فصلنامه علمی تخصصی ایده های نو در علوم، مهندسی و فناوری ، دوره ۲، شماره۲، تابستان ۱۳۹۷، ص ۷۵-۶۲
ISSN: 2588-3984
http://www.Tajournals.com

عنوان مقاله: تشخیص بیماری دیابت بر مبنای سیستم های فازی و الگوریتم بهینه‌سازی وال
دوره ۲ / شماره ۲ / تابستان ۱۳۹۷/ ص ۷۵- ۶۲

تشخیص بیماری دیابت بر مبنای سیستم های فازی و الگوریتم بهینه‌سازی وال

محسن خسروی۱ حسن زارعی۲ سارا خیری دونیقی۳

۱. هیئت علمی موسسه آموزش عالی شمس گنبد، گنبد، ایران(Mohsen.khosravi.sh@gmail.com)
۲ . هیئت علمی موسسه آموزش عالی شمس گنبد، گنبد، ایران(Hzarei1984@gmail.com)
۳دانشجوی ارشد مهندسی نرم افزار، موسسه آموزش عالی شمس گنبد، گنبد، ایران (Sara.kheiry96@gmail.com)

چکیده

تشخیص بیماری دیابت و یا آگاهی یافتن از احتمال بالای ابتلا به این بیماری همواره کار آسانی نخواهد بود. چرا که این بیماری علائم متعددی را بروز می‌دهد که بعضی از این علائم در سایر بیماری‌های دیگر نیز وجود دارند. بنابراین پزشک برای اتخاذ یک تصمیم مناسب، باید نتیجه‌ی آزمایش‌های بیمار و تصمیم‌هایی که در گذشته برای بیماران با وضعیت مشابه گرفته است، را بررسی کند. در این پژوهش ابتدا در مورد بیماری دیابت از مراکز درمانی داده جمع آوری شده است که هر داده دارای تعدادی ویژگی است، به این منظور روشی ترکیبی برمبنای سیستم‌های فازی و الگوریتم‌های بهینه سازی جهت تشخیص بیماری دیابت به کار گرفته‌ایم، که ابتدا با استفاده از سیستم های فازی تعداد افراد دیابتی و غیر دیابتی پیش‌بینی شده است سپس برای آنکه میزان تشخیص خطا کمتر شود از الگوریتم های فرا ابتکاری استفاده شده است که باعث بهبود و تکامل بخشیدن سیستم های فازی می‌شود تا کمترین میزان خطا را بدست آورد. در نهایت دقت به‌کارگیری روش ترکیبی را نسبت به سیستم فازی و همچنین نسبت به برخی الگوریتم‌های ارائه شده قبلی، می‌سنجیم. به‌کارگرفتن روش پیشنهادی در این پژوهش، در داده‌های مربوط به بیماری دیابت، به تشخیص زودرس و به تعویق انداختن این بیماری شایع در جوامع امروزی کمک می‌نماید. نتایج ارزیابی‌ها نشان می‌دهد که الگوریتم برای مجموعه داده‌ی دیابت دارای کارایی بسیار بالایی می‌باشد.

واژگان کلیدی: بیماری دیابت، سیستم‌های فازی، الگوریتم بهینه سازی وال، الگوریتم بهینه سازی ذرات، الگوریتم ژنتیک

مقدمه

افزایش استفاده از کامپیوترها در فعالیت‌های کسب و کار، منجر به رشد سریع پایگاه‌های اطلاعاتی و اجتماع داده‌ها توسط بیشتر سازمان‌ها شده است. روزانه حجم عظیمی از داده‌ها تولید شده و در پایگاه‌های مختلف داده ذخیره می‌شود. در سال‌های اخیر تمایل به جستجو برای کشف الگوهای تکرارپذیر به منظور بهبود در تصمیم گیری افزایش چشمگیری داشته است. همچنین کاوش در داده‌های تراکنشی جهت یافتن الگوهای پنهان و تکنیک‌های کشف دانش به منظور شناخت دقیق‌تر و بیشتر تراکنش‌ها، اهمیت بسزایی یافته است[۱]. در حوزه پزشکی و سلامت با افزایش استفاده از سیستم‌های جامع درمانی و پرونده‌های الکترونیک بیمار در بیمارستان‌ها و مراکز درمانی حجم انبوهی از اطلاعات مربوط بیماران و انواع بیماری‌ها مهیا می‌شود[۲]. استخراج دانایی از حجم عظیم داده‌های مرتبط با سوابق بیماری و پرونده‌های پزشکی افراد با استفاده از فرآیند داده کاوی می‌تواند منجر به شناسایی قوانین حاکم بر ایجاد، رشد و افت بیماری‌ها گردیده و اطلاعات ارزشمندی را به منظور شناسایی علل وقوع بیماری‌ها با توجه به عوامل محیطی حاکم در اختیار متخصصین و دست‌اندرکاران حوزه سلامت قرار دهد؛ که این امر در نهایت منجر به افزایش متوسط طول عمر افراد جامعه و ایجاد آرامش می‌گردد[۳].

آنچه مسلم است با افزایش سیستم‌های الکترونیک سلامت حجم داده‌های پزشکی هر روزه در حال افزایش است. اما این مجموعه داده‌های بزرگ به طور خام هیچ کاربردی ندارد برای آنکه بتوان از این داده‌ها ارزشی را استخراج کرد نیاز به تحلیل داده‌ها و تبدیل آن به اطلاعات و دانش، یک نیاز اساسی است. با توجه به چنین حجمی از داده‌ها استفاده از عامل انسانی به عنوان تشخیص‌دهنده الگوها و تحلیلگر داده‌ها پاسخگو نمی‌باشد؛ لذا داده کاوی روی داده‌های پزشکی از اهمیت بالایی برخوردار است. داده کاوی را می‌توان از جنبه‌های مختلف در پیشگیری یا تشخیص انواع بیماری، انتخاب روش‌های درمان بیماری، مدت زمان بستری بیمار و ... به کار برد.

