

بهبود روش طبقه‌بندی ماشین بردار پشتیبان در سیستم‌های اطلاعات مکانی فراگستر جهت پایش بیماران آسمی

ندا کفاش چرندابی^{۱*}، علی‌اصغر آل‌شیخ^۲

^۱ دانشجوی دکتری سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه

نصیرالدین طوسی

n_kaffash@yahoo.com

^۲ استاد دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی

alesheikh@kntu.ac.ir

(تاریخ دریافت مهر ۱۳۹۵، تاریخ تصویب آذر ۱۳۹۵)

چکیده

بیماری آسم یکی از مشکلات عمده در سلامت عمومی انسانها است که زندگی میلیون‌ها نفر در سراسر جهان را تحت تاثیر قرار داده است. نظر به اینکه این بیماری تحت تاثیر عوامل محیطی متفاوتی می‌باشد و روشی برای درمان قطعی مبتلایان وجود ندارد؛ شناسایی مناطق خطر تشدید آسم فرد بیمار و اعلام هشدارهای لازم در زمان مناسب گام موثری در مدیریت و پایش این بیماری به حساب می‌آید. یکی از جدیدترین و کاربرپسندترین بسترهای محاسباتی در پایش زندگی روزانه افراد و به‌ویژه بیماران، سیستم‌های اطلاعات مکانی فراگستر می‌باشند؛ زیرا این سیستمها خدمات خود را به کاربران به صورت زمینه‌آگاه و با کاهش تعامل با کاربر ارائه می‌دهند. شرایط و ویژگی‌های این سیستم‌های جدید، تطبیق و بهبود روشهای مدلسازی مورد استفاده در آنها را ضروری می‌سازد. این تحقیق با بهبود روش ماشین بردار پشتیبان، به پیش‌بینی وضعیت آسم بیماران در چارچوب بسترهای محاسباتی فراگستر می‌پردازد. روش پیشنهادی با قابلیت بهره‌گیری از داده‌های بدون برچسب و برچسب‌دار در فضای زمینه، فرآیند پیش‌بینی وضعیت بیماری آسم را بهبود می‌بخشد. برای پیاده‌سازی مدل پیشنهادی، داده‌های موردنیاز شامل انواع آلاینده‌ها، دما، رطوبت، فشار هوا، مکان افراد، فاصله از پارک‌ها، خیابان‌ها و اطلاعات شخصی در ۲۷۰ موقعیت مختلف برای بیماران مبتلا به انواع آسم اخذ گردید. برچسب داده‌های جمع‌آوری شده، وضعیت بیماری آسم بود که توسط پیک فلومتر در برخی از موقعیت‌ها ثبت شد. برای ارزیابی مدل پیشنهادی، نتایج پیش‌بینی با نتایج الگوریتم اصلی ماشین بردار پشتیبان و الگوریتم کلونی مورچه‌ها در حالات مختلف مقایسه گردید. به طور میانگین، نتایج حاکی از دقت ۹۰ درصدی روش پیشنهادی و بهبود ۱۲ درصدی نسبت به روشهای مذکور بوده است.

واژگان کلیدی: سیستم اطلاعات مکانی فراگستر، آسم، پیش‌بینی نیمه نظارتی، SVM، ACO

۱- مقدمه

پیشرفت علم و فناوری در بخش‌های گوناگون منجر به توسعه‌ی ابزارهای مطالعاتی بشر شده است. ابزارهای آنالیز و مدلسازی داده‌های مکانی نیز از این امر مستثنی نبوده و در سالهای اخیر گسترش فراوانی پیدا کرده‌اند. سیستم اطلاعات مکانی فراگستر (UbiGIS)^۱ نام انتخابی برای سیستم‌های جدید ارائه شده در حیطه بهره‌گیری از محاسبات فراگستر در مسائل و کاربردهای مرتبط با اطلاعات مکانمند می‌باشد [۱، ۲].

اصطلاح محاسبات فراگستر، در سال ۱۹۸۸ توسط مارک وایزر برای نخستین بار ارائه شد. بر مبنای نظریه‌ی وایزر، فناوریهای رایانه‌ای آینده به سمتی حرکت می‌کنند که در محیط و اشیا پیرامون بشر موجود و قابل دسترس باشند تا به افراد در فعالیتهای روزمره کمک نمایند [۳]. بسترهای محاسباتی فراگستر، با داشتن روابط اشکار در درون سیستم^۲، اخذ خودکار^۳ دانش از محیط، زمینه آگاهی^۴ و فعال شدن خودکار سیستم بدون نیاز به اخذ دستور و فرمان از کاربر^۵ محیط‌های هوشمندی را به وجود می‌آورند [۴، ۵].

در بستر UbiGIS در هر مکان، در هر زمان، با هر ابزاری، تحت هر شبکه‌ای، هرسرویسی برای هر کاربر ارائه می‌گردد. در UbiGIS کاربر و محیط پیرامون او اهمیت پیدا می‌کنند. به بیان دیگر، UbiGIS نسل چهارم سیستم‌های اطلاعات مکانی می‌باشد که با اضافه شدن زمینه^۶ آگاهی و سادگی در استفاده^۷ به نسل اطلاعات مکانی همراه و تحت اینترنت، به وجود آمده است [۱، ۲]. سیستم‌های فراگستر از ابعاد گوناگونی قابل بررسی می‌باشند که از این میان استنتاج^۸ و پیش‌بینی زمینه از اهمیت بیشتری برخوردار است [۶].

بهره‌گیری از روشها و مدل‌های مختلف ارائه شده در بستر محاسبات فراگستر، نیازمند آن است که ویژگی این سیستم‌ها در فرآیند مدلسازی زمینه لحاظ گردد. یکی از مهم‌ترین ویژگی این سیستم‌ها اخذ داده از منابع مختلف

نظیر انواع حسگرها، پردازنده‌ها و پایگاه داده‌ها است. وجود حسگرهای مختلف سبب می‌شود، حجم بالایی از داده‌ها با سرعت بالا و هزینه کم جمع‌آوری گردد. لیکن در اغلب پایگاه داده‌هایی که وجود دارند، برچسب^۹ کلاسها برای تعداد اندکی از داده‌ها مشخص است. این در حالی است که برچسب زنی تمام داده‌ها کاری دشوار، پرهزینه و زمان‌بر است که نیاز به افراد متخصص دارد [۴، ۷]. لذا در بستر UbiGIS، بهره‌گیری از روشهایی توانمند با قابلیت مدلسازی حجم بالای داده (با برچسب و بدون برچسب) و به تبع آن کاهش تعامل کاربران حائز اهمیت فراوانی است. در این مطالعه منظور از برچسب، کلاس یا شناسه مشخص داده‌ها (نظیر داده‌های آموزشی) در فرآیند پیش‌بینی و طبقه‌بندی می‌باشد [۷].

مزایا و قابلیت‌های فراوان سیستم‌های فراگستر، بهره‌گیری از چنین بسترهای توانمندی را در حیطه‌های کاربردی مختلف به ویژه مسائل مرتبط با بهداشت، سلامت و بیماری‌ها ضروری می‌نماید. یکی از بیماریهای شایع و مزمن به ویژه در شهرهای بزرگ و آلوده نظیر تهران، بیماری آسم می‌باشد. از یک سو، تعدد مبتلایان به این نوع از آلرژی و محدوده سنی گسترده بیماران، یکی از دلایل متقن برای برنامه‌ریزی در جهت مدلسازی و پایش دقیق این بیماری است. از سوی دیگر، مکانمند بودن عوامل موثر در بروز و تشدید این بیماری، مدلسازی آن در حیطه علوم و فنون جدید نظیر UbiGIS را ضروری می‌سازد [۸، ۹]. لذا این مقاله به پیش‌بینی وضعیت تشدید بیماری آسم بر مبنای روشهای توانمند در بستر UbiGIS می‌پردازد.

