کشف دانش موجود در متن دفترچه های بیمه با کمک مدل ترکیبی فازی ژنتیک

زینب شیرکول[۱]، احسان امیری [۲,*]

۱- گروه کامپیوتر، دانشکده مهندسی ، واحد شیراز، دانشگاه آزاد اسلامی، شیراز، ایران
shirkool.t@yahoo.com
۲- مدرس مدعو گروه کامپیوتر، دانشگاه جهرم، جهرم، ایران
[*]نویسنده مسئول: e.e.amiri@gmail.com

خلاصه

به دنبال پیشرفت‌های حاصله در زمینهٔ ساخت پردازنده‌های پرسرعت، اثر محدودکننده‌ی نحوه‌ی ورود داده‌ها به داخل کامپیوتر بر سرعت انتقال اطلاعات، بیش از پیش نمود یافته است. با به‌کارگیری دستگاه‌های پردازشی علاوه بر دستیابی به‌سرعت بالاتر در مرحله‌ی داده‌خواهی، امکان استفاده از قابلیت‌های پیش‌پردازشی آن‌ها و تغییر فرمت داده‌ها نیز فراهم می‌آید. با توجه به اهمیت موضوع و کاری که در این زمینه انجام‌گرفته است، نیاز به کشف دانش از ویژگی های موجود با کمک انتخاب ویژگی مناسب برای طبقه‌بندی متون به‌خوبی احساس می‌شود. در این مقاله از الگوریتم ژنتیک و منطق فازی برای ارائه روشی جهت طبقه‌بندی متون موجود در دفترچه‌های بیمه استفاده شده است. این سیستم بر پایه‌ی چندین مرحله شکل‌گرفته است. این مراحل شامل فاز یادگیری که مجموعه‌ای از متون آموزشی را برای استخراج ویژگی‌های دسته‌ها بررسی می‌کند تا خصوصیات هر دسته باشد؛ فاز تست سیستم، برای طبقه‌بندی متون دسته‌بندی نشده بکار می‌روند. دقت روش پیشنهادی روی مجموعه جمع‌آوری‌شده‌ای از چندین دفترچه بیماران مورد آزمایش قرار گرفته است. نتایج حاصله نشان می‌دهد که طبقه‌بندی پیشنهادی دقتی در حدود ۹۸ درصد دارد.

کلمات کلیدی: کشف دانش، دسته‌بندی متن، الگوریتم ژنتیک، منطق فازی.

۱. مقدمه

کشف دانش به معنای کشف و یافتن دانش جدید و نوآورانه است. این فرآیند شامل تحقیق، آزمایش، تجربه و تحلیل داده‌ها به منظور یافتن راه‌حل‌های جدید و بهبود فرآیندهای موجود است. در این فرآیند، افراد با استفاده از تخصص و تجربه خود سعی می‌کنند به راه‌حل‌هایی نوآورانه و کارآمد در برابر چالش‌های مختلف دست یابند. کشف دانش یک فرآیند پویا و پیچیده است که نیازمند مشارکت فردی و گروهی برای پیدا کردن راه‌حل‌های خلاقانه و نوآورانه است. این فرآیند در انواع زمینه‌های علمی، فنی، اجتماعی و اقتصادی انجام می‌شود.

[*] **Corresponding author:** استاد مدعو دانشگاه جهرم
Email: e.e.amiri@gmail.com

یکی از مهمترین مواردی که در کشف دانش بایستی بدان توجه کرد، کشف الگوهایی است که در هر مسئله وجود دارد. داروهای موجود در دفترچه های بیمه افراد که به صورت الکترونیکی ارائه می شوند، حاوی الگوی مشخصی از میزان سلامتی فرد هستند. کشف این الگو با کمک روش های هوشمند باعث بهبود نتایجی شده است. از میان روش های هوشمند موجود، الگوریتم‌های تکاملی در مقایسه با سایر الگوریتم‌های بهینه‌سازی برتری‌هایی دارند که موجب شده است به طور گسترده مورداستفاده قرار بگیرند. به‌عنوان مثال، این الگوریتم‌ها، نیاز به معرفی کامل مسئله نداشته و تنها با داشتن اطلاعات جزئی در مورد تعریف مسئله می‌توانند کار خود را انجام دهند. همچنین محدودیتی در مورد تابع شایستگی ندارند و لزومی ندارد که این تابع تعریف خاصی باشد.