اهداف تحقیق

۱. ارائه یک روش ترکیبی توسط الگوریتم فازی و وال برای بالا بردن دقت تشخیص بیماری دیابت.
۲. انتخاب ویژگی‌های مناسب و مفید، و حذف ویژگی‌های غیر مهم، به منظور کاهش ابعاد مسأله و پیچیدگی‌های محاسباتی، و افزایش دقت مسأله
۳. بهبود کیفی طبقه بندی داده‌ها
۴. افزایش دقت تشخیص بیماری دیابت
۵. مقایسه روش پیشنهادی با برخی روش‌های ارائه شده قبلی در زمینه تشخیص بیماری دیابت

دستیابی به این اهداف باعث تشخیص درست و به‌موقع بیماری دیابت گردیده و درنتیجه درمان سریع‌تر صورت می‌گیرد و از عواقب پرخطر آن پیشگیری می‌شود.

روش تحقیق

نوع روش تحقیق: بررسی و مطالعه منابع علمی در حوزه‌های تشخیص بیماری دیابت، داده کاوی، الگوریتم‌های طبقه‌بندی داده کاوی، سیستم فازی و الگوریتم‌های تکاملی.

تعیین استراتژی حل مسأله به‌صورت دقیق و مشخص نمودن فلوچارت کلی روش پیشنهادی.

انتخاب نمونه مورد مطالعه مناسب و به‌دست آوردن داده‌ها (۱۲۷۴ نمونه از داده های استاندارد مورد مطالعه بیماران دیابتی گنبد).

تحلیل معایب و مشکلات موجود و ارائه پیشنهاداتی برای رفع مشکلات.

شبیه سازی پیش‌پردازش و روش ترکیبی پیشنهادی در محیط نرم افزار متلب و انجام این شبیه‌سازی‌ها بر روی مجموعه داده مورد مطالعه، استفاده از معیارهای ارزیابی استاندارد مانند ضریب تعیین، مربعات خطا و...

مقایسه روش پیشنهادی با برخی روش‌های دیگر که قبلاً در جهت تشخیص بیماری دیابت، به‌کار گرفته شده است.

روش گردآوری اطلاعات

- استفاده از مقالات معتبر نمایه شده در IEEE و ELSEVIER.
- استفاده از کتب مربوطه
- استفاده از پایان نامه‌های مرتبط موجود در کتابخانه ملی
- استفاده از مجموعه داده استاندارد آماده در زمینه بیماری دیابت موجود در اینترنت
- استفاده از یک یا چند مرکز درمانی جهت تهیه مجموعه داده

روش تجزیه و تحلیل اطلاعات

داده کاوی برای بررسی داده‌های زیاد که هریک از این داده‌ها یک سری ویژگی دارد و به دست آوردن رابطه بین ویژگی‌های آن‌ها استفاده می‌شود. از بین الگوریتم‌های داده کاوی، الگوریتم‌های دسته‌بندی برای این تحقیق مناسب‌تر می‌باشند. زیرا در این الگوریتم‌ها نمونه‌ها دارای برچسب کلاس هستند و هدف تعیین برچسب کلاس یک نمونه جدید می‌باشد. قصد داریم از الگوریتم های و سیستم‌های فازی در بهبود دقت طبقه‌بندی داده‌های مربوط به بیماری دیابت استفاده نموده، به تشخیص این بیماری بپردازیم و نتیجه را با برخی الگوریتم‌های ارائه شده پیشین، مقایسه نماییم.

در این تحقیق، جهت شبیه سازی الگوریتم‌ها و نتایج، از نرم‌افزار MATLAB 2015a استفاده می‌نماییم.

پیشینه تحقیقات در زمینه تشخیص دیابت

استفاده از تکنیک‌های داده‌کاوی در پزشکی یکی از زمینه‌های پرکاربرد داده‌کاوی محسوب می‌شود که در سال‌های اخیر تحقیقات و مطالعات زیادی پیرامون آن انجام شده است. هنگامی که تعداد پارامترها در تشخیص بیماری، زیاد شود ممکن است تشخیص بیماری حتی برای یک متخصص خبره پزشکی نیز به سختی امکان پذیر باشد. همین دلیل موجب شده است که در چند دهه اخیر، ابزار تشخیص کامپیوتری با هدف کمک به پزشک، مورد استفاده قرار گیرد تا به نحوی بی‌نظمی را از داده‌ها خارج کند و همچنین خطاهای احتمالی ناشی از خستگی و یا بی تجربگی فرد خبره را کاهش دهد و داده‌های پزشکی مورد نیاز را در زمان کمتر و با جزئیات بیشتر در اختیار پزشک قرار دهد. از دیدگاه داده کاوی، پیش‌بینی در تشخیص بیماری، جزء مسائل دسته‌بندی داده‌ها محسوب می‌شود.

در مطالعات مربوط به تشخیص بیماری دیابت، برخی مشخصات عده‌ای از بیماران دیابتی و افراد معمولی به الگوریتم‌های طبقه-بندی داده کاوی داده می‌شود. این الگوریتم‌ها می‌توانند مدل‌هایی را برای طبقه‌بندی بیماران به دو دسته "بیمار دیابتی" و "افراد سالم"، ایجاد نمایند. از این مدل‌های ایجاد شده می‌توان به منظور طبقه‌بندی مراجعان جدید و افراد مشکوک به بیماری دیابت استفاده نمود و بیماری یا سلامت افراد جدید را پیش‌بینی نمود.

در سال ۲۰۰۶ آقای سو و همکارانش [۴] بر پایه چهار روش داده‌کاوی: شبکه‌های عصبی مصنوعی، درخت تصمیم، رگرسیون لجستیک و قواعد وابستگی و با استفاده از عکس‌های سه بعدی بدن توانستند با دقت قابل قبولی بیماری دیابت را در مراجعان تشخیص دهند. در این مطالعه، عکس‌های سه بعدی و دو بعدی از تمامی اعضای بدن افراد مختلف (بیماران دیابتی و افراد سالم) گرفته شد. سپس از این عکس‌ها متغیرهایی مانند: سطح شکم، محیط پاها، حجم دست‌ها و... استخراج شد. این متغیرها به چهار الگوریتم مورد نظر داده شدند و در نهایت با استفاده از این الگوریتم‌ها محققان توانستند بر پایه عکس‌های سه بعدی و دو بعدی

بدن با دقت ۸۹٪ پیش بینی کنند که آیا فرد مورد نظر به بیماری دیابت نوع دوم مبتلاست یا خیر. یکی از مزیت‌های این روش تشخیصِ بیماری دیابت نوع دوم، در این است که افراد مشکوک به بیماری نیازی به انجام تست‌های خونی ندارند. همچنین این محققان توانستند قواعدی را استخراج کنند که با استفاده از آن بتوانند بیماری دیابت نوع دوم را در افراد تشخیص دهند. به طور مثال آنها قواعدی مانند "اگر حجم شکم= x و حجم پا = Yو مساحت سینه = Z باشد آنگاه با احتمال ۹۰٪ بیمار دیابت نوع دوم دارد" را از داده استخراج نمودند.