در این تحقیق به سه ویژگی مهم سیستم‌های فراگستر توجه شده است. ویژگی اول زمینه‌آگاه بودن این سیستم‌ها است؛ یعنی پایش بیماران آسم و پیش‌بینی وضعیت آسم هر بیمار بر مبنای داده‌های زمینه‌ای (محیطی و فردی) هر بیمار بوده است. ویژگی دوم سیستم‌های فراگستر قابلیت مدلسازی بیماری در هر مکان و هر زمان می‌باشد که با توجه به داده‌های جمع‌آوری شده برای این تحقیق انجام شده است. ویژگی سوم بهبود روشهای مدلسازی زمینه به کار رفته در بستر سیستم‌های فراگستر به کمک داده‌های بدون برچسب در فضای زمینه بوده است.

^۱ Ubiquitous Geospatial Information System

^۲ Transparent interaction

^۳ Automated capture

^۴ Context awareness

^۵ Proactive and reactive

^۶ Context

^۷ Ease of Use

^۸ Reasoning

^۹ Label

نامه ارشد خود، پس از رتبه‌بندی کلاسهای بیماری با روش ویکور^۱ در فرآیندی مشابه با تحقیق [۱۵] به کمک شبکه عصبی به پیش‌بینی تشدید وضعیت بیماران آسمی پرداخت [۸]. Lee و همکارانش نیز به بررسی وضعیت بیماران آسم بر مبنای داده‌های آلودگی و هواشناسی به کمک الگوریتم‌های داده‌کاوی پرداختند [۱۶].

در تحقیقات پیشین، جهت مدل‌سازی زمینه‌های مسائل مختلف، غالباً از روشهای ساده‌ای استفاده شده است که به شدت تحت تاثیر نظر کارشناسان و برنامه‌نویسان است. به بیان ساده‌تر مدل‌سازی بر مبنای چند قانون تعیین شده توسط کارشناس انجام شده است و به شرایط مسئله و زمینه‌های کاربر کمتر توجه شده است. نکته قابل توجه اینکه در بکارگیری روشهای مختلف در بستر سیستم‌های فراگستر، به ویژگی‌های این سیستم‌ها توجهی نشده است. وجود حجم بالایی از داده‌های حسگرهای مختلف، اخذ سریع و ارزان داده‌های بدون برچسب و کاهش تعامل کاربر جزو ویژگی‌های بارز بسترهای محاسباتی فراگستر بوده است که در تحقیقات پیشین کمتر مورد توجه قرار گرفته است [۴، ۱۷]. در این تحقیق کاهش تعامل کاربر، به معنی کاهش فرآیند برچسب زنی توسط افراد متخصص برای داده‌ها و کاهش تعداد دفعات استفاده از ابزارهای سنجش وضعیت آسم برای بیمار می‌باشد.

روشهای نیمه‌نظارتی دسته‌ای از روشهای طبقه‌بندی می‌باشند که با بهره‌گیری از تعداد زیادی از داده‌های بدون برچسب و تعداد کمتری از داده‌های برچسب‌دار به طبقه‌بندی مسائل مختلف می‌پردازند. یکی از روشهایی که قابلیت بالایی در پیش‌بینی و طبقه‌بندی در حالت نظارت شده، نیمه نظارتی و نظارت نشده را دارا است، الگوریتم ماشین بردار پشتیبان^۲ می‌باشد. برای بهره‌گیری از این الگوریتم در حالت نیمه نظارتی، مدل‌های مختلفی توسعه داده شده است. لیکن در میان انبوه این مدل‌ها، به بهره‌گیری از راهکاری ترکیبی از روشهای گراف مبنا و SVM تاکید بسیار شده است [۷، ۱۷]. این تحقیق بر آن است که با پیشنهاد یک روش ترکیبی، بر مبنای داده‌های آلاینده‌های هوا، داده‌های هواشناسی، اطلاعات شخصی کاربران و موقعیت کاربران به پایش بیماران آسمی بپردازد. در این راستا به کمک بهره‌گیری از قوانین

در سالهای اخیر بکارگیری از مدل‌های زمینه‌آگاه جهت توسعه سیستم‌های بهداشت فراگستر افزایش قابل توجهی پیدا کرده است. به طور نمونه Choi و همکارانش [۱۰] یک میان افزار طراحی کرده‌اند که با اتصال به شش داده اصلی شامل ضربان قلب، دمای بدن، حالت چهره، دمای اتاق، زمان و مکان به آموزش و پیش‌بینی علایق کاربر در کارهای خانه می‌پردازد. به این صورت که تاریخچه این اطلاعات نگهداری می‌شود و آموزش و پیش‌بینی الگوهای کاربر بر مبنای این اطلاعات توسط شبکه عصبی انجام می‌گردد. سپس خروجی کار که به عنوان مثال شامل تغییر وضعیت صدا، هوا، نور و غیره است، ارائه می‌گردد. Feki و همکارانش [۱۱] یک الگوریتم ترکیبی بر مبنای منطق فازی و الگوریتم Q-learning جهت پیش‌بینی خودکار فعالیت‌های کاربر به صورت زمینه‌آگاه ارائه کرده‌اند. این تحقیق به طور نمونه برای کمک به یک فرد معلول در خانه هوشمند پیاده سازی شده است. Yuan و Herbert یک مدل فازی مبنا و یک واسطه استنتاج‌گر زمینه‌آگاه ارائه داده‌اند که یک سرویس کنترل درمانی شخصی برای بیماران مزمن فراهم می‌کند [۹]. تحقیق Lo و همکارانش [۱۲] یک سرویس مراقبت درمانی زمینه‌آگاه و مکان آگاه به صورت فراگستر ارائه کرده است که ضمن دریافت علائم حیاتی بیماران به کمک حسگرها، توصیه‌های لازم را جهت کنترل رژیم غذایی، کنترل وزن، مراقبت بهداشتی و ارسال نمایه به آمبولانس یا پزشک انجام می‌دهد.

در تحقیق دیگری [۱۳] یک چارچوب زمینه‌آگاه برای مراقبت درمانی از افراد مبتلا به آلزایمر توسعه داده شده است. در این تحقیق از روشهای هوش جمعی برای استنتاج نایقینی موجود در داده‌های زمینه استفاده شده است. Elgazzar و همکارانش [۱۴] نیز یک سرویس تحت اینترنت همراه برای پایش فراگستر افراد دارای بیماری‌های مزمن نظیر بیماری‌های قلبی ارائه داده‌اند که نیاز به پایش پیوسته‌ی علائم حیاتی و وضعیت آنها وجود دارد. در این تحقیق برای علائم حیاتی حد آستانه‌هایی معین شده است تا با رد شدن از آنها سیستم به کاربر هشدار دهد. Chu و همکارانش یک سیستم فراگستر برای پایش بیماران آسم پیشنهاد کردند. در این تحقیق بر مبنای داده‌های آلاینده‌های هوا نظیر CO, NO_x, O₃, SO₂، داده‌های هواشناسی شامل رطوبت، دما و فشار هوا موقعیت بیماران آسمی را به کمک روش شبکه عصبی پیش‌بینی نمودند [۱۵]. کشته‌گر در پایان

^۱ Vikor

^۲ Support Vector Machine (SVM)

فرمول مرز تصمیم‌گیری خطی در حالت کلی مطابق رابطه (۱) است [۲۰، ۲۱].

$$w.x + b = 0 \quad (1)$$

x یک نقطه روی مرز تصمیم‌گیری، w یک بردار n بعدی عمود بر مرز تصمیم‌گیری و $\frac{b}{\|w\|}$ فاصله مبدأ تا مرز

تصمیم‌گیری است. اولین مرحله برای محاسبه مرز تصمیم‌گیری بهینه، پیدا کردن نزدیک‌ترین نمونه‌های آموزشی دو کلاس است. در مرحله بعد فاصله آن نقاط از هم در راستای عمود بر مرزهایی که دو کلاس را به طور کامل جدا می‌کنند، محاسبه می‌شود. مرز تصمیم‌گیری بهینه، مرزی است که حداکثر حاشیه را داشته باشد. مرز تصمیم‌گیری بهینه با حل مسئله بهینه‌سازی (۲) محاسبه می‌شود [۲۰، ۲۱].

