تمام عیب‌ها، به یک سیستم جامع می‌رسیم. در بعضی از پژوهش‌های قبلی برای تشخیص دیابت، یک سیستم‌خبره فازی (نوع اول) طراحی شده است که ۴ ویژگی بعنوان ورودی و یک خروجی برای سیستم درنظر گرفته شده است. شکل (۱۰) نمای کلی این سیستم‌ها و شکل(۱۱) بخش فازی‌سازی ورودی‌های آن را نشان می‌دهد.



شکل۱۰.نمای کلی سیستم‌خبره فازی طراحی شده با ۴ ویژگی برای تشخیص دیابت

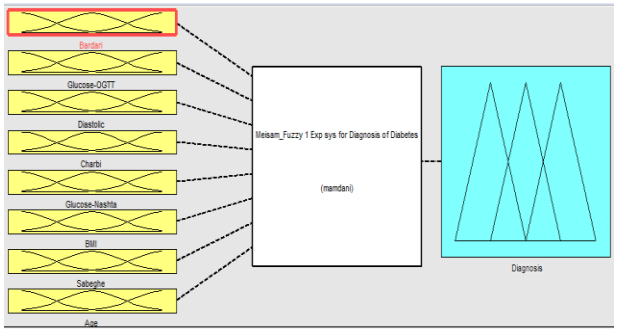


شکل۱۱.استفاده از فازی نوع اول در طراحی ورودی‌ها و خروجی، برای تعیین توابع عضویت آن‌ها

با نگاهی به نتایج این سیستم‌ها متوجه این عیب می‌شویم که آن‌ها در بعضی مواقع فقط برای تشخیص یک نوع دیابت مفید هستند و نمی‌توان با آن‌ها انواع دیابت را تشخیص داد. بنابراین در [۱۴]، بر اساس آخرین معیارهای انجمن دیابت آمریکا [۱]، با استفاده از مجموعه‌داده PIDD، طبق جدول (۱) تعداد ۸ ویژگی برای تشخیص دیابت انتخاب شده و سیستم تشخیص دیابت با فازی نوع اول و ۸ ویژگی پیاده‌سازی شده است. شکل(۱۲) نمای کلی سیستم طراحی شده را نشان می‌دهد.

جدول۱. ویژگی‌های فرد مبتلا به دیابت از نظر انجمن دیابت آمریکا

شماره	توضیح هر ویژگی
۱	تعداد دفعات بارداری
۲	غلظت گلوکز پلاسمای خون در آزمایش تحمل گلوکز خوراکی ۲ ساعته
۳	پایین‌ترین مقدار فشارخون بین دو تپش قلب
۴	ضخامت چربی زیر پوست
۵	میزان قندخون بعد از سرم انسولین دوساعته
۶	میزان سلامت وزن فرد با توجه به قدش (BMI)
۷	پیشینه قبلی و یا خانوادگی دیابت
۸	سن



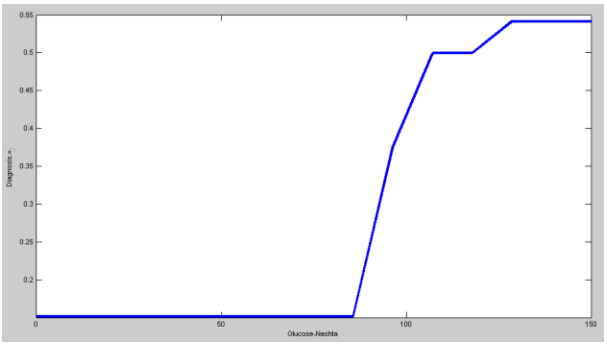
شکل۱۲.نمای کلی سیستم طراحی شده با فازی نوع اول و بهبود ویژگی‌ها

۴-۱- ایجاد پایگاه‌دانش

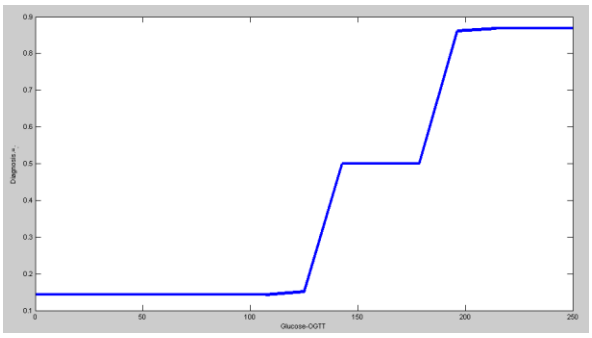
بعد از اینکه توابع عضویت برای متغیرها تعیین شدند، سپس براساس تفکر انسانی، روش درمان و دستورپزشک با ترکیب‌کردن ورودی‌ها و استفاده از آن‌ها در جملات شرطی If-Then، می‌توان تمام قواعدممکن را به‌وجود آورد. برای این سیستم باید بیش از ۵۰۰۰ قاعده تعریف کرد که یک پایگاه‌دانش جامع ولی حجیم ایجاد می‌شود. اما عیب این سیستم، بالا بودن تعداد قوانین فازی در پایگاه‌دانش آن است، زیرا تعداد بالای قوانین، همیشه باعث بهبود نتیجه سیستم نمی‌شود، بلکه گاهی اوقات باعث بالارفتن زمان محاسباتسیستم و درنتیجه، پایین آمدن سرعت می‌شود.