همچنین در سال ۲۰۱۰ براکات و همکارانش با استفاده از روش ماشین بردار پشتیبان توانستند دقت تشخیص بیماری را بهبود دهند و با دقت ۹۴٪ بیماری دیابت نوع دو را در افراد تشخیص دهند. آنها داده‌های تحقیق خود را از داده‌های مربوط به ۴۶۸۲ نفر مراجعه کننده استخراج کردند و متغیرهای آنها شامل: جنسیت، BMI، فشار خون، کلسترل و قند خون، بود [۵].

در سال ۲۰۱۰ عده‌ای از محققان تایلندی [۶] با استفاده از درخت تصمیم توانستند با دقت بیش از ۹۰٪ سندرم متابولیک را در افراد تشخیص دهند آن‌ها داده‌های مربوط به ۵۶۳۸ نفر را به این منظور وارد الگوریتم درخت تصمیم نمودند. در انتهای تحقیق نیز قواعدی را کشف کردند که به وسیله آن‌ها بتوان با دقت بیشتری تشخیص را در بیماران انجام داد.

گروهی از محققان همچنین توانستند مجموعه‌ای از قواعد را برای تشخیص بیماری دیابت نوع دوم استخراج نمایند که این قواعد می‌تواند در آینده جایگزین دستورالعمل‌های تشخیص پزشکی باشد و پزشکان را در تشخیص بیماری دیابت یاری نماید[۷].

همچنین در سال‌های اخیر گروه‌های مطالعاتی از جمله مصطفی فتحی گنجی و همکارانش [۸] دقت تشخیص بیماری دیابت نوع دوم را با استفاده از الگوریتم‌ها و روش‌های داده‌کاوی بهبود بخشیده‌اند.

باقرزاده خیابانی و همکاران، جهت تشخیص بیماری دیابت نوع ۲، مدلی ترکیبی ارائه داده‌اند که از ترکیب سه الگوریتم طبقه‌بندی بیز ساده، درخت تصمیم و ماشین بردار پشتیبان تشکیل گردیده است. طبق نتایج به‌دست آمده توسط آن‌ها، دقت به دست آمده ناشی از این مدل از دقت تک تک طبقه‌بندکننده‌های ترکیب شده، بالاتر است [۹].

همچنین اسمیت و همکارانش در [۱۰] از یک شبکه عصبی برای تشخیص زودهنگام شروع دیابت استفاده کردند که به دقت ۷۶ درصد رسیدند.با وجود تحقیقات و مطالعات بسیاری که در زمینه تشخیص بیماری دیابت نوع۲ با استفاده از داده کاوی، به‌کار برده شده است، هنوز روشی قطعی به منظور تشخیص این بیماری، معرفی نشده است و ادامه تحقیقات در این زمینه، با توجه به توسعه روز افزون تکنیک‌های مبتنی بر داده کاوی ضروری به نظر می‌رسد. به همین دلیل هدف محقق از این تحقیق، ارائه یک روش ترکیبی جدید برای بالا بردن دقت تشخیص بیماری دیابت است.

برای تشخیص، خوشه بندی و طبقه بندی مسائل از شیوه‌های مختلفی می شود استفاده کرد. تشخیص دیابت هم با توجه به ماهیت باینری بودن خروجی آن به صورت یک مساله طبقه بندی در نظر گرفته می شود. در این تحقیق الگوریتم های تشخیص فازی را برای این مساله انتخاب کردیم که نتایج آن در این فصل آورده شده است. در ابتدا در مورد خود داده های دیابت بحث می-کنیم و سپس تمام فاکتور های مهم ارزیابی را بیان می‌کنیم و بعد از آن به شرح عملکرد سه سیستم فازی، عصبی فازی و کامیزفازی می‌پردازیم و بعد توسعه و بهبود سیستم با الگوریتم های فراابتکاری و درانتها مقایسه بین آنها انجام می‌دهیم.

مجموعه داده دیابت شهرستان گنبد کاووس

اطلاعات مربوط به دیابت از دو منبع بدست می‌آید: از یک دستگاه ضبط الکترونیکی اتوماتیک و دیگری سوابق موجود روی کاغذ.دستگاه اتوماتیک دارای یک ساعت داخلی است که به وقایع برچسب زمانی تخصیص می‌دهد، در حالیکه اطلاعات موجود بر روی کاغذ فقط زمان منطقی مانند صبحانه ناهار شام و زمان قبل از خواب را ارائه میدهد. برای اطلاعات کاغذی زمان‌های ۸:۰۰ برای صبحانه، ۱۲:۰۰ برای ناهار، ۱۸:۰۰ برای شام و ۲۲:۰۰ برای زمان قبل از خواب ارائه شده است. بنابر این اطلاعات روی کاغذ

فصلنامه علمی تخصصی ایده های نو در علوم، مهندسی و فناوری ، دوره ۲، شماره۲، تابستان ۱۳۹۷، ص ۷۵-۶۲
ISSN: 2588-3984
http://www.Tajournals.com