$$\min_{w,b} \frac{1}{2} \|w\|^2, Y_i(w^T \cdot x_i + b) - 1 \geq 0, i = 1, \dots, L \quad (2)$$

بهینه‌سازی فرمول (۲) کار دشواری است که با استفاده از ضرایب لاگرانژ^۳ قابل حل می‌شود. تا بدین جا مساله با فرض تفکیک خطی دو کلاس حل شد. حال آنکه در مسائل دنیای واقعی، کلاس‌ها با هم همپوشانی داشته و جدا کردن کلاس‌ها به وسیله مرز تصمیم‌گیری خطی، با خطا همراه خواهد بود. برای حل این مشکل، روش SVM در ابتدا داده‌ها را از فضای اولیه R^n با استفاده از یک تبدیل غیر خطی ϕ ، به فضای R^m با ابعاد بیشتر منتقل می‌کند تا کلاس‌ها تداخل کمتری با یکدیگر داشته باشند. سپس در فضای جدید با استفاده از جایگزینی x_i با $\phi(x_i)$ و در نظر گرفتن مقداری خطا؛ مرز تصمیم‌گیری بهینه محاسبه می‌شود. در این حالت یافتن مرز تصمیم‌گیری بهینه و حل مسئله بهینه‌سازی مطابق رابطه (۳) خواهد بود [۲۰، ۲۱]. λ_i ها ضرایب لاگرانژ می‌باشند و C یک عدد ثابت است. اگر $C \rightarrow \infty$ ، مسئله بهینه‌سازی جهت یافتن مرز برای کلاس‌های با تداخل بسیار زیاد و اگر $C \rightarrow 0$ ، مسئله بهینه‌سازی به سمت یافتن مرز بهینه جدا کننده برای دو کلاس با تداخل بسیار کم، مناسب خواهد بود.

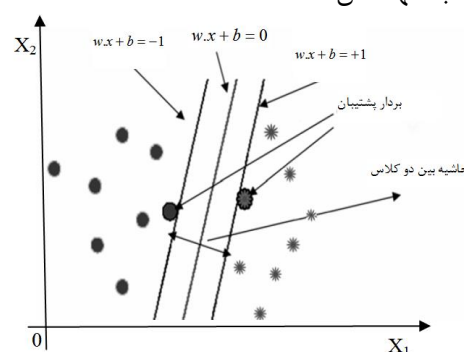
گراف و نزدیک‌ترین همسایگی در فضای زمینه^۱، الگوریتم SVM بهبود یافت و پیش‌بینی وضعیت تشدید بیماری افراد مختلف در هر زمان، هر مکان و در شرایط مختلف انجام شد. برای بررسی و ارزیابی روش پیشنهادی، نتایج با الگوریتم نظارت شده SVM و الگوریتم ACO^۲ مورد مقایسه و بررسی قرار گرفت.

۲- مبانی نظری

این بخش به بیان مفاهیم دو روش مورد استفاده در این تحقیق، الگوریتم ماشین بردار پشتیبان و الگوریتم بهینه‌سازی کلونی مورچه‌ها می‌پردازد.

۲-۱- الگوریتم ماشین بردار پشتیبان

ماشین بردار پشتیبان یک روش توانمند برای کلاسه‌بندی و رگرسیون است که اولین بار توسط Vapnik معرفی شد. الگوریتم رایج ماشین بردار پشتیبان، یک الگوریتم نظارت شده است که در اصل یک طبقه‌کننده دودویی است. روش SVM سعی دارد یک فراضحه ایجاد نماید که فاصله هر کلاس را تا فراضحه حداکثر کند. داده‌های نقطه‌ای که به فراضحه نزدیک‌ترند، برای اندازه‌گیری این فاصله بکار می‌روند. از این رو، این داده‌های نقطه‌ای، بردارهای پشتیبان نام گرفتند [۱۸]. در شکل ۱ دو کلاس و بردارهای پشتیبان مربوط به آنها نشان داده شده است.



شکل ۱- مرز خطی بهینه برای دو کلاس جدا از هم [۱۸]

با فرض دو کلاسه بودن داده‌ها $(x_i, i=1, \dots, L)$ ، داده وجود دارد که با $Y_i = \pm 1$ برچسب زده می‌شوند. برای محاسبه مرز تصمیم‌گیری دو کلاس کاملاً جدا از هم، از روش حاشیه بهینه استفاده می‌شود [۱۸، ۱۹].

^۳ Lagrange multiplier

^۱ Context space

^۲ Ant Colony Optimization (ACO)

$$Y_i(w.x_i + b) - 1 = 0 \quad \lambda_i, i=1, \dots, L \quad (5)$$

$$f(x, w, b) = \text{sgn}(w.x + b) \quad (6)$$

ماشین بردار پشتیبان اصالتاً یک طبقه‌بندی کننده دودویی است. حال آنکه در اکثر مسائل مختلف و کاربردهای موجود در دنیای واقعی بیش از دو کلاس وجود دارد. برای استفاده از ماشین بردار پشتیبان در حالت چند کلاسه دو روش کلی وجود دارد [۲۰، ۲۱]. در مورد M کلاس، در روش اول که آن را روش یکی در مقابل بقیه^۶ می‌نامند، طبقه‌بندی کننده دودویی که هر کدام برای جدا کردن یک کلاس از بقیه کلاس‌ها می‌باشد طراحی می‌شود. بنابراین در این حالت برای آموزش هر طبقه‌بندی کننده از تمام نقاط آموزشی M کلاس استفاده می‌گردد. چون در این حالت از تمام نقاط آموزشی برای آموزش هر طبقه‌بندی کننده استفاده می‌شود، بنابراین زمان آموزش هر طبقه‌بندی کننده در این حالت بسیار زیاد است و در عمل به ندرت از آن استفاده می‌شود. روش دوم که آن را دسته‌بندی کردن زوج-زوج^۷ می‌نامند، $\binom{M}{2} = \frac{M(M-1)}{2}$ طبقه‌بندی کننده دودویی که در واقع برابر با تمامی جفت‌های مختلف M کلاس می‌باشد، طراحی می‌شود. در این روش چون هر طبقه‌بندی با نقاط آموزشی کمتری آموزش داده می‌شود، زمان آموزش آن بسیار کمتر از حالت قبل است [۲۲، ۷].

مهمترین مزایای روش SVM قابلیت طبقه‌بندی مسائل خطی و غیرخطی؛ مسائل نظارت شده، نیمه نظارتی و نظارت نشده؛ مسائل چند کلاسه؛ کارآمد برای داده‌هایی با توزیع نامنظم و مبهم و گیر نیافتادن در بهینه محلی می‌باشد [۲۰، ۲۱]. لذا در این تحقیق برای طبقه‌بندی و پیش‌بینی وضعیت بیماران آسمی انتخاب شده است.

۲-۲- الگوریتم کلونی مورچه‌ها

این خانواده از روش‌ها اولین بار توسط Dorigo به عنوان یک راه‌حل چند عامله^۸، برای حل مسائل بهینه‌سازی ترکیبی، ارائه شد [۲۳]. الگوریتم کلونی

$$\lambda_1, \dots, \lambda_L \left[-\frac{1}{2} \sum_{i=1}^L \sum_{j=1}^L \lambda_i Y_i(\varphi(x_i), \varphi(x_j)) Y_j \lambda_j + \sum_{i=1}^L \lambda_i \right] \quad (3)$$

$$C \geq \lambda_i \geq 0, i=1, \dots, L$$

$$\sum_{i=1}^L \lambda_i Y_i = 0$$

به طور معمول در رابطه فوق، به جای استفاده از $\varphi(x_i), \varphi(x_j)$ از یک تابع کرنل مطابق $K(x_i, x_j) = \varphi(x_i) \varphi(x_j)$ استفاده می‌شود. $K(x_i, x_j)$ در واقع یک تابع در فضای اولیه می‌باشد که برابر با ضرب داخلی دو بردار در فضای ویژگی است. برای معادل بودن تابع $K(x_i, x_j)$ با ضرب داخلی دو بردار در فضای ویژگی، باید $K(x_i, x_j)$ یک تابع معین و مثبت متقارن بوده و در شرط مرسر^۱ صدق کند [۱۹، ۲۰، ۲۱].