علاوه بر این موارد، چون این الگوریتم‌ها دارای جمعیتی از موجودات هستند، روی بخش‌های مختلفی از جمعیت به طور موازی کار می‌کنند پس احتمال کمتری برای قرارگرفتن در بهینه‌های محلی دارند. این قابلیت این دست از الگوریتم‌ها اجازه می‌دهد که کار بهینه‌سازی را به طور موازی روی چندین بخش جمعیت انجام شود. به‌عنوان نتیجه این ویژگی گفته می‌شود که الگوریتم‌های تکاملی روی جمعیت‌های مختلف شلوغ خوب عمل می‌کند. این‌گونه از جمعیت‌ها دارای تعداد زیادی بهینه‌های محلی هستند. الگوریتم‌های تکاملی در این بهینه‌های محلی گیر نمی‌کنند و معمولاً می‌توانند درنهایت به بهترین جواب برسند [۵].

در زمان خواندن اطلاعات دارویی توسط پرستاران و مسئولین داروخانه‌ها معمولاً اشتباهاتی صورت می‌گیرد که در بعضی مواقع این اشتباهات جبران‌ناپذیر هستند. به همین سبب داشتن سیستمی که بتواند انسان را از این اشتباهات جدا کند می‌تواند به روند بهبودی سریع‌تر بیماران کمک کند. سیستمی که در اینجا معرفی می‌شود یک سیستم کشف دانش برای جلوگیری از اشتباه دارویی است. همان‌طور که از عنوان تحقیق مشخص است این سیستم رفتاری مشابه با یک سیستم بازشناسی نوری حروف ساده دارد ولی انجام این کار با دشواری‌هایی همراه است.

در این روش، در آغاز کار تعدادی از جواب‌ها به‌صورت تصادفی حدس زده‌شده، سپس تابع هدف برای هر یک از این جواب‌ها محاسبه و نسل اول ایجاد خواهد شد. اگر هیچ‌یک از معیارهای خاتمه بهینه‌سازی دیده نشوند، ایجاد نسل جدید شروع خواهد شد. اعضا برحسب میزان شایستگی‌شان برای تولید جواب‌ها انتخاب می‌شوند. این افراد به‌عنوان والدین و با بازترکیب جواب‌ها را تولید می‌نمایند. سپس تمامی جواب‌ها با یک مقدار معینی از احتمال، تغییر ژنتیکی می‌یابند. اکنون میزان شایستگی (برازندگی) جواب‌ها تعیین و در اجتماع جایگزین والدین شده و جواب‌های جدید را ایجاد می‌نمایند. این چرخه آن‌قدر تکرار می‌شود تا یکی از معیارهای پایان بهینه‌سازی کسب شود.

روش‌هایی که در مقالات دیگر آمده است عموماً به دسته‌بندی متن پرداخته‌اند ولی با الکترونیکی شدن تجویز دارو، عملا مبحث کشف الگو نمود بیشتری یافته است که دیگران به آن نپرداخته‌اند درحالی‌که می‌توان با انتخاب ویژگی‌های بهتر، دسته‌بندی بهتری انجام داد. یکی از موارد موجود دسته‌بندی متن را بر اساس درخت تصمیم انجام داده است و با توجه به رویه‌ای که درخت تصمیم دارد، مجبور به انتخاب تمام ویژگی‌ها جهت دسته‌بندی متن شده است درحالی‌که بسیاری از این ویژگی‌ها کاربردی در دسته‌بندی ندارند.

۲. مروری بر مطالب مرتبط

جوس لیوس و همکاران در سال ۲۰۱۸ بر روی تشخیص الگوی شبکه رنگ‌دانه در تصاویر درماتوسکوپی بر اساس طبقه‌بندی فازی پیکسل‌ها کارکرده‌اند. این کار سه مجموعه مربوط به فازی را ایجاد می‌کند و درنتیجه، سه تصویر احتمالی که آن‌ها را طراحی می‌کنند تولید می‌شوند. استخراج بعد از ویژگی‌های مختلف محاسبه‌شده و از ترکیبی از ماسک‌های

تصویر استخراج‌شده و سپس از تصاویر احتمالی مربوط به کاهش آلفا به دست می‌آید. این روش بر روی یک پایگاه داده از ۸۷۵ تصویر آزمایش‌شده است که تا حد زیادی بزرگ‌ترین مورداستفاده در حالت‌های پیشرفته برای شناسایی شبکه رنگ‌دانه استخراج‌شده از اطلس عمومی درماتوسکوپی می‌باشد. به دست آوردن نتایج AUC نشان از ۸۸ ٪ دقت و ۹۰,۷۱ ٪ حساسیت و ویژگی خاص ۸۳,۴۴ ٪ دارد[۱۱].