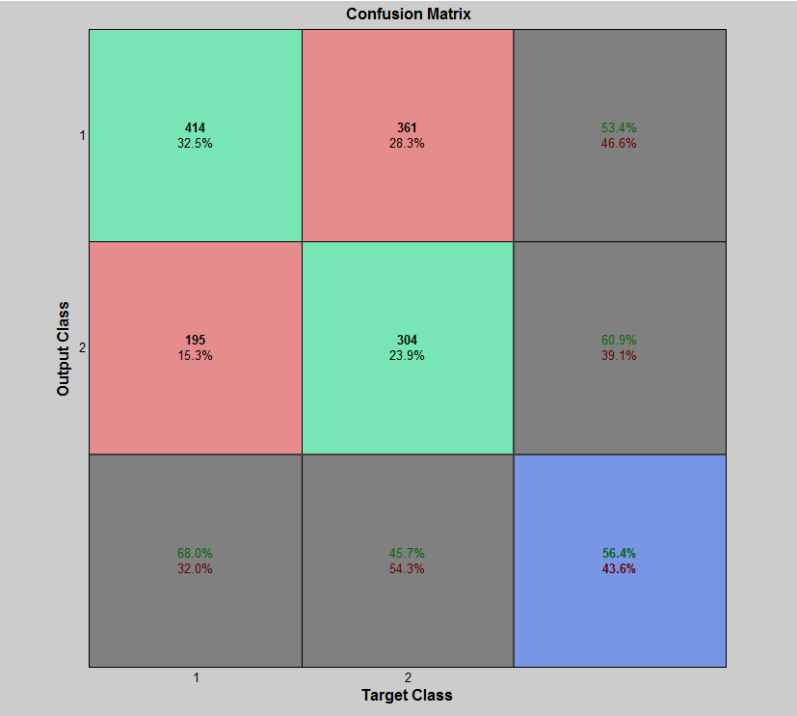
دارای یک زمان های یکنواخت و ساخته شده هستند در حالیکه که اطلاعات الکترونیکی دارای زمان‌های واقعی تر است. اطلاعات مربوط به مجموعه داده از تعداد ۱۲۷۴ بیمار شهر گنبد کاووس بدست آمده است که به طور خلاصه در جدول زیر قابل مشاهده می‌باشد.

جدول ۱: مجموعه داده مورد استفاده در این پژوهش

نام مجموعه داده	تعداد ویژگی ها	تعداد نمونه ها
دیابت شهر گنبد	۲۰	۱۲۷۴

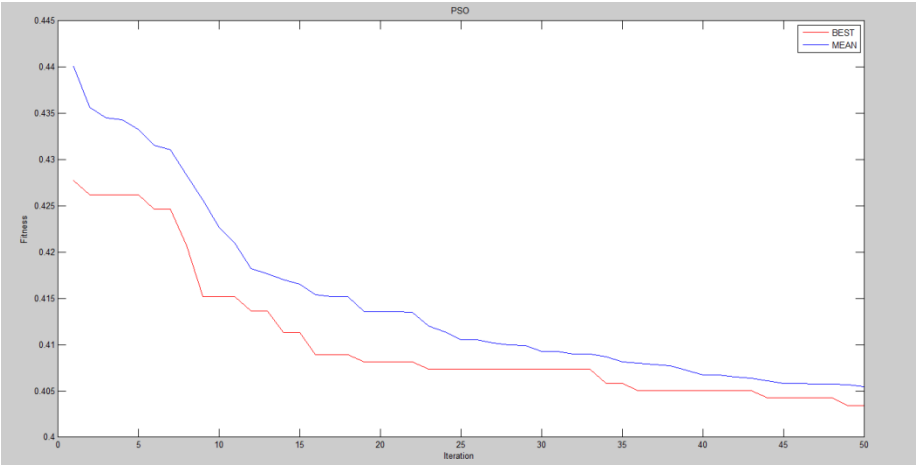
عملکرد سیستم فازی و سیستم فازی توسعه داده شده با الگوریتم های تکاملی

دراین پژوهش ۸۰٪ دیتا را به عنوان داده آموزش و مابقی را به عنوان داده تست انتخاب کردیم. شیوه انتخاب داده‌ها هم به صورت تصادفی است. نرم افزار مورد استفاده در این تحقیق نرم افزار MATLAB نسخه ۲۰۱۵ می باشد. در پژوهش‌های انجام شده در زمینه‌ی پیش بینی خطا در نرم افزار، مدل‌های کلاس بندی مبتنی بر متریک های نرم افزاری، یک ماژول نرم افزاری را به عنوان مستعد خطا و یا غیر مستعد خطا پیش بینی می‌کنند. برای سیستم فازی در این تحقیق از تابع آماده متلب genfis2 استفاده کرده ایم. که بر مبنای خوشه بندی عمل می‌کند. در این تابع شعاع پوشش را برابر ۵ در نظر گرفتیم نتایج بدست آمده به شرح زیر است.



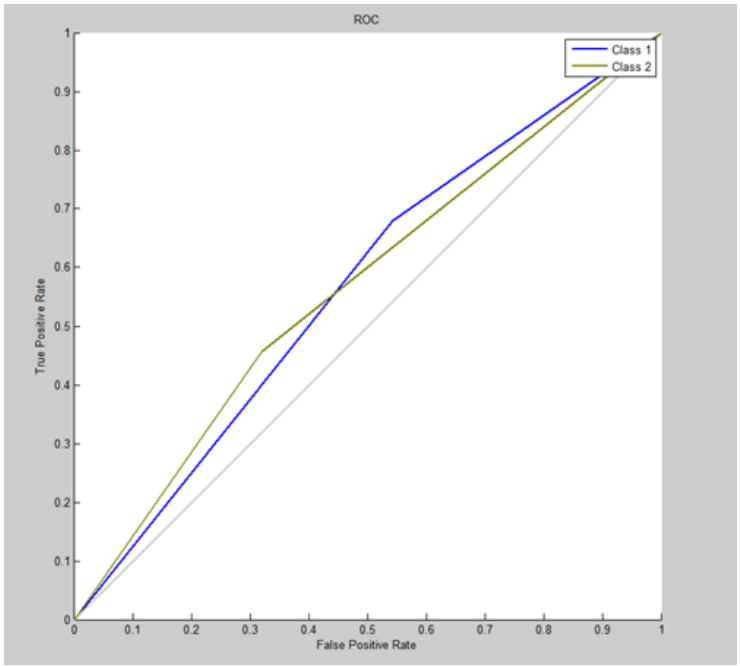
شکل ۱: مقادیر کانفیوژن برای سیستم فازی

فصلنامه علمی تخصصی ایده های نو در علوم، مهندسی و فناوری ، دوره ۲، شماره۲، تابستان ۱۳۹۷، ص ۷۵-۶۲
ISSN: 2588-3984
http://www.Tajournals.com