متداولترین توابع کرنل که در این شرط صدق می‌کنند؛ کرنل خطی^۲ $(K(x_i, x_j) = x_i \cdot x_j)$ ، کرنل چندجمله‌ای^۳ $(K(x_i, x_j) = (\tau + x_i \cdot x_j)^p, \tau \geq 0)$ ، کرنل پایه شعاعی^۴ $(K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}})$ و کرنل حلقوی^۵ $(K(x_i, x_j) = \tanh(k_1 x_i \cdot x_j + k_2))^p$ هستند. میزان کارایی ماشین‌های بردار پشتیبان با توجه به نوع تابع کرنل متفاوت می‌باشد. اینکه کدام تابع کرنل برای کدام سری داده مناسب است، به درستی معلوم نیست. بنابراین، باید بهترین تابع کرنل برای هر سری داده بررسی و تعیین شود [۲۰، ۲۱].

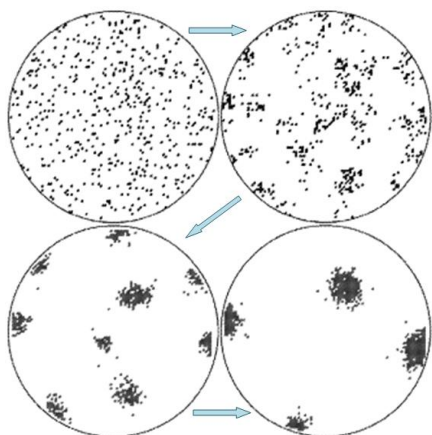
با حل معادله (۳) و یافتن ضرایب لاگرانژ، w با استفاده از رابطه (۴) محاسبه می‌شود.

$$w = \sum_{i=1}^L \lambda_i Y_i x_i \quad (4)$$

λ_i بردارهای پشتیبان بزرگتر از صفر و λ_i سایر نقاط صفر خواهد بود. بنابراین با توجه به صفر بودن λ_i مربوط به داده‌هایی که بردار پشتیبان نیستند، با تعدادی محدودی از نقاط آموزشی، مرز تصمیم‌گیری قابل حصول است. پس از یافتن w، با استفاده از روابط (۵) و (۶) مقدار b نهایی (با میانگین‌گیری از b های حاصل) و فرمول نهایی طبقه بندی بدست می‌آید.

^۱ Mercer
^۲ Linear Kernel
^۳ Polynomial Kernel
^۴ Radial basis function (RBF) Kernel
^۵ Sigmoidal Kernel

^۶ One vs All (OvA)
^۷ One vs One (OvO)
^۸ Multi Agent



شکل ۲- شمایی از رفتار خوشه‌بندی توسط مورچه‌های واقعی [۲۵]

مدلهای فراوانی برای توابع احتمال رهاسازی و برداشتن عناصر توسط مورچه‌ها ارائه شده است. مدل بکار رفته در این تحقیق مطابق روابط (۷) و (۸) بوده است که تحت عنوان مدل پایه برای این دو تابع احتمال شناخته شده‌اند [۲۵].

$$P_p = \left(\frac{k_1}{k_1 + f} \right)^2 \quad (7)$$

$$P_d = \left(\frac{f}{k_2 + f} \right)^2 \quad (8)$$

که P_p احتمال برداشتن یک عنصر، P_d احتمال رها کردن یک عنصر، f نسبت تشابهی دریافتی از همسایگان مجاور مورچه، k_1 و k_2 مقادیر حد آستانه‌ای می‌باشند.

۳- روش پیشنهادی

روشهای مدلسازی زمینه (به طور اخص، پیش‌بینی زمینه) از منظر بهره‌گیری از داده‌های مختلف به سه دسته روشهای نظارت شده، روشهای نظارت نشده و روشهای نیمه نظارتی تقسیم می‌شوند. روشهای نظارت شده به پیش‌بینی و طبقه‌بندی آن دسته از مسائلی می‌پردازد که در آن ورودی و خروجی مشخص است. به بیان ساده‌تر، ناظری وجود دارد که اطلاعاتی را در اختیار یادگیرنده قرار می‌دهد، و به این ترتیب روشها به دنبال یافتن ارتباط بین ورودی و خروجی می‌باشند. در روشهای نظارت نشده برخلاف مدلهای نظارت شده، داده‌های مشخصی از قبل وجود ندارد. هدف ارتباط ورودی و خروجی نیست، بلکه تنها دسته‌بندی بر مبنای ساختاری خاص انجام می‌شود. اما در روشهای نیمه نظارت شده هم از داده‌های برچسب

مورچه‌ها بر این اصل بنا نهاده شده که تعاملات محلی، محدود و ساده اعضای یک دسته و جمعیت با محیط، منتهی به یک رفتار جمعی هوشمندانه می‌شود. این تعاملات غالباً غریزی بوده و بدون نظارت انجام می‌گیرند و نتیجه آنها غالباً یک رفتار پیچیده و هوشمندانه جمعی و بطور خاص انجام بعضی بهینه‌سازی‌های پیچیده است. از مهمترین مزایای هوش جمعی می‌توان به عدم وجود کنترل متمرکز، توانایی سازگاری با تغییرات محیط پیرامون، تعاملات توزیع شده موجودات، سرعت انتقال و کارکرد موازی اشاره نمود [۲۳، ۲۴]. در سال‌های اخیر کاربردهای فراوانی از به‌کارگیری این الگوریتم در مسائل بهینه‌سازی پیچیده نظیر انتخاب اشیاء و تشخیص حالات، خوشه‌بندی داده‌ها، مسیریابی وسایل نقلیه و تکنیک‌های فراگیری ماشین ارائه گردیده است [۲۵].

اصول خوشه‌یابی بر اساس کلونی مورچه‌ها با روش‌های سنتی متفاوت است. در حالی که روشهای متداول مثل K-mean یک دسته‌بندی واضح از داده‌ها را ایجاد می‌کنند، خوشه‌یابی بر پایه کلونی مورچه‌ها به طور ضمنی این قسمت‌بندی را تولید می‌کند. اولین خوشه‌یابی و مرتب سازی بر پایه کلونی مورچه‌ها توسط Deneubourg و همکارانش در سال ۱۹۹۰ معرفی شد [۲۶]. این الگوریتم بر اساس رفتار مورچه‌ها در مرتب‌سازی لارو^۱ و تمیز کردن اجساد طراحی شده است (شکل ۲). در این خوشه‌بندی نظارت نشده، مورچه‌ها مانند عامل‌های ساده‌ای شبیه-سازی شدند که به طور اتفاقی در یک صفحه دو بعدی حرکت می‌کنند. عناصر داده‌ای که بایست خوشه‌یابی شوند، در این محیط پخش شده، توسط عامل‌ها جابجا می‌شوند. به بیان ساده مورچه‌ها با کنار هم قرار دادن داده‌های مشابه در این صفحه، آنها را مرتب می‌کنند که در شکل ۲ نشان داده شده است. تصمیم به برداشتن یا رها کردن یک عنصر تصادفی است؛ لیکن این تصمیم تصادفی تحت تاثیر نحوه پراکندگی عناصر می‌باشد تا احتمال رهاکردن یک عنصر در محیطی با داده‌های پیرامونی مشابه، افزایش یابد. به همین ترتیب، در یک محیطی که عنصر داده با عناصر متفاوت احاطه شده باشد یا هیچ عنصر دیگری نباشد، احتمال برداشتن آن افزایش خواهد یافت [۲۵، ۲۶].