اسما آزرال و همکاران در سال ۲۰۱۷ بر روی تشخیص سریع لبه تصویر بر اساس فایبر شیدر ویولت و اوسو آستانه کارکرده‌اند. تشخیص لبه یک مرحله بحرانی در بسیاری از سیستم‌های دید در کامپیوتر است، مانند تقسیم‌بندی تصویر و تشخیص شیء. همان‌طور که مشکل است لبه‌های تصویر را با دقت و پیچیدگی کم تشخیص دهیم، مناسب است که روش‌های جدیدی را برای تشخیص لبه پیدا کنیم. در این مقاله، از حد آستانه و اتزو برای تشخیص لبه‌ها در چند مقیاس با پیچیدگی کم استفاده شده است، زیرا ضرایب شدت این موجک بر روی نقاط لبه و تنها شامل عملیات محاسباتی هستند. اول، تصویر با استفاده از فیلتر دوجانبه بسته به تخمین نویز، نرم شده است. دوم، ضرایب افراطی FSW بر اساس آستانه Otsu انتخاب می‌شوند. درنهایت، نقاط لبه با استفاده از یک الگوریتم لینک لبه پیش‌بینی برای لبه‌های تصویر می‌شود. اثربخشی روش پیشنهادشده توسط نتایج تجربی پشتیبانی می‌شود که نشان می‌دهد این روش سریع‌تر از بسیاری از روش‌های رقابتی است و می‌تواند در برنامه‌های در حال اجرا مورداستفاده قرار گیرد [۱۲].

ژو در سال ۲۰۲۳ بسیاری از متدهای میان رشته ای کشف دانش و داده کاوی را معرفی کرده اند. حوزه میان رشته‌ای کشف دانش و داده‌کاوی از نیاز به داده‌های بزرگ که نیازمند روش‌های تحلیلی جدید فراتر از رویکردهای آماری سنتی برای کشف دانش جدید از داده‌کاوی است، پدید آمد. این رویکرد نوظهور یک فرآیند تحقیق دیالکتیکی است که هم قیاسی و هم استقرایی است. رویکرد داده کاوی به طور خودکار یا نیمه خودکار تعداد بیشتری از پیش‌بینی‌کننده‌های مشترک، تعاملی و مستقل را برای رسیدگی به ناهمگونی علّی و بهبود پیش‌بینی در نظر می‌گیرد. به جای به چالش کشیدن رویکرد ساخت مدل مرسوم، نقش مکمل مهمی در بهبود تناسب مدل، آشکارسازی الگوهای پنهان معتبر و قابل توجه در داده‌ها، شناسایی اثرات غیرخطی و غیرافزودنی، ارائه بینشی در مورد تحولات داده‌ها، روش‌ها و تئوری دارد. و اکتشافات علمی را غنی می کند. زمانی که ساختار مدل صریح نامشخص است و دستیابی به الگوریتم‌هایی با عملکرد خوب دشوار است، یادگیری ماشینی با یادگیری و بهبود از داده‌ها، مدل‌ها و الگوریتم‌ها را می‌سازد. جدیدترین پیشرفت، ترکیب این پارادایم جدید مدل‌سازی پیش‌بینی‌کننده با رویکرد کلاسیک رگرسیون‌های تخمین پارامتر برای تولید مدل‌های بهبودیافته است که توضیح و پیش‌بینی را ترکیب می‌کنند[۱۵].

آلوز و همکاران در سال ۲۰۲۲ بر روی کشف الگوی مناسب کشف دانش از داروهای بیماران کار کرده اند. ثابت شده است که خط لوله کشف داروی مرسوم برای بیماری های نادر ناپایدار است. در اینجا، در مورد پیشرفت‌های اخیر در استخراج دانش زیست پزشکی که برای کشف درمان‌های بیماری‌های نادر استفاده می‌شود، بحث شده است. داده‌های شیمی‌شناسی فعلی مرتبط با بیماری‌های نادر خلاصه شده و دیدگاهی در مورد اثربخشی یادگیری ماشین (ML) و استخراج نمودار دانش زیست‌پزشکی در کشف داروی بیماری‌های نادر ارائه شده است. قدرت این روش‌ها با استفاده از مطالعه موردی کوردوما نشان داده شده است[۱۶].