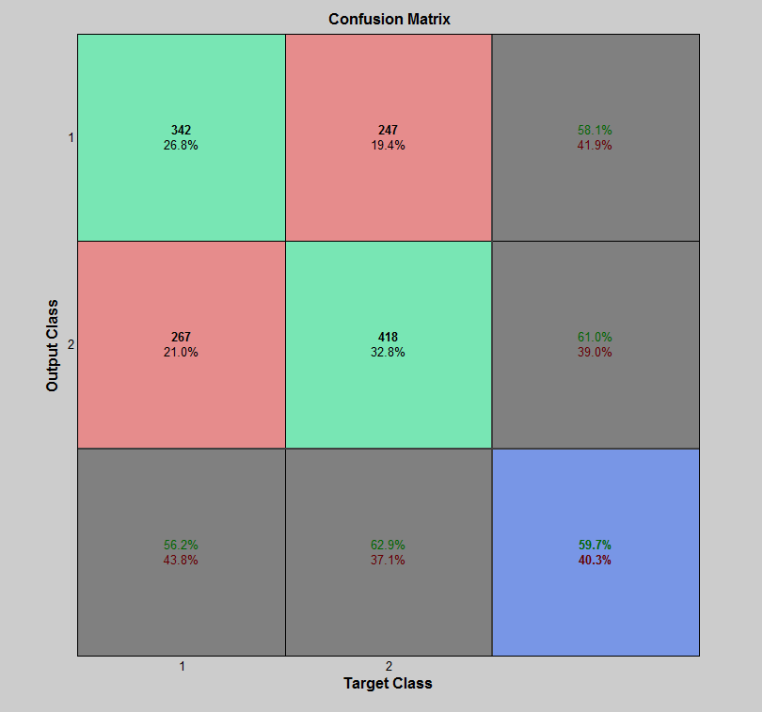


شکل ۲: نمودار همگرایی الگوریتم PSO برای بهبود سیستم فازی

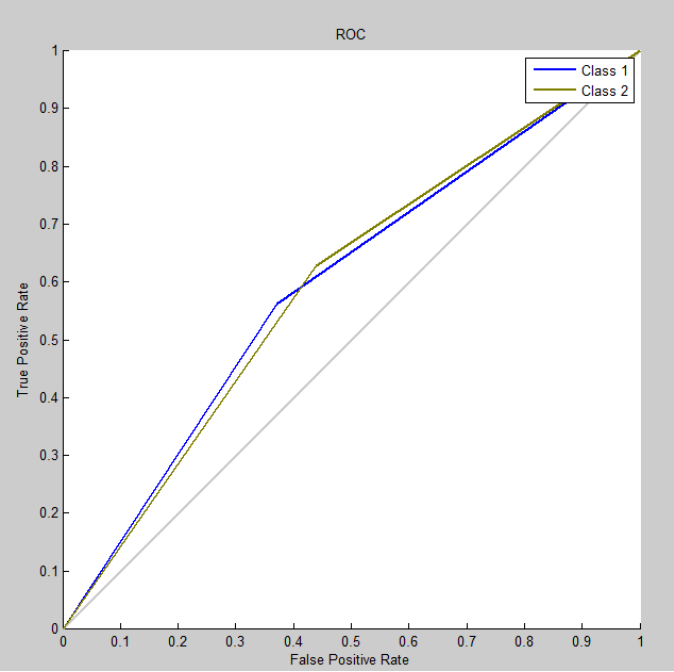
برای بهبود عملکرد سیستم فازی از الگوریتم های فراابتکاری استفاده کردیم بدین شکل بعد ساخته شدن سیستم فازی مقادیر توابع عضویت توسط الگوریتم بهبود می‌یابد به طوری در نتیجه نهایی دارای دقت بهتری نسبت به حالت اولیه باشیم. نتایج بدست امده از الگوریتم ها به شرح زیر است. برای الگوریتم PSO تعداد ذرات را برابر ۱۰ و تعداد تکرار را برابر ۱۰۰ در نظر گرفتیم همچنین پارامترهای W C1 C2 به ترتیب ۱ ۲ ۲ در نظر گرفته شده است.



شکل ۳: نمودار ROC برای سیستم فازی

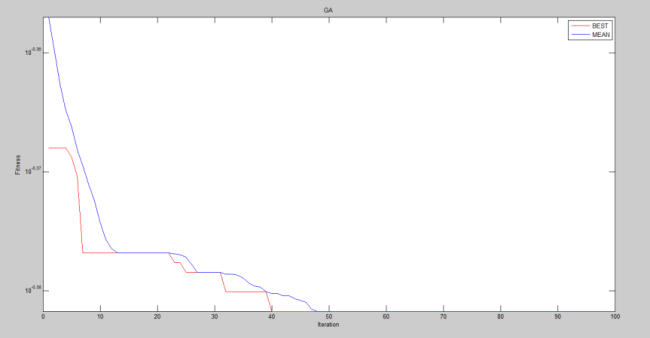


شکل ۴: نمودار کانفیوژن بدست آمده از سیستم فازی بهبود یافته با الگوریتم PSO



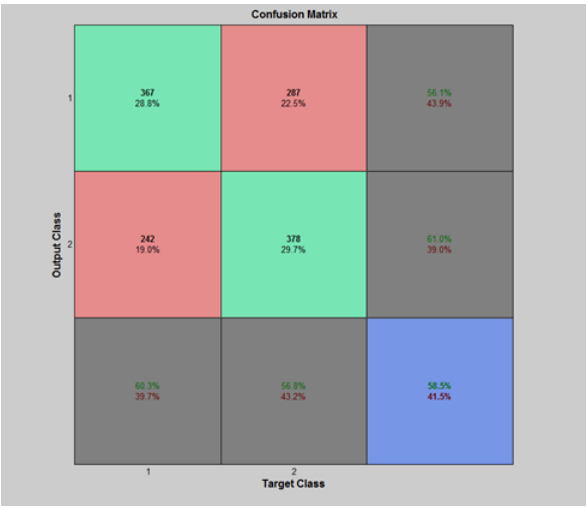
شکل ۵: نمودار ROC بدست امده از سیستم فازی بهبود یافته با الگوریتم PSO

فصلنامه علمی تخصصی ایده های نو در علوم، مهندسی و فناوری ، دوره ۲، شماره۲، تابستان ۱۳۹۷، ص ۷۵-۶۲
ISSN: 2588-3984
http://www.Tajournals.com

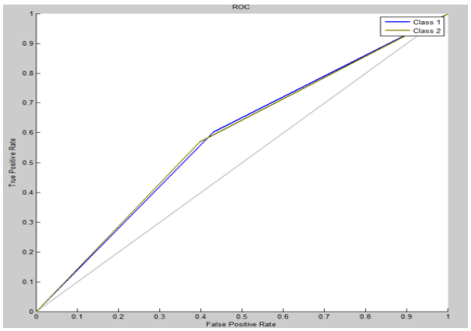


شکل ۶: نمودار همگرایی الگوریتم ژنتیک برای بهبود سیستم فازی.