^۱ larvae

خورده و هم از داده‌های برچسب نخورده به صورت همزمان استفاده می‌شود تا دقت یادگیری مقدار ی بهبود یابد. لذا ضمن اینکه تهیه داده‌های بدون برچسب سریع، کم هزینه و راحتتر می‌باشد، استفاده از فرآیندهای نیمه نظارتی منجر به بهبود دقت طبقه‌بندی نیز می‌شود [۲۷، ۷].

در این تحقیق با توجه به ویژگی‌های بسترهای محاسباتی فراگستر و وجود انبوه داده‌های بدون برچسب، توسعه و بهره‌گیری از روشهای نیمه نظارتی مورد توجه قرار گرفته است. لذا برای پیش بینی وضعیت آسم بیماران، یک مدل ترکیبی برای روش غیرخطی چندکلاسه نیمه نظارتی SVM پیشنهاد شده است. نکته قابل توجه اینکه فرآیند پیش‌بینی نیاز به داده‌های آموزشی دارد. داده‌های آموزشی، داده‌هایی هستند که برچسب آنها مشخص است [۱۷]. در این تحقیق برچسب داده‌ها، مقادیر PEF^۱ بیماران است که توسط پیک فلومتر^۲ اندازه‌گیری شده است. اسپیرومتری^۳ و پیک فلومتر دو آزمون رایج برای سنجش بازدم بیماران آسمی هستند که هر کدام شاخص‌های گوناگونی دارند. دلیل انتخاب پیک فلومتر، ارزان بودن، سهولت استفاده و حمل آسان آن برای بیماران است [۲۸].

برای حل الگوریتم نیمه نظارتی SVM نیاز به حل همزمان پارامترهای فراصفحه و بردار برچسب‌ها می‌باشد. به بیان ساده‌تر، معادله (۳) و بردار برچسب‌ها که مجهول می‌باشند، بایستی تعیین شوند؛ با این شرط که برچسب معلوم نقاط آموزش در اولویت باشند. برای حل این مسئله ترکیبی غیرمحدب، راه‌حل‌های گوناگونی پیشنهاد شده است [۷]. این راه‌حل‌ها با دو استراتژی کلی بهینه‌سازی ترکیبی یا بهینه‌سازی پیوسته به حل مسئله می‌پردازند. روشهای زیادی برای هر دسته پیشنهاد شده‌اند که هر کدام مزایا و معایب خاص خود را دارند. یکی از مدل‌های پیشنهادی برای حل این مسئله، استفاده از روشهای گراف‌مینا و الگوریتم SVM به صورت ترکیبی می‌باشد [۲۷، ۷].

در این تحقیق، برای حل مسئله غیرمحدب الگوریتم نیمه نظارتی SVM، یک فرآیند ترکیبی از روشهای گراف

مینا، SVM و فضای زمینه پیشنهاد شد. بدین ترتیب که از مزایای فضای زمینه در توزیع مناسب داده‌ها، روش گراف‌مینا و مزایای همسایگی در محاسبه بردار برچسب‌های مجهول و از روش SVM در حل پارامترهای فراصفحه برای طبقه‌بندی داده‌ها استفاده شد. مدل پیشنهادی با شرایط محیط محاسباتی فراگستر سازگارتر است؛ چراکه از حجم انبوه داده‌های بدون برچسب استفاده بهینه کرده و تعامل کاربر را کاهش می‌دهد [۷].

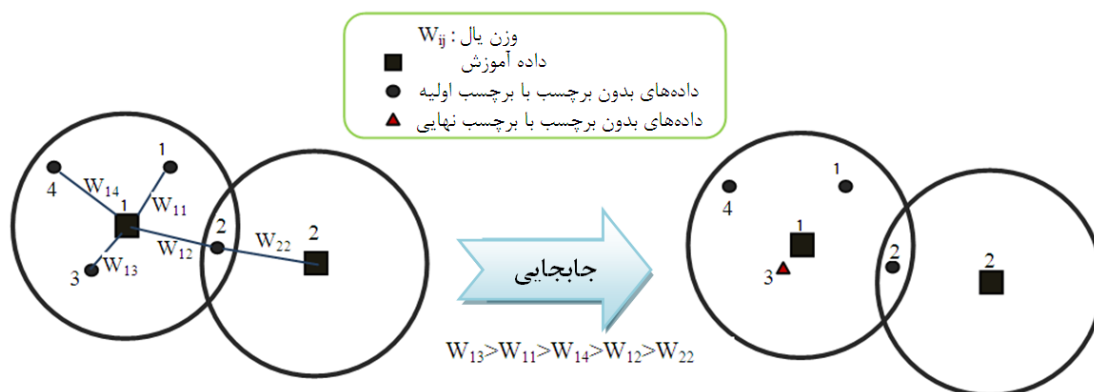
روش پیشنهادی از پنج گام تشکیل شده است. در گام نخست، بر مبنای داده‌های آموزشی (L) و داده‌های بدون برچسب (U) گراف مسئله تشکیل می‌شود. سپس کرنل مناسب برای الگوریتم SVM با توجه به شرایط مسئله انتخاب شده و با اجرای روش SVM برچسب‌های اولیه به مجموعه U اختصاص داده می‌شود. در گام بعدی، k نزدیکترین همسایه برای هریک از اعضای مجموعه L تعیین و گراف همسایگی تشکیل می‌شود. وزن یالهای این گراف (W_{ij}) برابر معکوس فاصله اقلیدسی انتخاب شد. پس یالهای با وزن بزرگتر مربوط به داده‌های بدون برچسبی می‌باشند که به داده آموزشی نزدیکتر هستند. در گام چهارم، نقاط بدون برچسب در همسایگی هر نقطه آموزشی، که وزن بیشتری دارند، برچسب نقطه آموزشی مربوطه را گرفته و جابجا می‌شوند. به عنوان مثال در شکل ۳، فرآیند برچسب زنی و جابجایی نقاط بدون برچسب برای ۶ نقطه نشان داده شده است. در این مثال، وزن تمام نقاط بدون برچسب از نقاط آموزشی حساب شده و نقطه بدون برچسب با وزن بیشتر (نقطه شماره ۳) وزن نقطه آموزش ۱ را گرفته و سپس جابجا می‌شود.

این مرحله با هدف حل بردار مجهول برچسب‌ها و نزدیکی نقاط بدون برچسب به نقاط آموزشی در فضای زمینه انجام می‌شود. در گام نهایی فضای زمینه بروز شده و روش SVM اجرا می‌گردد. این فرآیند تا رسیدن به حدآستانه مورد قبول (σ_0) ادامه می‌یابد. در نتیجه با اجرای روش پیشنهادی، فضای زمینه طبقه‌بندی شده و وضعیت آسم بیماران تخمین زده می‌شود. شبه کد روش پیشنهادی مطابق شکل ۴ می‌باشد. در خلال روش پیشنهادی، ضمن برچسب‌دهی به داده‌های بدون برچسب، دقت طبقه‌بندی داده‌های زمینه‌ای نیز افزایش می‌یابد.