۴-۲- نمودار گرافیکی خروجی بخشی از سیستم

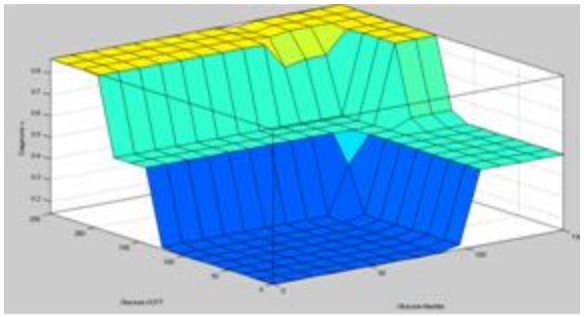
شکل‌های (۱۳)، (۱۴) و (۱۵)، نتایج ترکیب دو آزمایش تحمل گلوکزخوراکی دو ساعته OGTT و آزمایش گلوکزخون ناشتا در تشخیص دیابت را نشان می‌دهد [۱۴].



شکل ۱۳. آزمایش گلوکز خون ناشتا (FBG test)



شکل۱۴.آزمایش تحمل گلوکز خوراکی دوساعتهOGTT



شکل۱۵.نتیجه ترکیب دو آزمایش تحمل‌گلوکزخوراکی دوساعتهOGTT و آزمایش گلوکزخون ناشتا درتشخیص‌دیابت

۴-۳- معایب سیستم فازی نوع اول برای تشخیص دیابت

بعد از طراحی این سیستم‌خبره فازی، که پایگاه‌دانش آن توسط یک گروه از کارشناسان‌خبره ایجادشده بود و با مقایسه‌کردن نمودارهای خطی آن، با دو سیستم‌خبره فازی دیگری که توسط گروه دیگری از کاشناسان خبره طراحی شده بود، متوجه این عیب سیستم می‌شویم که هر سه سیستم برای یک هدف یکسان طراحی شده‌اند ولی نتیجه آن‌ها باهم متفاوت است. یکی از دلیل‌هایش متفاوت‌بودن نظر و سلیقه شخصی هر دسته از کارشناسان می‌باشد و دلیل دیگرش، تفاوت نژادها و تفاوت در وضعیت‌بدنی افرادمختلف در قاره‌های مختلف می‌باشد. همچنین ممکن است داده‌های ما دارای نویز باشند، بنابراین خود مرزهای

توابع عضویت نیز با یک عدم قطعیت مواجه می‌شوند و نمی‌توان برای آن‌ها یک مقدارِ عددیِ قطعی تعریف کرد، در نتیجه دقت سیستم کاهش می‌یابد. در جدول(۲)، معایب سیستم فازی نوع اول برای تشخیص دیابت نشان داده شده است.

جدول۲. معایب سیستم فازی نوع اول برای تشخیص دیابت

معایب سیستم طراحی شده با فازی نوع اول
مشکل تعیین درجه عضویت یا مقدار فازی برای یک عدد از محور X
عدم قطعیت در توابع عضویت به‌دلیل نویزی بودن داده‌های ورودی و در نتیجه، دقت پایین سیستم
تفاوت نتایج در سیستم‌های مختلف با هدف یکسان و کارشناسان خبره مختلف، به‌دلیل متفاوت‌بودن نظر و سلیقه شخصی هر دسته از کارشناسان و نیز تفاوت نژادها و تفاوت در وضعیت‌بدنی افرادمختلف در قاره‌های مختلف
عدم تفسیر ساده و فهم قوانین فازی به‌دلیل تعداد بسیارزیاد قوانین فازی در پایگاه‌دانش و نیزبالارفتن زمان محاسبات سیستم و پایین‌آمدن سرعت

۵- روش پیشنهادی

در این بخش، با استفاده از ترکیب علوم داده‌کاوی، فازی نوع دوم و سیستم‌خبره، یک سیستم‌خبره فازی نوع دوم را برای کمک به پزشک در تشخیص بیماری دیابت و یا جایگزین شدن به‌جای پزشک برای تشخیص خودکار دیابت ارائه می‌شود.

ابتدا با استفاده از تکنیک کاوش قواعد انجمنی، پایگاه‌دانش را بهبود داده و سپس با استفاده از فازی نوع دوم باعث بهبود یافتن مرزهای توابع عضویت و همچنین بهبود یافتن نقاط پارتیشن‌بندی می‌شویم که با این‌کار، دقت سیستم را تا حد بالایی افزایش می‌دهیم. با نگاهی ریزبینانه‌تر، طبق جدول (۳) این سیستم معایب سیستم‌های طراحی‌شده قبلی که با روش‌های ترکیبی و مختلف از قبیل فازی نوع اول، شبکه های عصبی، الگوریتم های ژنتیکی و .... طراحی شده بودند را برطرف می‌کند.

۵-۱- ایجاد پایگاه‌دانش قوی به‌وسیله تکنیک کاوش قواعد انجمنی [20]

در این بخش از انواع تکنیک‌های داده‌کاوی، تکنیک کاوش قواعد انجمنی برای بهینه‌سازی پایگاه‌دانش، استخراج و انتخاب بهترین و قوی‌ترین قواعد از پایگاه‌دانش استفاده می‌شود تا با این کار هم قدرت تفسیر پایگاه‌دانش بالا رود و هم زمان محاسبات پایین بیاید که در نتیجه سرعت سیستم بالا می‌رود. روش پیشنهادی در ایجاد یک پایگاه‌دانش قوی، برگرفته از روش به‌کارگرفته شده توسط زنگ و همکارانش [۱۵] می‌باشد، با این تفاوت که در آنجا از فازی نوع اول برای بهینه‌سازی مرزهای توابع عضویت استفاده شده است. مشکل این روش این است که آنها مجموعه داده ورودی را بخش‌بندی و سپس تا حدودی آن را فازی‌سازی کرده اند ولی در عمل دو نقطه پارتیشن‌بندی x_0 و x_1 ، عددی و قطعی هستند (رجوع شود به شکل (۱۷)) که ما برای رفع این مشکل آن‌را با فازی نوع دوم پیاده‌سازی می‌کنیم تا به‌جای داشتن دو نقطه در محل برخورد مرزهای دو تابع عضویت، یک محدوده از نقاط را در محل برخورد آن‌ها داشته باشیم.