برای الگوریتم ژنتیک تعداد کروموزوم برابر ۱۰ و تعداد تکرار برابر ۱۰۰ و درصد تقاطع و جهش به ترتیب ۰,۹ و ۰,۱ در نظر گرفته شده است. نتایج حاصل از این الگوریتم به شرح زیر می باشد.



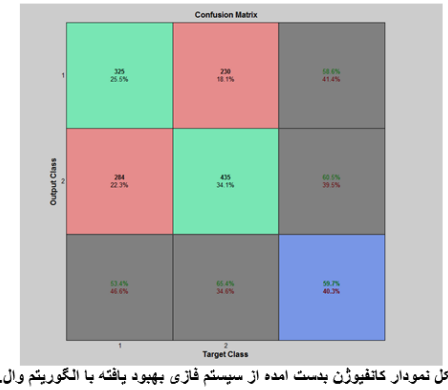
شکل ۷: نمودار کانفیوژن بدست امده از سیستم فازی بهبود یافته با الگوریتم ژنتیک



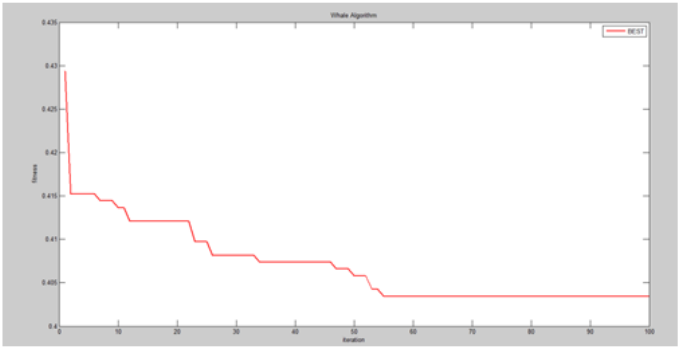
شکل ۸: نمودار ROC بدست امده از سیستم فازی بهبود یافته با الگوریتم ژنتیک.

فصلنامه علمی تخصصی ایده های نو در علوم، مهندسی و فناوری ، دوره ۲، شماره۲، تابستان ۱۳۹۷، ص ۷۵-۶۲
ISSN: 2588-3984
http://www.Tajournals.com

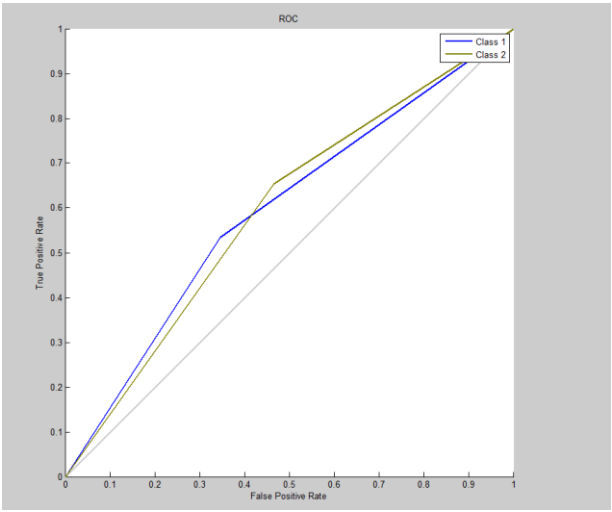
برای الگوریتم وال تعداد وال برابر ۱۰ و تعداد تکرار برابر ۱۰۰ در نظر گرفته شده است. نتایج حاصل از این الگوریتم به شرح زیر می باشد.



شکل نمودار کانفیوژن بدست امده از سیستم فازی بهبود یافته با الگوریتم وال.



شکل ۱: نمودار همگرایی الگوریتم وال برای بهبود سیستم فازی.



شکل ۱۱: نمودار ROC بدست امده از سیستم فازی بهبود یافته با الگوریتم وال.

همچنین نتایج بدست آمده از مقایسات کلی در جدول زیر آورده شده است.

جدول۲: جدول نتایج بدست آمده از ترکیب سیستم فازی با الگوریتم های تکاملی

	وال	PSO	ژنتیک	فازی
نرخ کلاس بندی اشتباه	0.088697	0.410518	0.0949765	**0.10204082**
میزان هزینه کلاس بندی اشتباه	0.10503	0.612025	0.1095724	**0.14622743**
نرمال شده میزان هزینه کلاس بندی اشتباه	0.052515	0.306012	0.0547862	**0.07311371**
حساسیت	0.866864	0.576512	0.8556851	**0.8832**
نرخ اختصاصی	0.961538	0.599719	0.962585	**0.91217257**
دقت	0.911303	0.589482	0.9050235	**0.89795918**
صحت	0.962233	0.53202	0.9638752	**0.90640394**
فراخوانی مجدد	0.866864	0.576512	0.8556851	**0.8832**
معیار ترکیبی F	0.912062	0.553373	0.9065637	**0.89465154**
سازگاری	0.716362	0.242243	0.6873178	**0.77071926**
AUC	0.914201	0.588116	0.9091351	**0.89768629**
معیار توازن	0.902009	0.587952	0.8945802	**0.89666584**
زمان اجرا	98.83984	7.441711	97.995446	**9.51113711**

همان طور که از نتایج پیدا است هر ۳ الگوریتم باعث بهبود سیستم فازی شده اند و در این بین الگوریتم وال بهترین عملکرد و بعد از الگوریتم PSO و ژنتیک قرار دارند. بنابراین تا این مرحله بهترین دقت بدست امده ۰,۹۱۱ است.