^۱ Peak Expiratory Flow

^۲ Peak Flow meters

^۳ Spirometry



شکل ۳- فرآیند برچسب زنی و جابجایی نقاط بدون برچسب در روش پیشنهادی

Initialize: Create Context Space (Context 1, Context 2)

Construct D graph ($D = L \cup U$)

Initialized SVM Kernel and parameters

Run SVM algorithm: $SVM(\omega, \xi, b)$

Step 1: assign initial Label for U set

Loop until $|f(\omega^k, \xi^k) - f(\omega^{k-1}, \xi^{k-1})| < \sigma_0$

Step 2: Select k nearest neighborhood

Step 3: Calculate edge weight ($W_{ij} = (1/1 + \text{dist}(i, j))$)

Step 4: Change node with max W_{ij} position toward nearest training data

Step 5: assign initial Label for U set

Step 6: Run SVM ($f(\omega^k, \xi^k) = \frac{1}{2} \|\omega^k\| + C \sum_{i=1}^D \xi_i^k$)

Go to loop

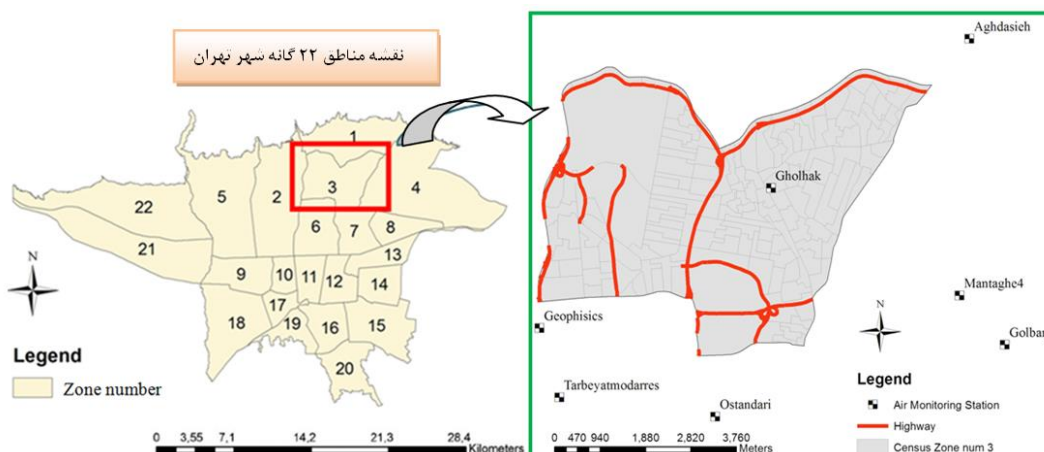
Step 7: Calculate PEF

شکل ۴- شبه کد روش پیشنهادی

۴- پیاده‌سازی و ارزیابی

شناخته شده است. به این دلیل، منطقه ۳ شهر تهران برای پیاده‌سازی انتخاب گردید که در شکل ۵ نشان داده شده است.

در سالهای اخیر، شهر تهران به عنوان یکی از آلوده‌ترین شهرهای جهان که نرخ بالایی از بیماری آسم را دارا می‌باشد،



شکل ۵- منطقه مطالعاتی به همراه بزرگراه‌ها و ایستگاه‌های پایش آلودگی هوا

داده‌های زمینه‌ای موثر در بیماری آسم تعیین و از پایگاه داده‌ها و منابع مختلف جمع‌آوری شد. به طور دقیق‌تر،

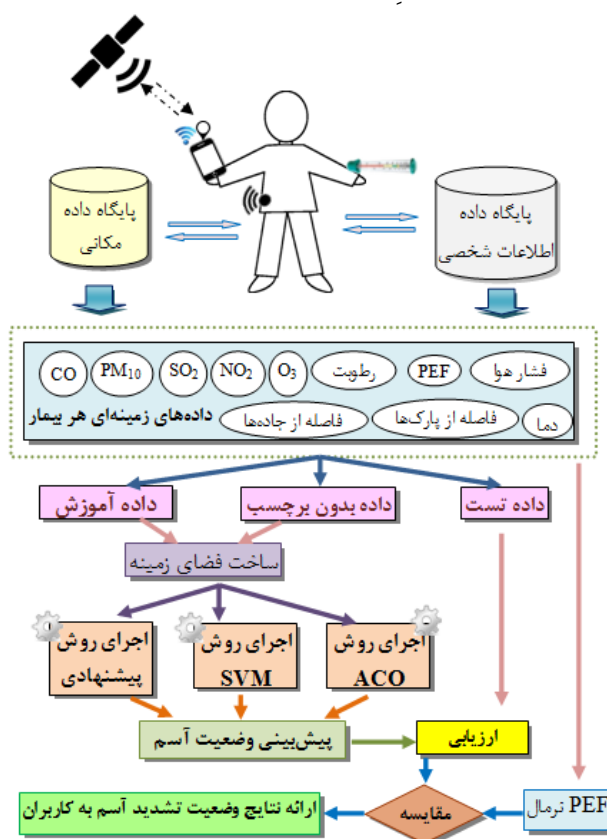
مراحل اجرایی این تحقیق مطابق فلوچارت شکل ۶ می‌باشد. بدین ترتیب که، پس از انتخاب منطقه مطالعاتی،

اندازه‌گیری و ثبت شد. لازم به ذکر است که به دلیل خودداری از افزایش طول مقاله، توضیحات مربوط به نحوه تعیین پارامترهای محیطی موثر در بیماری آسم و میزان اهمیت آنها حذف شد. علاقه‌مندان می‌توانند برای مطالعه بیشتر به [۸، ۲۸] مراجعه نمایند.

برای پیاده‌سازی، داده‌های مذکور برای بیماران دارای انواع مختلف آسم (آسم آلرژیک، آسم فصلی و آسم شغلی) در ۲۷۰ موقعیت مختلف جمع‌آوری شد (در فصل پاییز و زمستان). در این تحقیق سه بیمار (دو مرد و یک زن) مورد پایش قرار گرفتند و داده‌های مربوط به آنها جمع‌آوری گردید. داده‌های زمینه‌ای مربوط به یک مرد مبتلا به آسم آلرژیک، ۳۵ ساله با ۱۸۲ سانتی‌متر قد در ۱۱۰ وضعیت مختلف که ۳۰ وضعیت آن داده‌های بدون برچسب بوده‌اند. از ۸۰ داده‌ی برچسب‌دار ۴۰ داده برای آموزش و ۴۰ داده به عنوان داده تست استفاده شدند. یک خانم آرایشگر ۳۲ ساله با ۱۶۷ سانتی‌متر قد به عنوان نفر دوم که دارای آسم شغلی بود انتخاب گردید. داده‌های زمینه‌ای او در ۸۰ وضعیت مختلف (۳۰ داده برای آموزش، ۲۰ داده بدون برچسب و ۳۰ داده برای تست) جمع‌آوری شد. داده‌های زمینه‌ای نفر سوم که یک مرد ۴۲ ساله با ۱۷۹ سانتی‌متر قد مبتلا به آسم فصلی بود، مشابه نفر دوم در ۸۰ وضعیت مختلف اخذ و ثبت گردید. بنابراین در مجموع به ترتیب ۱۰۰، ۷۰ و ۱۰۰ داده آموزشی، داده بدون برچسب و داده تست جمع شد. سپس فضای دو بعدی زمینه با دو محور C1 و C2 ساخته شد. محور افقی زمینه (C1) نشانگر مجموع وزندار PM_{10} ، NO_2 ، SO_2 ، O_3 ، رطوبت، دمای بیش از ۱۵ درجه سلسیوس و فاصله از پارک‌ها می‌باشد. با افزایش ارزش این محور، وضعیت آسم خطرناک می‌شود. محور عمودی زمینه (C2) مجموع وزندار فشار هوا، فاصله از جاده‌ها و دمای کمتر از ۱۵ درجه سانتی‌گراد است. با کاهش مقادیر این محور، مقادیر PEF بیماران کاهش خواهد یافت و وضعیت سلامتی آنها در معرض خطر قرار خواهد گرفت.

برای بررسی دقیق‌تر داده‌های افراد با انواع مختلف آسم از هم تفکیک شده و برای هر کدام فضای زمینه بطور مجزا تشکیل شد. حال نوبت به اجرای روش‌های طبقه بندی؛ روش SVM، روش ACO و روش پیشنهادی این تحقیق می‌رسد. الگوریتم این روش‌ها در فضای برنامه نویسی Matlab® 7.12.0 کدنویسی شد. با توجه به چهار کرنل اصلی مورد استفاده در روش SVM بایستی اذعان کرد که

آلاینده‌های هوا شامل؛ SO_2 ، O_3 ، NO_2 و PM_{10} از ایستگاه‌های پایش آلودگی هوا جمع‌آوری شد (۷ ایستگاه پایش هوای واقع در پیرامون منطقه مطالعاتی در شکل ۵ نشان داده شده‌اند). برای این آلاینده‌ها از نقشه‌های میانگین به کمک روش درونیایی کریجینگ استفاده شد. با توجه به محدوده تغییرات آلاینده‌ها و اهمیت آنها در وقوع آسم، از نقشه‌های میانگین روزانه برای آلاینده‌های NO_2 و PM_{10} و برای سایر آلاینده‌ها از نقشه‌های میانگین ماهانه استفاده شد. به دلیل نرخ بالای تغییرات آلاینده CO، اهمیت و تاثیر آن در بیماری آسم؛ داده‌های آنی آلاینده CO به کمک حسگر MQ9 قابل حمل همراه شخص اخذ شد.