۵-۱-۱- روند اجرا و انتخاب قوی‌ترین قواعد [21]

مجموعه‌داده PIDD که شامل اطلاعات ۷۶۸ بیمار دیابتی است را از سایت UCI گرفته که طبق جدول (۴) این مجموعه‌داده، شامل ۸ ویژگی می‌باشد که از نظر انجمن دیابت آمریکا برای تشخیص بیماری دیابت مهم هستند و نیز شامل یک فیلد باینری (فیلد ۹) برای تشخیص دیابتی بودن (مقدار۱) یا دیابتی نبودن (مقدار۰) می‌باشد. طبق شکل (۱۶)، مجموعه‌داده گرفته شده، براساس مقادیر مرتب‌سازی می‌شود و برای هریک از ویژگی‌ها (به‌جز ستون آخر که به‌صورت باینری می‌باشد) ابتدا محدوده مقدار آن‌ها را به پنج بازه تقسیم می‌کنیم. برای هر یک از ویژگی‌ها، سه مجموعه فازی FSL, FSM, FSR تعریف می‌شود.

[20] Association Rule Mining
[21] Strong Rule

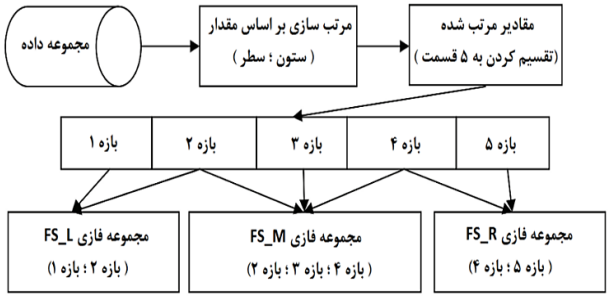
جدول۳. معایب سیستم‌های طراحی شده قبلی در تشخیص دیابت و راه حل بهبود دادن آن‌ها توسط روش پیشنهادی

احساس کمبود و عیب برای سیستم طراحی شده با فازی نوع اول	ارائه راه حل با استفاده از فازی نوع دوم و تکنیک‌های داده‌کاوی
عدم تفسیر ساده و فهم قوانین فازی به‌دلیل تعداد بسیارزیاد قوانین فازی در پایگاه‌دانش و نیزبالارفتن زمان محاسبات سیستم و پایین آمدن سرعت	استفاده از تکنیک کاوش قواعد انجمنی به‌منظور ارائه سطح بالاتری از دانش استخراج شده ازپایگاه‌دانش مذکور و تفسیر ساده قوانین موجود در پایگاه‌دانش
مشکل تعیین درجه عضویت یا مقدار فازی برای یک عدد از محور x	ایجاد یک محدوده فازی به‌صورت برش‌های عمودی برای تعریف‌شدن درجه عضویت
عدم قطعیت در توابع عضویت به‌دلیل نویزی بودن داده‌های ورودی و در نتیجه، دقت پایین سیستم	ایجاد توابع عضویت بالایی(UMF) و پایینی (LMF) به‌عنوان تلرانس کم و زیادشدن برای تعریف درجه عضویت بصورت فازی با برش‌های عمودی و محدوده‌ای
تفاوت نتایج در سیستم‌های مختلف با هدف یکسان و کارشناسان خبره مختلف، به‌دلیل متفاوت‌بودن نظر و سلیقه شخصی هر دسته از کارشناسان و نیز تفاوت نژادها و تفاوت در وضعیت‌بدنی افراد در قاره‌های مختلف	استفاده از توابع عضویت فازی نوع دوم جهت پوشش دادن تمام نظرات و تشخیص‌ها، با استفاده از توابع عضویت بالایی(UMF) و پایینی (LMF)

جدول۴. ویژگی‌های فرد مبتلا به دیابت ازنظر انجمن دیابت آمریکا

ویژگی	توضیح هر ویژگی
۱	تعداد دفعات بارداری
۲	غلظت گلوکز پلاسمای خون در یک آزمایش تحمل گلوکز خوراکی دو ساعته
۳	پایین‌ترین مقدار فشارخون بین دو تپش قلب (برحسب میلیمتر جیوه، mm Hg)
۴	ضخامت چربی زیر پوست
۵	میزان قندخون بعد از سرم انسولین دوساعته
۶	میزان سلامت وزن فرد با توجه به قدش (BMI)
۷	پیشینه قبلی و یا خانوادگی دیابت
۸	سن
۹	دیابت دارد یا ندارد ؟ (متغیر باینری ۰ یا ۱)

L و M و R موقعیت مجموعه‌های فازی (چپ، وسط، راست) را مشخص می‌کنند [۱۵]. سپس کل بازه اول را به FSL اختصاص می‌دهیم، یعنی در بازه اول، درجه عضویت FSL برابر با یک است و باقی درجات عضویت صفر هستند؛ به‌طور مشابه، کل بازه سوم را به FSM و کل بازی پنجم را به FSR اختصاص می‌دهیم. بازه دوم بین FSL و FSM مشترک است(با نامx_0) و بازه چهارم بین FSM و FSR (با نامx_1) مشترک است. قسمت مقدم[۲۲] قوانین به‌صورت<attribute ID, fuzzy set ID> می‌باشد. مقدار attribute ID از ۱ تا ۸ می‌باشد که بیانگر شماره ویژگی‌هاست و مقدار fuzzy set ID از ۰ تا ۲ می‌باشد که مقدار ۰ برای FSL ، ۱ برای FSM و ۲ برای FSR می‌باشد.