الکساندر و همکاران در سال ۲۰۲۳ بر روی کشف دارو تمرکز کرده اند. هوش مصنوعی (AI) این پتانسیل را دارد که فرآیند کشف دارو را متحول کند و کارایی، دقت و سرعت را بهبود بخشد. با این حال، کاربرد موفقیت‌آمیز هوش مصنوعی به در دسترس بودن داده‌های با کیفیت بالا، رسیدگی به نگرانی‌های اخلاقی، و شناخت محدودیت‌های رویکردهای مبتنی بر هوش مصنوعی بستگی دارد. در این مقاله، مزایا، چالش ها و معایب هوش مصنوعی در این زمینه بررسی شده و راهبردها و رویکردهای ممکن برای غلبه بر موانع موجود ارائه شده است. استفاده از افزایش داده ها، هوش مصنوعی قابل توضیح، و

ادغام هوش مصنوعی با روش های تجربی سنتی، و همچنین مزایای بالقوه هوش مصنوعی در تحقیقات دارویی نیز مورد بحث قرار می گیرد. به طور کلی، این بررسی پتانسیل هوش مصنوعی در کشف دارو را برجسته می کند و بینش هایی را در مورد چالش ها و فرصت های تحقق پتانسیل آن در این زمینه ارائه می دهد. نکته از نویسندگان انسانی: این مقاله برای آزمایش توانایی ChatGPT، یک ربات چت مبتنی بر مدل زبان GPT-۳،۵، در زمینه کمک به نویسندگان انسانی در نوشتن مقالات مروری ایجاد شده است. متن تولید شده توسط هوش مصنوعی به عنوان نقطه شروع استفاده شد و توانایی آن برای تولید خودکار محتوا ارزیابی شد. پس از انجام یک بررسی کامل، نویسندگان انسانی عملاً نسخه خطی را بازنویسی کردند و تلاش کردند تا تعادلی بین پیشنهاد اصلی و معیارهای علمی حفظ کنند[۱۷].

رکمالشمی و همکاران در سال ۲۰۱۷ بر روی ارزیابی جهت لبه تصویر از طریق منطق فازی کارکرده‌اند. در این مقاله، یک برآورد پیش‌فرض جهت لبه تصویر لبه با استفاده از رویکرد منطق فازی پیشنهادشده است. ایده پشت جهت برآورد این است که برای طراحی تصویربرداری لبه به‌کاررفته برای افزایش تصویر فوق‌العاده رزولوشن. محاسبه لبه تصویر جهت‌گیری پیکسل بر اساس انواع مختلف شیب در زوایای مختلف از پیکسل لبه خاص اندازه‌گیری می‌شود. گرایش از لبه پیکسل به ۹ زاویه طبقه‌بندی‌شده برای اطمینان از کیفیت برآورد جهت. مزیت فاز پیشنهادشده بر اساس روش این است که محاسبه جهت‌گیری در زبان انسانی با کمک قوانین فازی ساده انجام می‌شود. استفاده از عضویت ساده تابع و قوانین کمک می‌کند تا طراحی چارچوب برآورد کاری مناسب را طراحی کنند. نتایج تجربی نشان می‌دهد که مبتنی بر فازی است ارزیابی جهت‌گیری حتی اگر تصویر توسط نویز تأثیر می‌گذارد، بهتر عمل می‌کند [۱۹].

ایزابل بلوچ در سال ۲۰۱۵ بر روی مجموعه‌های فازی برای پردازش و درک تصویر کار کرده است. این مقاله، یک مرور کلی از مدل‌ها و روش‌های مبتنی بر مجموعه‌های فازی برای پردازش تصویر و درک تصویر را پیشنهاد می‌دهد[۲۰].

۳. روش تحقیق

الگوریتم ژنتیک، الهامی از علم ژنتیک و نظریه تکامل داروین است و بر اساس بقای برترین‌ها یا انتخاب طبیعی استوار است. یک کاربرد متداول الگوریتم ژنتیک، استفاده از آن به‌عنوان تابع بهینه‌سازی است. الگوریتم ژنتیک ابزار سودمندی در بازشناسی الگو، انتخاب ویژگی، درک تصویر و یادگیری ماشینی است [۱۹-۱۴]. در الگوریتم‌های ژنتیکی، نحوه تکامل ژنتیکی موجودات زنده شبیه‌سازی می‌شود.