شکل ۶- فلوچارت مراحل اجرایی تحقیق

اطلاعات شخصی بیماران شامل؛ جنسیت، سن، قد و تاریخچه بیماری جمع‌آوری گشت و در پایگاه داده شخصی ذخیره گردید. پایگاه داده GIS حاوی داده‌های مکان مبنا با داده‌هایی شامل؛ منطقه مطالعاتی، خیابان‌ها، پارک‌ها و فضای سبز در مقیاس ۱:۲۰۰۰ تشکیل شد. همچنین مختصات کاربر در هر مکان، داده‌های هواشناسی شامل؛ دما، رطوبت و فشار هوا به کمک حسگرهای تلفن هوشمند همراه (سامسونگ گلکسی S4) جمع‌آوری و ذخیره گردید. اطلاعات PEF بیماران نیز به کمک پیک فلومتر قابل حمل،

کرل خطی و حلقوی با توجه به توزیع داده‌های تحقیق مناسب نبودند. لذا به بررسی کرل چندجمله‌ای و RBF پرداخته شد. مقادیر مختلف برای پارامترهای این دو کرل امتحان گردید و در نهایت بهترین کرل انتخاب شد. با توجه به شرایط منطقه مطالعاتی، داده‌های موجود و اجرای الگوریتم با کرل‌ها و پارامترهای مختلف، کرل چندجمله‌ای با $\tau = 2.3$ انتخاب شد. الگوریتم ACO با ۳۰ جمعیت اولیه، در ۱۰۰ تکرار و $k_1 = k_2 = 2.15$ اجرا شد. برای طبقه‌بندی به مقادیر نرمال PEF هر بیمار نیاز است که از روی جنسیت، قد و سن قابل حصول می باشد [۲۹]. سپس با مقایسه مقادیر تخمین زده شده PEF برای هر بیمار (خروجی روش‌های طبقه‌بندی) با مقادیر نرمال PEF، وضعیت آسم هر بیمار در چهار کلاس حاد، خطرناک، هشدار و نرمال به ترتیب با برچسب ۴-۳-۲ و ۱ طبقه‌بندی می‌شود. مبنای مقایسات و طبقه‌بندی داده‌ها در چهار طبقه مذکور مطابق جدول ۱ می‌باشد [۲۸].

لذا بردار برچسب‌های مجهول در روش پیشنهادی، $Y_i = 1; 2; 3; 4$ می‌باشد که به مجموعه داده‌های بدون برچسب اختصاص می‌یابد.

نتایج طبقه‌بندی وضعیت آسم حاصل از سه روش مذکور در شکل ۷ آمده است. علامت مثلث جهت نمایش داده‌های تست و علامت دایره برای نمایش داده‌های آموزشی استفاده شد. علامت ستاره در روش SVM نشان‌دهنده داده‌های بدون برچسبی هستند که به کمک این روش طبقه‌بندی شده‌اند. این علامت در روش ACO بیانگر داده‌های آموزشی و بدون برچسبی هستند که توسط عامل‌ها جابجا شده‌اند تا خوشه‌ها را تشکیل دهند.

با تفسیر بصری نتایج به دست‌آمده از طبقه‌بندی می‌توان اذعان داشت که در داده‌های آسم آلرژیک به دلیل توزیع همگن‌تر داده‌ها در فضای زمینه به نسبت دو نوع دیگر آسم، دقت طبقه‌بندی‌ها بالاتر است (به ویژه روش پیشنهادی و روش ACO). حال آنکه در آسم شغلی به دلیل توزیع ناهمگن (که ناشی از تشدید آسم در مناطق خاص مربوط به شغل بیمار بوده است) دقت طبقه‌بندی اندکی کمتر بوده است. همچنین در آسم فصلی نیز روش پیشنهادی و روش SVM با دقت خوبی طبقه‌بندی را انجام داده‌اند. با مقایسه نتایج حاصل از پیش‌بینی و مقادیر PEF نرمال برای هر فرد بر مبنای قوانین مذکور در جدول ۱، نتایج پیش‌بینی وضعیت آسم در اختیار کاربران قرار

می‌گیرد (مطابق شکل ۸). منظور از کاربر می‌تواند، خود بیمار، مراقب درمانی وی و یا پزشک متخصص باشد. برای بررسی دقیق‌تر نتایج از داده‌های تست استفاده شد. طبقه‌بندی برای داده‌های تست به کمک این سه روش انجام شده بود که با مقایسه کلاس PEF تخمینی از سه روش با مقادیر اندازه‌گیری شده برای آنها، دقت طبقه‌بندی برآورد شد. برای اینکار از ماتریس خطا^۱ استفاده گردید [۳۰].

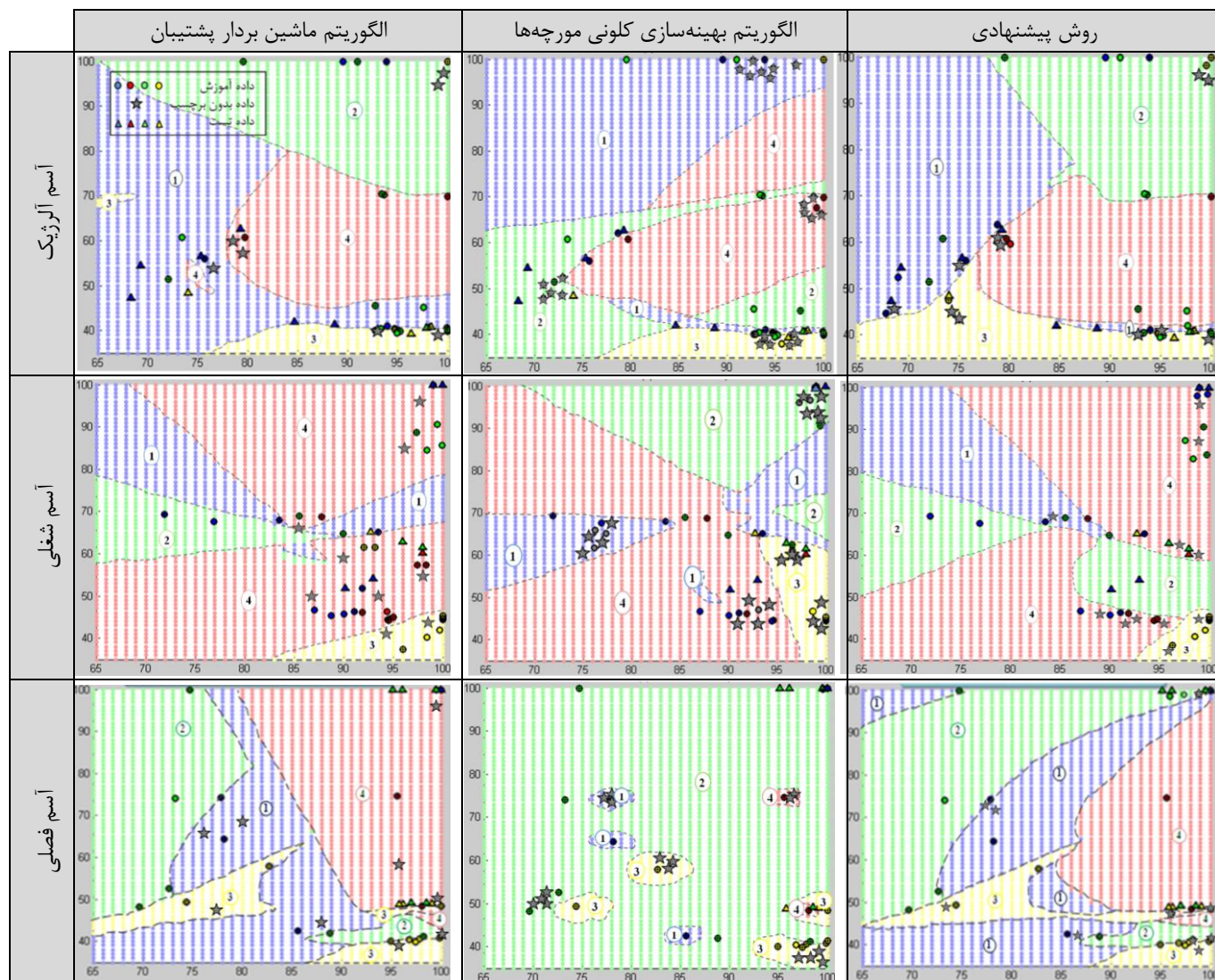
با توجه به چهار کلاس آسم مورد بررسی در این تحقیق، ابعاد ماتریس خطا 4×4 بوده است. عناصر سطری کلاس اندازه‌گیری شده برای PEF بیماران و عناصر ستونی کلاس تخمینی برای PEF بیماران (به دست آمده از هریک از سه روش مذکور) است. برای به دست آوردن دقت کلی^۲ طبقه‌بندی، مجموع عناصر قطری به تعداد کل داده‌های تست تقسیم شد. شاخص‌های دیگری نیز برای ارزیابی دقت به کمک ماتریس خطا وجود دارد که تحت عنوان دقت تولید کننده، دقت کاربر و ضریب کاپا شناخته شده‌اند و به کمک ماتریس خطا برآورد می‌شوند [۳۰]. در این تحقیق برای برآورد دقت کلی، ده سناریو مختلف با تغییر پارامتر τ کرل چندجمله‌ای الگوریتم SVM، الگوریتم پیشنهادی و k_1, k_2 الگوریتم ACO تعریف شد. سپس میانگین دقت کلی در هر سناریو برای هر سه نوع آسم و برای هر روش محاسبه گردید و در شکل ۹ نشان داده شد.