شکل۱۶. روند بهینه‌سازی فازی قواعد به‌دست‌آمده از ترکیب ورودی‌ها در پایگاه‌دانش

تابع عضویت برای بخش مشترک اول(x_0) به‌صورت زیر می‌باشد:

$$\mu_M(x) = r(x)^2(-2r(x)+3)$$
$$\mu_L(x) = 1 - \mu_M(x)$$
$$r(x) = x^{\alpha} \qquad (۱)$$
$$r(x_0) = 0.5$$

برای بخش مشترک دوم (بازه چهارم، یعنی x_1) نیز داریم:

$$\mu_R(x) = r(x)^2(-2r(x)+3)$$
$$\mu_M(x) = 1-\mu_R(x) \qquad (۲)$$
$$r(x_1) = 0.5$$

برای هر یک از ویژگی‌ها، نقاط پارتیشن مجموعه‌های فازی به‌صورت زیر محاسبه می‌شود:

$$x_0 \leftarrow 0.5 * (DP_0 + DP_1);$$
$$x_1 \leftarrow 0.5 * (DP_2 + DP_3); \qquad (۳)$$

[22] Antecedent

۵-۱-۲- روش کار کاوش قواعد انجمنی

در کد سیستم پیشنهادی، یک ماتریس دوبعدی به‌نام D تعریف شده است که اطلاعات نمونه‌ها (۷۶۸ نفر) را در خود ذخیره می‌کند. برای سنجش قواعد انجمنی از سه معیار ناحیه پوشش[۲۳]، درجه اطمینان[۲۴]، ضریب تاثیر[۲۵] استفاده می‌شود (D مجموعه تراکنش‌ها می‌باشد، $\mu_x(t)$ میزان تعلق آیتم را نشان می‌دهد). برای قانون A→B که A و B هر یک می‌توانند شامل چند دسته موارد[۲۶] باشند داریم:

ناحیه پوشش، به‌صورت زیر محاسبه می‌شود:

$$Supp(A \to B) = \frac{Supp(A \cup B)}{|D|} = \frac{\sum_{t \in D} \prod_{x \in A \cup B} \mu_x(t)}{|D|}$$

$$= P\,(AUB) = count\,(A) / n \qquad (۴)$$

درجه اطمینان،به‌صورت زیر محاسبه می‌شود:

$$Conf(A \to B) = \frac{Supp(A \cup B)}{Supp(A)} = \frac{\sum_{t \in D} \prod_{x \in A \cup B} \mu_x(t)}{\sum_{t \in D} \prod_{x \in A} \mu_x(t)} \qquad (۵)$$

ضریب تاثیر،به این صورت محاسبه می‌شود:

$$CF(A \to B) =$$

$$\begin{cases} \frac{Conf(A \to B) - Supp(B)}{1 - Supp(B)}, & Conf(A \to B) >= Supp(B) \\ \frac{Conf(A \to B) - Supp(B)}{Supp(B)}, & Conf(A \to B) < Supp(B). \end{cases} \qquad (۶)$$

۵-۱-۳- تابع کیفیت قواعد و انتخاب قواعد قوی‌تر

در ابتدا، مقادیر minSupp، minConf و minCF توسط کاربر تنظیم می‌شود. یک تابع $\phi(r)$ تعریف شده که کیفیت یک قانون را تعیین می‌کند. برای هر یک از قوانین مانند r ابتدا Sup(r) و Conf(r) و CF(r) مطابق با روابطی که در قسمت قبل توضیح داده شد، محاسبه شده و سپس برای به‌دست آوردن کیفیت قوانین داریم از رابطه ۷ استفاده میکنیم[15]:

$$\varphi(r) = \min \left\{ \begin{array}{l} \mathrm{Supp}(r) - \mathrm{minSupp} \\ \mathrm{Conf}(r) - \mathrm{minConf} \\ \mathrm{CF}(r) - \mathrm{minCF} \end{array} \right\} \qquad (۷)$$

[23] Support
[24] Confidence
[25] Certainty Factor
[26] ItemSet

قانون A→B را یک قانون قوی می‌نامیم اگر معیارهای ناحیه پوشش، درجه اطمینان و ضریب تاثیر برای این قانون از مقدار حد آستانه‌ای که کاربر تعیین کرده است بیشتر باشد. یعنی باید $\phi(r) > 0$ باشد. در حالت کلی، قواعدی قوی‌تر و بهتر هستند که:

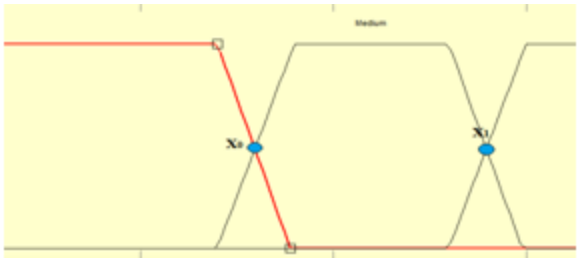
۱- ناحیه پوشش آن‌ها بزرگتر از MST[۲۷] باشد.

۲- درجه اطمینان آن‌ها بزرگتر ازMCT[۲۸] باشد . اگردرجه اطمینان بیشتر باشد، آن قاعده اصلی‌تر و بهتر است.

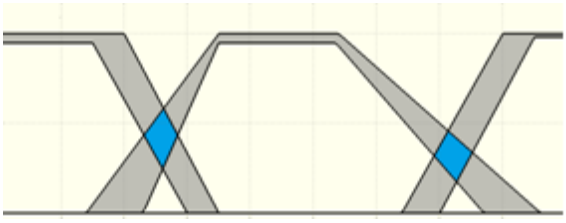
در نهایت مجموعه [F, R] به‌عنوان مجموعه‌ای بهینه‌شده از دسته موارد و مجموعه بهینه‌شده از قواعد به‌دست می‌آید.