ارزش الگوریتم پیشنهادی وقتی دو چندان می‌شود که بدانیم این روش مبتنی بر قانون است. چرا که دسته‌بندی‌های مبتنی بر قانون دارای قابلیت تفسیر بسیار بالایی هستند و کاربری که از آن‌ها استفاده می‌کند به راحتی می‌تواند دانش استخراج شده را تفسیر کند. البته یک دسته بند مبتنی بر قانون در صورتی دارای قابلیت تفسیر مناسب است که تعداد قوانین کم باشد تا کاربر بتواند در زمان مناسب آن‌ها را تفسیر کند. در الگوریتم پیشنهادی تابع برازش به گونه‌ای تعریف شد که دسته‌بند فازی از لحاظ تعداد قوانین نیز فشرده باشد.

عملکرد سیستم ANFIS و سیستم ANFIS توسعه داده شده با الگوریتم های تکاملی

برای سیستم ANFIS در این تحقیق از تابع آماده متلب genfis3 استفاده کرده ایم. که بر مبنای خوشه بندی fcm عمل میکند. در این تابع تعداد خوشه برابر ۲ و نوع سیستم فازی را سوگنو قرار داده ایم. نتایج بدست آمده به شرح زیر است. برای بهبود عملکرد سیستم ANFIS از الگوریتم های فراابتکاری استفاده کردیم بدین شکل بعد ساخته شدن سیستم فازی اولیه ANFIS

ان سیستم را پرورش می دهد و بعد از آن مقادیر توابع عضویت توسط الگوریتم بهبود می یابد به طوری در نتیجه نهایی دارای دقت بهتری نسبت به حالت اولیه باشیم. نتایج بدست امده از الگوریتم ها به شرح زیر است.

جدول ۳: جدول نتایج بدست آمده از ترکیب سیستم ANFIS با الگوریتم های تکاملی

معیارهای ارزیابی	وال	PSO	ژنتیک	ANFIS
نرخ کلاس بندی اشتباه	0.395604	0.401099	0.400314	**0.40423862**
میزان هزینه کلاس بندی اشتباه	0.584929	0.584341	0.5959149	**0.59809002**
نرمال شده میزان هزینه کلاس بندی اشتباه	0.292464	0.29217	0.2979575	**0.29904501**
حساسیت	0.585089	0.572059	0.5857886	**0.57654723**
نرخ اختصاصی	0.622527	0.62963	0.6111908	**0.61363636**
دقت	0.604396	0.598901	0.599686	**0.59576138**
صحت	0.592775	0.638752	0.5550082	**0.58128079**
فراخوانی مجدد	0.585089	0.572059	0.5857886	**0.57654723**
معیار ترکیبی F	0.588907	0.603569	0.5699831	**0.57890433**
سازگاری	0.195439	0.08216	0.2428904	**0.18260784**
AUC	0.603808	0.600844	0.5984897	**0.5950918**
معیار توازن	0.603366	0.599808	0.5982889	**0.59466735**
زمان اجرا	9.212057	9.389817	4.2511936	**0.47401299**

همان طور که از نتایج پیدا است هر ۳ الگوریتم باعث بهبود سیستم ANFIS شده اند و در این بین الگوریتم وال بهترین عملکرد و بعد از الگوریتم PSO و ژنتیک قرار دارند. بنابراین تا این مرحله بهترین دقت بدست امده ۰,۶۰۴ است که از روش قبلی نتیجه بهتری است.

عملکرد سیستم FCM و سیستم FCM توسعه داده شده با الگوریتم های تکاملی

برای سیستم FCM در این تحقیق از تابع آماده متلب fcm استفاده کرده ایم. که بر مبنای خوشه بندی fcm عمل میکند. نتایج بدست آمده به شرح زیر است. برای بهبود عملکرد سیستم FCM از الگوریتم های فراابتکاری استفاده کردیم بدین شکل بعد ساخته شدن سیستم فازی اولیه FCM مقادیر مرکز دستها توسط الگوریتم بهبود می یابد به طوری که در نتیجه نهایی دارای دقت بهتری نسبت به حالت اولیه باشیم. نتایج بدست امده از الگوریتم ها به شرح زیر است. همچنین نتایج بدست آمده از مقایسات کلی در جدول زیر آورده شده است.

جدول ۴: جدول نتایج بدست آمده از ترکیب سیستم FCM با الگوریتم های تکاملی

	وال	PSO	ژنتیک	FCM
نرخ کلاس بندی اشتباه	0.431711	0.475667	0.4395604	**0.51177394**
میزان هزینه کلاس بندی اشتباه	0.647549	0.709869	0.6572104	**0.7566719**
نرمال شده میزان هزینه کلاس بندی اشتباه	0.323775	0.354935	0.3286052	**0.37833595**
حساسیت	0.567506	0.502935	0.5744681	**0.46624804**
نرخ اختصاصی	0.568698	0.538663	0.5555556	**0.51020408**
دقت	0.568289	0.524333	0.5604396	**0.48822606**
صحت	0.407225	0.422003	0.3103448	**0.48768473**
فراخوانی مجدد	0.567506	0.502935	0.5744681	**0.46624804**
معیار ترکیبیF	0.474187	0.458929	0.4029851	**0.47672552**
سازگاری	0.341699	0.170039	0.4263199	**-0.0675039**
AUC	0.568102	0.520799	0.5650118	**0.48822606**
معیار توازن	0.568101	0.520466	0.564909	**0.48775436**
زمان اجرا	5.119123	5.581426	2.63916	**9.99248206**

همان طور که از نتایج پیدا است هر ۳ الگوریتم باعث بهبود سیستم FCM شده اند و در این بین الگوریتم وال بهترین عملکرد و بعد از آن الگوریتم ژنتیک و سپس PSO قرار دارند. بنابراین تا این مرحله بهترین دقت بدست امده ۰,۵۶۸ است که از روش قبلی نتیجه بدتری است.

نتیجه گیری

نتایج بدست آمده نشان می‌دهد دقت روش پیشنهادی در مجموعه داده‌های دیابت سیستم فازی بهبود یافته با الگوریتم وال بهتر یا حداقل قابل رقابت با، سیستم فازی عصبی و FCM توسعه یافته با الگوریتم‌های ژنتیک و PSO و وال است که در نهایت بهترین ترکیبی که در این سه سیستم فازی بدست آمد ترکیب سیستم فازی بهبود یافته با الگوریتم وال بود که دقت بدست آمده نزدیک ۹۲ درصد است. همچنین با توجه به این که روش پیشنهادی مبتنی بر قانون است، بنابراین الگوریتم دارای قابلیت تفسیر مناسبی است.