الگوریتم‌های ژنتیکی را می‌توان یک روش بهینه‌سازی تصادفی جهت‌دار دانست که به‌تدریج به سمت نقطه بهینه حرکت می‌کند. در مورد ویژگی‌های الگوریتم ژنتیک در مقایسه با دیگر روش‌های بهینه‌سازی می‌توان گفت که الگوریتمی است که بدون داشتن هیچ‌گونه اطلاعی از مسئله و هیچ‌گونه محدودیتی بر نوع متغیرهای آن برای هرگونه مسئله‌ای قابل اعمال است و دارای کارایی اثبات‌شده‌ای در یافتن بهینه کلی است. توانایی روش در حل مسائل پیچیده، است که روش‌های کلاسیک نیستند و یا دریافتن بهینه کلی قابل‌اطمینان نیستند[۱۱].

به‌طورکلی، الگوریتم‌های ژنتیکی از اجزاء زیر تشکیل می‌شوند:

- کروموزوم : که یک جواب را شامل می‌شود.
- جمعیت : مجموعه کروموزوم‌ها و در نتیجه مجموع جواب‌ها را شامل می‌شود.
- تابع برازندگی : تابع اصلی در حل مسئله مورد نظر می‌باشد.
- عملگرهای ژنتیکی: عملگرهای حل مسئله را شامل می‌شوند.

- عملگر انتخاب: این عملگر از بین راه حل‌های موجود در یک جمعیت، تعدادی راه‌حل را برای تولیدمثل انتخاب می‌کند.
- عملگر آمیزش: عملگر آمیزش بر روی یک زوج راه‌حل از نسل مولد عمل کرده و یک زوج راه‌حل جدید تولید می‌کند.
- عملگر جهش: این عملگر یک ژن از یک راه حل را به طور تصادفی انتخاب نموده و سپس محتوای آن ژن را تغییر می‌دهد.

پس از اتمام عمل جهش، راه‌حل‌های تولیدشده به‌عنوان نسل جدید شناخته‌شده و برای دور بعد اجرای الگوریتم ارسال می‌شوند.

1 0 1 0 **1** 0 1 1 1 0

شکل ۱ – نمونه کروموزوم

در شکل ۲ یک الگوریتم ژنتیکی مبتنی بر منطق فازی، و در شکل ۳ نمودار گردشی الگوریتم ژنتیکی پیشنهادی نشان‌داده‌شده است.

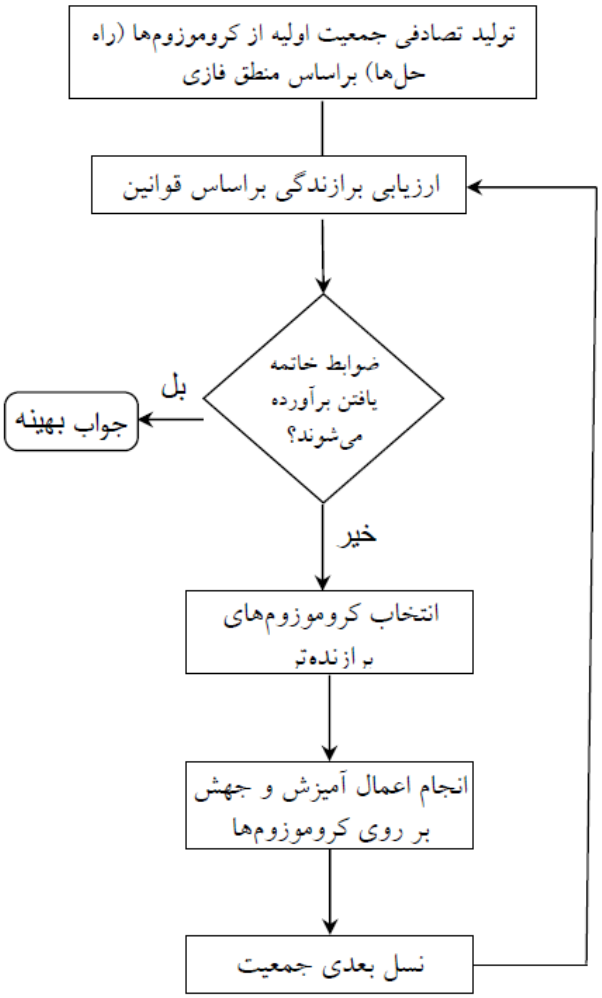
قبل از این‌که یک الگوریتم ژنتیکی بتواند اجرا شود، ابتدا باید نحوه نمایش مناسبی برای حل مسئله موردنظر پیدا شود. همچنین یک تابع برازندگی نیز باید ابداع شود تا به هر راه‌حل کدگذاری شده ارزشی را نسبت دهد. در طی اجرا، والدین برای تولیدمثل انتخاب می‌شوند و با استفاده از عملگرهای آمیزش و جهش باهم ترکیب می‌شوند تا فرزندان جدیدی تولید کنند. این فرایند چندین بار تکرار می‌شود تا نسل بعدی جمعیت تولید شود. سپس این جمعیت بررسی می‌شود و درصورتی‌که ضوابط همگرایی برآورده شوند، فرایند فوق خاتمه می‌یابد.

```
BEGIN  /* genetic algorithm */
    generate initial population logic
    compute fitness of each individual using rule function
    WHILE NOT finished DO
    BEGIN  /* produce new generation */
        FOR population_size / 2 DO
        BEGIN  /* reproductive cycle */
            select two individuals from old generation for mating
                /* biassed in favour of the fitter ones */
            recombine the two individuals to give two offspring
            compute fitness of the two offspring using fuzzy logic
            insert offspring in new generation
        END
        IF population has converged THEN
            finished := TRUE
    END
END
```