۵-۲- برطرف‌کردن عیب روش قواعد انجمنی با فازی نوع دوم

روش پیشنهادی ما در ایجاد پایگاه‌دانش قوی، برگرفته ازروش زنگ و همکارانش بود با این تفاوت که طبق شکل (۱۷) در آنجا از فازی نوع اول برای بهینه‌سازی مرزهای توابع عضویت استفاده شده است. همانطور که قبلا اشاره شد عیب این روش این است که آنها مجموعه داده ورودی را بخش‌بندی و سپس تاحدودی آن را فازی‌سازی کرده اند ولی در عمل دو نقطه پارتیشن‌بندی x_0 و x_1، عددی و قطعی هستند. در ادامه، طبق شکل (۱۸) ما برای رفع این مشکل آن‌را با فازی نوع دوم پیاده‌سازی می‌کنیم تا به‌جای داشتن دو نقطه در محل برخورد مرزهای دو تابع عضویت، یک محدوده از نقاط را در محل برخورد داشته باشیم.



شکل۱۷. دونقطه پارتیشن‌بندیx0 و x1 ،عددی وقطعی هستند



[27] Minimum Support
[28] Minimum Confidence

شکل۱۸. دونقطه پارتیشن‌بندی x0 و x1 ، محدوده‌ای ازنقاط هستند

۵-۳-استفاده از فازی نوع دوم، برای رفع مشکل تاثیرِ منفیِ عدم قطعیت در مرزهای توابع‌عضویت

طبق آنچه که قبلا شرح داده شد، برای افزایش دقت یک سیستم می‌توان با استفاده از رویکرد فازی نوع اول، پارامترهای ورودی که دارای مقادیرِ عددیِ قطعی[۲۹] هستند را فازی کرد و برای هر متغیر ورودی، توابع عضویت «کم»، «متوسط» و «زیاد» را ایجاد کرد، ولی از آنجایی‌که با درنظرگرفتن مرزهای قطعی برای توابع عضویت «کم»، «متوسط» و «زیاد» در فازی نوع اول، نمی‌توان به جواب دلخواه و دقیقی رسید و با عدم قطعیت در مرزهای توابع عضویت فازی نوع اول مواجه هستیم و همچنین ممکن است داده‌های ورودی نویزی باشند و با توجه به این نکته مهم که نژاد و شرایط بدنی افراد در نقاط مختلف دنیا باهم فرق می‌کند، نتیجه حاصله از سیستم‌های خبره مختلف طراحی شده توسط کارشناسان خبره مختلف باهم فرق می‌کند. بنابراین در این سیستم پیشنهادی، با استفاده از رویکرد فازی نوع دوم و با ایجاد توابع عضویت بالایی[۳۰] و پایینی[۳۱] برای هر یک از توابع عضویت، خود توابع عضویت را نیز فازی می‌کنیم. با این‌کار، علاوه‌بر اینکه تاثیر منفی اختلاف نظر کارشناسان خبره بر روی نتیجه سیستم از بین می‌رود، عدم‌قطعیت در مرزهای توابع عضویت فازی نوع اول نیز از بین می‌رود و نقاط پارتیشن‌بندی نیز بهبود می‌یابدکه در نتیجه، دقت سیستم افزایش می‌یابد. بنابراین، همان سیستم فازی نوع اول پیاده‌سازی شده در [۱۴] را طبق آنچه در شکل(۱۹) مشاهده می‌کنید با استفاده از رویکرد فازی نوع دوم طراحی می‌کنیم. در این سیستم بوسیله سیستم استنتاج ممدانی (یک‌بار با استفاده از توابع گاوسی و بار دیگر با توابع ذوزنقه‌ای) عمل فازی‌سازی متغیرهای ورودی (ویژگی‌های بیماری) و ایجاد توابع عضویت بالایی و پایینی انجام میشود.

شکل۱۹. نمای کلی سیستم‌خبره فازی نوع دوم طراحی شده برای تشخیص دیابت

۵-۳-۱- مقداردهی توابع عضویت فازی نوع دوم برای ۸ ویژگی در تشخیص دیابت

برای سیستم پیشنهادی، طبق جدول(۵)، مقادیر تعریف شده برای ۸ ویژگی در تشخیص دیابت و توابع عضویت آن‌ها را آورده‌ایم. این مقادیر با توجه به نظر پزشکان مختلف تاحدودی باهم تفاوت دارند و این به‌دلیل تفاوت نژادها، تفاوت وضعیت بدنی، مکان زندگی، تفاوت جنسیت‌ها و سایر عوامل تاثیرگذار بر سیستم تشخیص است. لذا با استفاده از رویکرد فازی نوع دوم می‌توان یک تولرانسی برای مرزهای بالا و پایین این مقادیر به‌وجود آورد تا این تفاوتِ مقدار را برطرف کند و نتیجه واحدی از سیستم به‌دست آید که در ادامه آن را بررسی می‌کنیم.