با توجه به واقعیات موجود دیابت یکی از بیماری‌های رایج در جوامع امروزی تبدیل شده است که دارای عوارض خطرناکی می‌باشد. این بیماری اگر چه گونه‌ای از بیماری‌های قلبی محسوب نمی‌شود ولی اغلب سبب بیماری‌های قلبی می‌شود. تشخیص بیماری دیابت و یا آگاهی یافتن از احتمال بالای ابتلا به این بیماری همواره کار آسانی نخواهد بود. چرا که این بیماری علائم

متعددی را بروز می‌دهد که بعضی از این علائم در سایر بیماری‌ها نیز وجود دارند. بنابراین پزشک برای اتخاذ یک تصمیم مناسب، باید نتیجه‌ی آزمایش‌های بیمار و تصمیم‌های که در گذشته برای بیماران با وضعیت مشابه گرفته است، را بررسی کند. در این تحقیق روشی را برای واکشی دانش مطرح کردیم. دانش مورد نظر توسط فرایند داده‌کاوی و به کمک روش دسته‌بندی به دست آمده است. دانش حاصل شده توسط قوانین اگر – آنگاه فازی نمایش داده شد. دسته‌بند طراحی شده مبتنی بر سیستم فازی توسط معیارهای دقت و قابلیت تفسیر ارزیابی شد.برای استخراج توابع عضویت و قوانین فازی به طور هم‌زمان، یک الگوریتم مبتنی بر بهینه‌سازی وال پیشنهاد گردیده است. این الگوریتم دارای ویژگی‌هایی است که آن را از روش‌های استخراج دانش مبتنی بر بهینه‌سازی متمایز می‌کند. در الگوریتم پیشنهادی هر وال یک سیستم فازی را شکل می‌دهد به طوری که شامل پارامترهای توابع عضویت و قوانین فازی می‌شود. والهای دیگر سیستم‌های فازی متفاوتی را ایجاد می‌نمایند. در هر دور هر یک از والها با توجه به بهترین سیستم فازی طراحی شده توسط هر وال و همین طور کل والها سعی در ایجاد ویرایش در سیستم دسته‌بند فازی را دارند. نتایج بدست آمده نشان دادند در بین سیستم های پیش بینی مورد استفاده که فازی و فازی عصبی و FCM است بهترین عملکرد را سیستم فازی دارا است. همچنین به یقین رسیدیم که ترکیب با الگوریتم ها باعث بهبود عملکرد سیستم های پیش بینی می شود و در این بین بهترین عملکرد مربوط به الگوریتم وال است که قابل رقابت و یا حتی بهتر است.

پیشنهادات آتی

برای کارهای آتی در این زمینه می‌توان پیشنهادات زیر را در نظر گرفت:

- ✓ استفاده از داده های فازی
- ✓ استفاده از داده های خاکستری
- ✓ افزایش گستره داده ها به استان یا کشور
- ✓ استفاده از روش های ترکیبی سیستم های پیش بینی
- ✓ استفاده از الگوریتم های تکاملی دیگر
- ✓ چند هدفه سازی مساله
- ✓ اضافه کردن‌مبحث‌انتخاب ویژگی برای انتخاب کاراترین ویژگی ها
- ✓ تلاش برای حذف پارامترهایی که باید توسط کاربر تنظیم شوند. چرا که تنظیم این پارامترها می‌تواند از کاربر وقت زیادی بگیرد. به علاوه امکان اشتباه کاربر وجود دارد. می‌توان از یک روال خودکار برای بدست آوردن مقدار بهینه این پارامترها استفاده کرد.
- ✓ ارائه یک روش برای یادگیری قوانین غیر ساخت یافته. قوانین غیر ساخت یافته قوانینی هستند که از عملگرهای منطقی گوناگون مانند And و Or استفاده می‌کنند. استفاده از این عملگرها می‌تواند موجب بالا بردن قابلیت تفسیر قوانین گردد.
- ✓ در نظر گرفتن معیارهای مربوط به سادگی و قابلیت تفسیر مربوط به پایگاه قوانین به دست آمده برای تشخیص بیماری طوری که باعث کاهش دقت مجموعه قوانین نشود و مجموعه قوانین با تعداد کمتری را فراهم آورد.

منابع

1- O. Maimon and L. Rokach, "Introduction to knowledge discovery and data mining," in Data Mining and Knowledge Discovery Handbook, ed: Springer, 2010, pp. 1-15.
2- H. C. Koh and G. Tan, "Data mining applications in healthcare," Journal of Healthcare Information Management—Vol, vol. 19, p. 65, 2011.

ISSN: 2588-3984
http://www.Tajournals.com

3- R. Bellazzi and B. Zupan, (2008) "Predictive data mining in clinical medicine: current issues and guidelines," international journal of medical informatics, vol. 77, pp. 81-97.
4- Su, C.T., et. Al., (2006) " Data Mining for the Diagnosis of Type II Diabetes from Three Dimensional Body Surface Anthropometrical Scanning Data", Computers and Mathematics with Applications,1075-1092.
5- Barakat, N. H., Bradley, A. P. and Barakat, M. N. H. , (2010) "Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus". IEEE Transactions on Information Technology in Biomedicine, VOL. 14, NO. 4.
6- Worachartcheewan, A., et. Al., (2010) "Identification of metabolic syndrome using decision tree analysis", diabetes research and clinical practice 90. 1 5– 1 8.
7- Patil, B. M., Joshi, R.C., Toshniwal, D., (2010) "Association rule for classification of type -2 diabetic patients", Second International Conference on Machine Learning and Computing.
8- Saniee Abadeh, M., (2011) "A fuzzy classification system based on Ant Colony Optimization for diabetes disease diagnosis", Expert Systems with Applications, 38, 14650–14659.
9- Bagherzadeh Khiabani, F., Akhavan Niaki, S.T, (2013)" An Ensemble Model for Prediction of Diabetes Type 2", The Seventh Iran Data Mining Conference, Tehran, Iran.
10- Smith, J. W, Everhart, J. E, Dickson, W C, Knowler, W. C, and Johannes, R.S, (1998) "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus". In Proceedings of 12th Symposium on Computer Applications in Medical, Care, R. A. Greenes, Ed. IEEE Computer Society Press, pp. 261–265.