شکل ۲ – الگوریتم ترکیبی ژنتیک فازی

مطابق الگوریتم شکل ۲، مرحله اولیه مقداردهی تصادفی براساس منطق فازی است. مقادیر اولیه صفر و یک های منطق فازی هستند که با یک تعلق عضویت تصادفی شکل گرفته اند. در این مرحله مقادیر به صورت کروموزوم های صفر و یک نوشته خواهند شد و هر کروموزوم از ده ژن برای ده دسته از انواع داروهای موجود می باشد که با مقادیر صفر و یک

تشکیل می شود. هر کدام از مقادیر صفر و یک موجود، یک عدد تصادفی دریافت خواهند کرد که نشان دهنده عضویت آن ژن به هر کدام از ده دسته موجود می باشد. در صورتی که مقدار یک ثبت شده باشد، آن دارو با ضریب مشخص انتخاب می شود، در صورتی که مقدار صفر ثبت شده باشد، آن دارو انتخاب نمی شود. دیاگرام زیر، روند الگوریتم ژنتیک را بهبود انتخاب نشان می دهد.



شکل ۳ – دیاگرام ترکیبی ژنتیک فازی

در مراحل بعدی عملیات براساس قوانین فازی از میان داروها انتخاب صورت می گیرد. مراحل بعد شامل اعمال عملیات الگوریتم ژنتیک بر روی کروموزوم ها است. عملیات ترکیب تک نقطه ای، جهش تصادفی ۰,۳ و انتخاب تورمنت برای عملیات های ژنتیک درنظر گرفته شده است.

۴. ارزیابی دسته‌بندی متون

ارزیابی نتایج حاصل از مدل‌سازی با کمک مدل ارزیابی دقت صورت می‌گیرد. نتایج ارزیابی باعث بهبود مدل می‌شود و مدل را قابل استفاده می‌نماید. برای بررسی دقت مدل، ابتدا لازم است داده‌های موجود را به سه بخش آموزش، آزمایش و

اعتبارسنجی تقسیم کنیم. شاخص‌های مختلفی برای ارزیابی دقت روش‌های دسته‌بندی وجود دارد که می‌توان از آنها سود برد.

مطالب متعددی درون دفترچه‌ها وجود دارد که عبارتند از نسخه، مهر، بارکد، پس‌زمینه و مطالب چاپی. هر دفترچه و حتی هر نسخه نسبت به هرکدام از این موارد متفاوت است به همین سبب کار را برای تشخیص جامع سخت می‌کند.

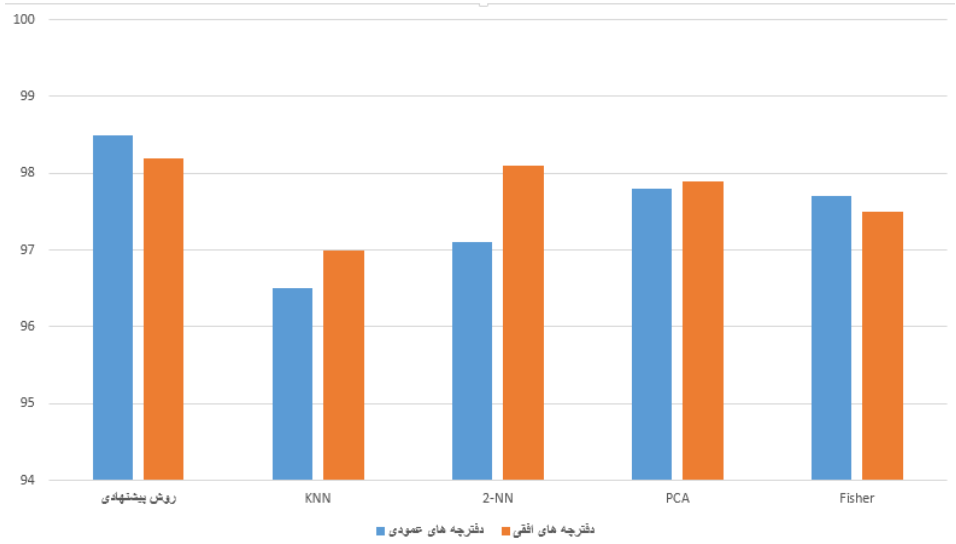
برای آنکه بتوانیم نسخه‌ها را مورد تشخیص قرار دهیم و درعین‌حال خود را درگیر تعدد موارد موجود نکنیم انتخاب آن‌ها را به‌صورت دستی انجام دادیم. از میان دفترچه‌های موجود دو نوع دفترچه و از میان نسخه‌های پزشکان مختلف، یک پزشک را انتخاب نمودیم.