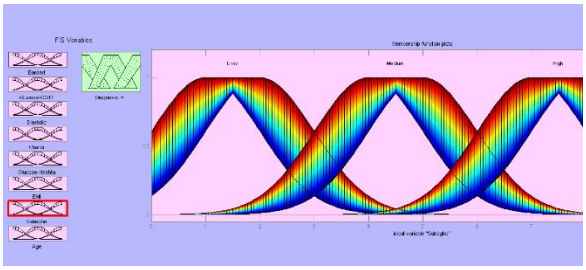
[30] UMF
[31] LMF

جدول۵. مقداردهی برای توابع عضویت ۸ ویژگی درسیستم تشخیص دیابت

نام ویژگی	دسته‌بندی مقادیر برای هر ویژگی	توابع عضویت
تعداد دفعات بارداری	0 - 3 3 – 5 > 5	Low Medium High
گلوکزخون بعد از ۲ ساعت غذا با آزمایش OGTT	0 – 139 140 – 199 > 200	Good Medium Bad
پایین‌ترین فشار خون (دیاستولیک)	0 - 80 90 – 99 100 – 109 > 110	Good Medium Bad Very Bad
چربی زیرپوست (با میانگین تولرانس ۳)	± 3 5.5 – 7.5 ± 3 10 – 11.5 ±3 17.5 – 20.5 ± 3 21.5 – 28 ± 3 29	Needed Very Good Good Medium Bad
میزان قند ناشتا	0 – 99 100 – 125 > 126	Good Medium Bad
وضعیت توده بدنی	± < 242 ± 242 ± > 242	Bad Good Bad
سابقه خانوادگی	<2 2 - 4 > 6	Low Medium High
سن	1 – 11 12 – 18 18 – 34 35 – 55 > 55	Baby Youngster Young Middle-aged Old

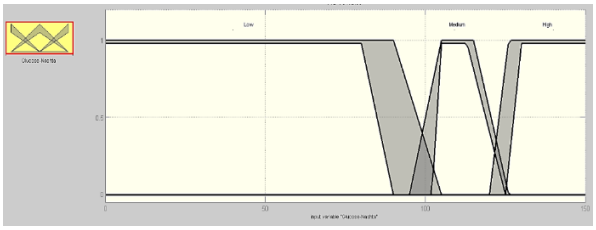
۵-۳-۲- طراحی ورودی‌ها و خروجی سیستم و رسم توابع عضویت آن‌ها

طبق آنچه که در شکل (۲۰) مشاهده می‌کنید با استفاده از رویکرد فازی نوع دوم و ابتدا با استفاده از شکل توابع گاوسی، عمل فازی‌سازی ورودی‌ها و خروجی و ایجاد توابع عضویت بالایی و پایینی را انجام می‌دهیم.



شکل۲۰. طراحی ورودی‌ها و خروجی سیستم و تعریف توابع عضویت گاوسی (فازی نوع دوم) برای آن‌ها

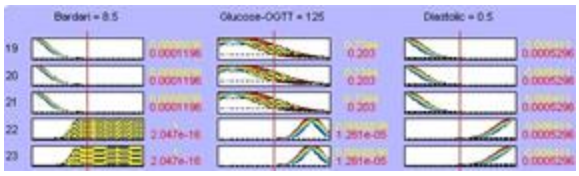
همچنین می‌توانیم طبق شکل(۲۱)، با استفاده از شکل تابع ذوزنقه‌ای و ایجاد توابع عضویت بالایی و پایینی نیز این کار را انجام دهیم. در سیستم پیشنهادی ما، هردو مدل طراحی جواب یکسانی را می‌دهد.



شکل۲۱. طراحی ورودی‌ها و خروجی سیستم و تعریف توابع عضویت ذوزنقه‌ای (فازی نوع دوم) برای آن‌ها

۵-۳-۳- پایگاه‌دانش در فازی نوع دوم

بعد از طراحی توابع، پایگاه‌دانش قوی طبق شکل (۲۲) ایجاد می‌شود. همان‌طور که مشاهده می‌کنید، در کنار هرشکل تابع عضویت، دوعدد (با رنگ‌های زرد و قرمز) قرار دارد که نشان‌دهنده توابع عضویت بالایی و پایینی می‌باشد.

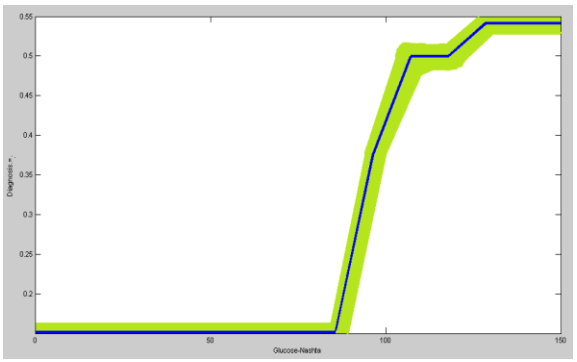


شکل۲۲. نمایش بخشی از قواعد پایگاه‌دانش درسیستم فازی نوع دوم بااستفاده از واسطگرافیکی

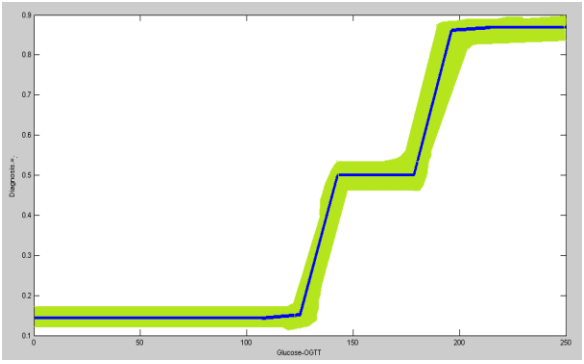
۵-۳-۴- نمایش بخشی از خروجی بهبودیافته سیستم با رویکرد فازی نوع دوم

با اعمال رویکرد فازی نوع دوم در توابع عضویت و مقایسه نمودارهای سیستم فازی نوع اول با نموداربه‌دست آمده از این سیستم پیشنهادی طراحی شده با فازی نوع دوم به این نتیجه می‌رسیم که با ایجاد توابع عضویت بالایی و پایینی برای هر یک از توابع عضویت، علاوه بر اینکه تاثیر منفی اختلاف نظر کارشناسان خبره بر روی نتیجه سیستم از بین می‌رود، عدم‌قطعیت در مرزهای توابع عضویت فازی نوع اول نیز از بین می‌رود و دقت سیستم نیز افزایش می‌یابد. شکل‌های (۲۳) و (۲۴)، بخشی از خروجی سیستم که مربوط به نتایج ترکیب دو آزمایش تحمل گلوکزخوراکی دو ساعته OGTT و آزمایش گلوکزخون ناشتا در تشخیص دیابت می‌باشد را با رویکرد فازی

نوع دوم و مجموعه‌ای از برش‌های عمودی و محدوده‌ای نشان می‌دهد.



شکل۲۳. نتیجه آزمایش گلوکزخون ناشتا با فازی نوع دوم



شکل۲۴. آزمایش تحمل گلوکزخوراکی دوساعتهOGTT با فازی نوع دوم

۶- محاسبه دقت با ماتریس آشفتگی

معیاردقت، به‌وسیله ماتریس آشفتگی به شکل زیر محاسبه می‌شود:

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP} X\ 100\ \% \qquad (۸)$$

جدول۶. ماتریس آشفتگی

کلاس واقعی	کلاس پیش‌بینی شده	
	بله	خیر
بله	TP	FN
خیر	FP	TN

بنابراین ماتریس آشفتگی برای روش پیشنهادی ما، به شکل زیر خواهد بود:

جدول۷.ماتریس آشفتگی برای روش پیشنهادی

	تشخیص مثبت (تشخیص داده می‌شود که فرد دیابت دارد)	تشخیص منفی (تشخیص داده می‌شود که فرد دیابت ندارد)
فرد واقعا مبتلا به دیابت است	TP	FP
فرد در واقعیت دیابت ندارد	FN	TN

با توجه به جدول (۷) برای فاز تست به دو جمعیت احتیاج داریم:

- در جمعیت اول می‌دانیم که جواب تشخیص برای آن‌ها باید درست باشد؛ یعنی می‌دانیم که فرد دیابت دارد. بنابراین از مجموعه‌داده، افرادی که مقدار ستون دیابتی برای آن‌ها ۱ است را در این جمعیت قرار می‌دهیم (تعداد این افراد ۲۶۸ نفر است).
- در جمعیت دوم می‌دانیم که جواب تشخیص برای آن‌ها باید غلط باشد؛ یعنی می‌دانیم فرد دیابت ندارد. پس از مجموعه‌داده، افرادی که مقدار ستون دیابتی برای آن‌هاصفر است را در این جمعیت قرار می‌دهیم. تعداد این افراد ۵۰۰ نفر است که ما به تعداد ۲۶۸ نفر از این ۵۰۰ نفر را انتخاب می‌کنیم تا با جمعیت اول برابر شود.

۶-۱- محاسبه دقت روش پیشنهادی (اندازه‌گیری معیار ارزیابیِ دقت)

دقت تشخیص با روش پیشنهادی، در جدول (۸) با استفاده از 10-fold cross validation برای هر فولد نشان داده شده است. میانگین کلی تمام مقادیر موجود در این ستون، برابر با عدد **۹۴,۸۸** درصد می‌باشد که نسبت به روش‌های برتر قبلی، دقت بالاتری را نشان می‌دهد.

جدول۸. مقادیر دقت به‌دست آمده از روش پیشنهادی

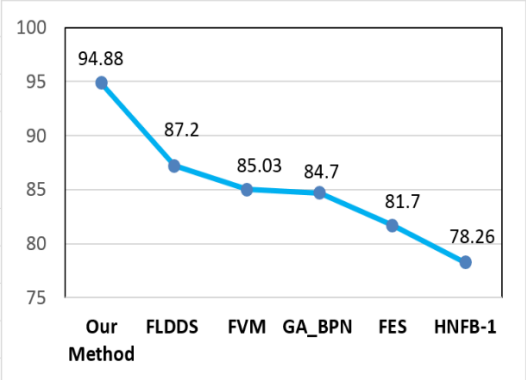
	روش پیشنهادی
Fold 1	94.96
Fold 2	95.14
Fold 3	94.4
Fold 4	94.58
Fold 5	94.58
Fold 6	95.14
Fold 7	94.96
Fold 8	94.96
Fold 9	94.58
Fold 10	95.52

۷- مقایسه دقت روش پیشنهادی با سایر روش‌ها

پس از به‌دست آمدن دقت روش پیشنهادی، آن را با سایر روش‌های استفاده‌شده در پژوهش‌های قبلی مقایسه می‌کنیم تا برتری روش ما نسبت به سایر روش‌ها اثبات شود (رجوع شود به جدول(۹) و شکل (۲۶)).

جدول۹. مقایسه دقت روش پیشنهادی با سایر روش‌ها

پژوهشگر	دقت(%)	روش
روش پیشنهادی	94.88	Optimization of Association Rule Mining by Type-2 Fuzzy Expert System
جین و رهجا [۴]	87.2	FLDDS for Very Young
کالپانا و کومار [۱۶]	85.03	FVM for Diabetes Decision Very Young
کارگودا و همکاران [۳]	84.7	GA_BPN
لی و ونگ [۱۷]	81.7	FES
کنکالوس و همکاران [۱۸]	78.26	HNFB-1