روش‌های معرفی‌شده را با KNN، ۲□□، ویولت و فیشر مورد بررسی قرار داده‌ایم. بانک اطلاعاتی مورداستفاده جهت دسته‌بندی متون در این گزارش، تصاویر نسخه‌های پزشکی موجود در دفترچه‌های بیمه است که شامل اعداد و نوشته‌های داروها است. برای آنکه نتیجه به‌دست‌آمده با دیگر روش‌ها مورد بررسی قرار گیرد، نتایج حاصله با دیگر بررسی‌ها مورد مقایسه قرار گرفته شد که جدول زیر نتیجه این مقایسه‌ها را نشان می‌دهد.

جدول ۱ – درصد مقایسه دسته‌بندی با دیگر روش‌ها

نحوه قرارگیری تصویر	روش پیشنهادی	KNN	۲-NN	PCA	Fisher
دفترچه‌های عمودی	۹۸.۰۵	۹۶.۵	۹۷.۱	۹۷.۸	۹۷.۷
دفترچه‌های افقی	۹۸.۲	۹۷.۰	۹۸.۱	۹۷.۹	۹۷.۵

نمودار زیر درصد دسته‌بندی را بر اساس نوع دفترچه نشان می‌دهد.



شکل ۴ – تفاوت تشخیص با دیگر روش‌ها

مشاهده می‌شود که روش پیشنهادی دسته‌بندی نتایج بسیار خوبی را بر روی نسخه موجود در دفترچه بیماران ارائه می‌کند به‌گونه‌ای که در حالت عمودی و افقی دقت آن از روش‌هایی مانند ۲-NN و PCA بهتر شده است. با بررسی بیشتر می‌توان به نتایج دیگری بر روی متون مختلف دیگر نیز رسید. به نظر می‌رسد که روش پیشنهادی انتخاب ویژگی‌های

بهتری نسبت به روش‌های ریاضی داشته باشد. این روش پیشنهادی دسته‌بندی متن را بر اساس انتخاب ویژگی‌های بهتر انجام داد.

۵. نتیجه‌گیری و پیشنهادات

نتایج دسته‌بندی نشان می‌دهد که دسته‌بندی صورت گرفته برای دفترچه‌های عادی دقتی در حدود ۹۸ درصد را به دست آورده است و برای دفترچه‌های افقی دقتی در حدود ۹۸ درصد را به دست آورده است. این نتایج نشان‌دهنده قدرت روش پیشنهادی است.

در این تحقیق معرفی بهترین الگو در زمینه‌ی انتخاب ویژگی جهت دسته‌بندی متن است که یافتن ویژگی با کمک روش‌های بهینه در مقالات مختلف بررسی‌شده که در این گزارش این ویژگی‌ها را برای مسئله دسته‌بندی متن انجام شده است. استفاده از این روش در دسته‌بندی متن در سطح مقالات بین‌المللی بسیار کم بوده و به نظر می‌رسد که می‌تواند مدل خوبی برای حل این مسئله باشد.

۶. مراجع

۱. Garcia-Arroyo, J. L., & Garcia-Zapirain, B. (۲۰۱۸). Recognition of pigment network pattern in dermoscopy images based on fuzzy رده‌بنـدی of pixels. Computer methods and programs in biomedicine, ۱۵۳, ٦۱-٦۹.

۲. Azeroual, A., & Afdel, K. (۲۰۱۷). Fast Image Edge Detection based on Faber Schauder Wavelet and Otsu Threshold. Heliyon, ۳(۱۲), e۰۰٤۸٥.

۳. El Hatri, C., & Boumhidi, J. (۲۰۱۷). Fuzzy deep learning based urban traffic incident detection. Cognitive Systems Research.

٤. Dadgostar, H., & Afsari, F. (۲۰۱٦). Image steganography based on interval-valued intuitionistic fuzzy edge detection and modified LSB. Journal of information security and applications, ۳۰, ۹٤-۱۰٤.

٥. Gonzalez, C. I., Melin, P., Castro, J. R., Castillo, O., & Mendoza, O. (۲۰۱٦). Optimization of interval type-۲ fuzzy systems for image edge detection. Applied Soft Computing, ٤۷, ٦۳۱-٦٤۳.

٦. Karakış, R., Güler, İ., Capraz, I., & Bilir, E. (۲۰۱٥). A novel fuzzy logic-based image steganography method to ensure medical data security. Computers in biology and medicine, ٦۷, ۱۷۲-۱۸۳.

۷. Mehranfar, A., Ghadiri, N., Kouhsar, M., & Golshani, A. (۲۰۱۷). A Type-۲ fuzzy data fusion approach for building reliable weighted protein interaction networks with application in protein complex detection. Computers in biology and medicine, ۸۸, ۱۸-۳۱.

٨. Ouma, Y. O., & Hahn, M. (٢٠١٧). Pothole detection on asphalt pavements from ٢D-colour pothole images using fuzzy c-means clustering and morphological reconstruction. Automation in Construction, ٨٣, ١٩٦-٢١١.

٩. Reshmalakshmi, C., & Sasikumar, M. (٢٠١٧). Image Edge Orientation Estimation via Fuzzy Logic. Materials Today: Proceedings, ٤(٢), ٤٢٧٤-٤٢٨٢.

١٠. Uguz, S., Sahin, U., & Sahin, F. (٢٠١٥). Edge detection with fuzzy cellular automata transition function optimized by PSO. Computers & Electrical Engineering, ٤٣, ١٨٠-١٩٢.

١١. Biju, V. G., & Mythili, P. (٢٠١٥). Fuzzy clustering algorithms for cDNA microarray image spots segmentation. Procedia Computer Science, ٤٦, ٤١٧-٤٢٤.

١٢. Liang, L. R., & Looney, C. G. (٢٠٠٣). Competitive fuzzy edge detection. Applied soft computing, ٣(٢), ١٢٣-١٣٧.

١٣. Lopez-Molina, C., De Baets, B., & Bustince, H. (٢٠١١). Generating fuzzy edge images from gradient magnitudes. Computer Vision and Image Understanding, ١١٥(١١), ١٥٧١–١٥٨٠.

١٤. Melin, P., Mendoza, O., & Castillo, O. (٢٠١٠). An improved method for edge detection based on interval type-٢ fuzzy logic. Expert Systems with Applications, ٣٧(١٢), ٨٥٢٧-٨٥٣٥.

١٥. Shu, X., & Ye, Y. (٢٠٢٣). Knowledge Discovery: Methods from data mining and machine learning. Social Science Research, ١١٠, ١٠٢٨١٧.

١٦. Alves, V. M., Korn, D., Pervitsky, V., Thieme, A., Capuzzi, S. J., Baker, N., ... & Tropsha, A. (٢٠٢٢). Knowledge-based approaches to drug discovery for rare diseases. Drug Discovery Today, ٢٧(٢), ٤٩٠-٥٠٢.

١٧. Blanco-Gonzalez, A., Cabezon, A., Seco-Gonzalez, A., Conde-Torres, D., Antelo-Riveiro, P., Pineiro, A., & Garcia-Fandino, R. (٢٠٢٣). The role of ai in drug discovery: challenges, opportunities, and strategies. Pharmaceuticals, ١٦(٦), ٨٩١.

١٨. Jacquey, F., Comby, F., & Strauss, O. (٢٠٠٨). Fuzzy edge detection for omnidirectional images. Fuzzy sets and Systems, ١٥٩(١٥), ١٩٩١-٢٠١٠.

١٩. Bouchet, A., Alonso, P., Pastore, J. I., Montes, S., & Díaz, I. (٢٠١٦). Fuzzy mathematical morphology for color images defined by fuzzy preference relations. Pattern Recognition, ٦٠, ٧٢٠-٧٣٣.

۲۰. Bloch, I. (۲۰۱۵). Fuzzy sets for image processing and understanding. Fuzzy Sets and Systems, ۲۸۱, ۲۸۰-۲۹۱.