شکل۲۶. مقایسه دقت روش پیشنهادی با سایر روش‌ها

۸- نتیجه‌گیری

علیرغم این که روش‌های به‌کارگرفته‌شده در پژوهش‌های قبلی برای تشخیص دیابت، با استفاده از مدل ریاضی دقیق، در شبیه‌سازی جواب خوبی را ارائه می‌دهند، بسیار احتمال دارد که در عمل موفق نبوده و در هنگام اعمال به بیمار واقعی (به‌جای مدل ریاضی دقیق)، با شکست مواجه شوند. در این پژوهش، ما روشی را برای تشخیص انواع بیماری دیابت ارائه دادیم که معایب ذکرشده در پژوهش های قبلی را تا حدی بهبود می‌دهد و دقت سیستم را افزایش می‌دهد. در ابتدا پایگاه‌دانش آن را با استفاده از تکنیک کاوش قواعد انجمنی بهبود دادیم و قواعد قوی‌تر را برای پایگاه‌دانش استخراج کردیم تا پایگاه‌دانش ساده‌تر و قابل فهم‌تری داشته باشیم. در این مرحله یک مشکل پیش رو داشتیم، زیرا از فازی نوع اول برای بهینه‌سازی مرزهای توابع عضویت استفاده شده بود و عیب آن این بود که مجموعه داده ورودی، بخش‌بندی و سپس تا حدودی فازی‌سازی شده بود ولی در عمل دو نقطه پارتیشن بندی x_0 و x_1 ، عددی و قطعی بودند. در نهایت ما برای رفع این مشکل، آن‌را با فازی نوع دوم پیاده‌سازی کردیم تا به‌جای داشتن دو نقطه در محل برخورد مرزهای دو تابع عضویت، یک محدوده از نقاط را در محل برخورد آن‌ها داشته باشیم.در آخر نیز با انجام آزمایشات، به میزان ۹۴,۸۸ درصد دقت توسط سیستم پیشنهادی دست یافتیم و با مقایسه آن با سایر روش‌های انجام‌شده در پژوهش‌های قبلی، برتری روش پیشنهادی نسبت به سایر روش‌های پیشین را اثبات کردیم.

۹- منابع و مراجع

[1] American Diabetes Association, “Standards of Medical Care in Diabetes: Classification and diagnosis of diabetes”, Sec-2: S8-S16, 2015.

[2] M. Lichman, “UCI Machine Learning Repository”, [online], Available: http:// archive.ics.uci.edu/ml. PIDD, *CA:University of California, School of Information and Computer Science*, 2013.

[3] A.G. Karegowda, A. Manjunath, M. Jayaram, "Application of genetic algorithm optimized neural network connection weights for medical diagnosis of Pima Indians diabetes", *International Journal on Soft Computing*, 2011.

[4] V. Jain, S.Raheja, "Improving the Prediction Rate of Diabetes using Fuzzy Expert System", *I.J. Information Technology and Computer Science*, 2015.

[5] N. Sprunk, AM. Garcia, R. Bauernschmitt, A. Knoll, "Evaluation of an adaptive algorithm for fuzzy type-2 control in blood pressure regulation. In Fuzzy Systems (FUZZ)", *International Conference on IEEE*, 2013.

[6] H.K. Jabbar, R.Z. Khan, "Survey on Development of Expert System in the Areas of Medical, Education, Automobile and Agriculture",*In Computing for Sustainable Global Development (INDIACom), 2nd International Conference on IEEE (pp. 776-780)*, 2015.

[7] Jackson, Peter, "Introduction to Expert Systems", *third edition. Addison-Wesley, Harlow, England*, 1999.

[8] J.C. Giarratano, G.D, Riley, "Expert Systems: Principles and Programming", *fourth Edition, Boston, MA*, 2004.

[9] L. A. Zadeh, "Fuzzy Sets, Information and Control", *Vol. 8, pp.338-353*, 1965.

[10] L. A. Zadeh, "The Concept of a Linguistic Variable and Its Application to Approximate Reasoning-I", *Information Sciences, Vol. 8, pp. 199-249*, 1975.

[11] E. Hisdal, "The IF THEN ELSE Statement and Interval-Valued Fuzzy Sets of Higher Type", *International Journal of Man-Machine Studies, Vol.15, pp.385-455*, 1981.

[12] S. Coupland, R. John, "Type-2 Fuzzy Logic and the Modeling of Uncertainty, Fuzzy Sets and Their Extensions: Representation, Aggregation and Models.",*Fuzzy sets and their extensions: Representation, aggregation and models, 3-22*, 2008.

[۱۳] مرادی فراهانی، حسین، عسگری، جواد، ذکری، مریم،"مروری بر منطق فازی نوع دوم: از پیدایش تا کاربرد"*نشریه علمی-ترویجی محاسبات نرم*، شماره ۳،صفحه ۲۲ تا ۴۳، ۱۳۹۲.

[۱۴] روشن فکر، میثم، عبائی، گلنوش، "بهبوددادنویژگی‌ها در طراحیسیستم‌خبره فازی برای تشخیص بیماری دیابت"،*چهارمین کنفرانس بین المللی در مهندسی برق و کامپیوتر*، تهران،۱۳۹۵.

[15] H. Zheng, J. He, G. Huang, Y. Zhang, "Optimized Fuzzy Association Rule Mining for Quantitative Data", *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2014.

[16] M. Kalpana, A. V. S. Kumar, "Fuzzy Expert System for Diabetes using Fuzzy Verdict Mechanism", *International Journal Advanced Networking and Applications, Volume: 03, Issue: 02, pp 1128-1134*, 2011.

[17] C. S. Lee, M. H. Wang, "Ontology-based intelligent healthcare agent and its application to respiratory waveform recognition", *Expert System Application, vol. 33, no.3, pp.606-619*, 2007.

[18] L. B. Goncalves, M. M. B. R. Vellasco, M. A. C.Pacheco, and F. J. de Souza, "Inverted hierarchical neuro-fuzzy BSP system: A novel neuro-fuzzy model for pattern classification and rule extraction in databases", *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Application and Reviews, vol.36, no. 2, pp 236-248*, 2006.