جدول ۱- شروط طبقه‌بندی وضعیت آسم در قیاس با PEF نرمال

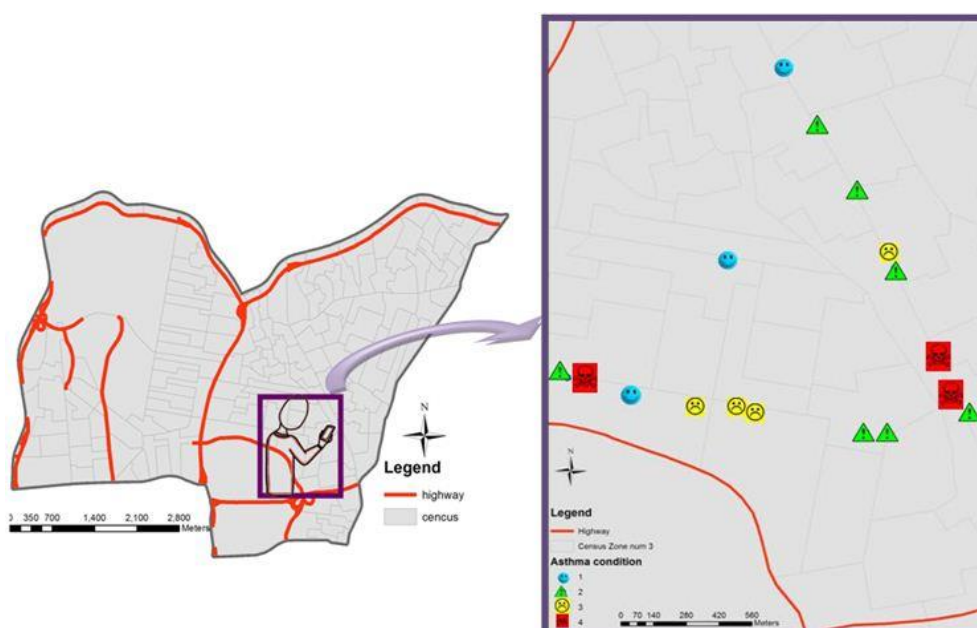
شماره برچسب	کلاس	سمبل	شرط
۴	حاد		اگر PEF پیش‌بینی شده مساوی یا کمتر از ۴۰٪ PEF نرمال باشد.
۳	خطرناک		اگر PEF پیش‌بینی شده مساوی یا کمتر از ۶۰٪ و بیش از ۴۰٪ PEF نرمال باشد.
۲	هشدار		اگر PEF پیش‌بینی شده مساوی یا کمتر از ۸۰٪ و بیش از ۶۰٪ PEF نرمال باشد.
۱	نرمال		اگر PEF پیش‌بینی شده بیشتر از ۸۰٪ PEF نرمال باشد.

^۱ Error matrix

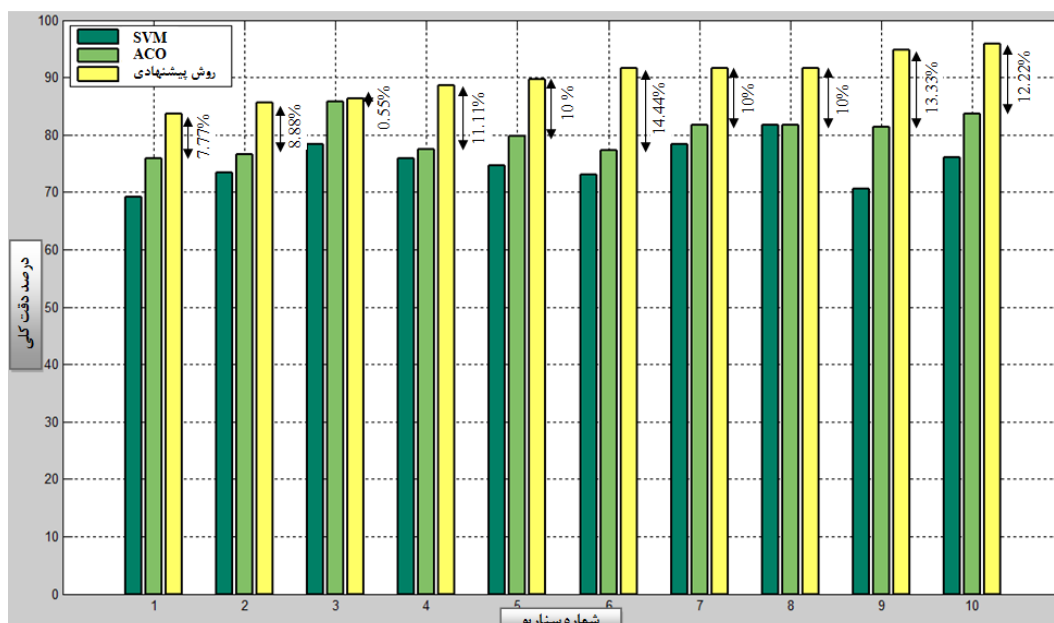
^۲ Overall Accuracy



شکل ۷- طبقه‌بندی حاصل از روش پیشنهادی، ACO و SVM برای داده‌های انواع آسم



شکل ۸- نمونه‌ای از نتایج پیش‌بینی وضعیت آسم در ۱۷ موقعیت مختلف برای بیمار مبتلا به آسم آلرژیک



شکل ۹- نمودار مقایسه دقت کلی روشهای SVM، ACO و روش پیشنهادی

جدول ۲- نتایج حاصل از ماتریس خطا برای کل داده‌های تحقیق

روش پیشنهادی	ACO	SVM	
میانگین دقت کلی (درصد)	۸۹.۹۴٪	۷۵.۱۱٪	۸۰.۱۱٪
حداکثر دقت کلی (درصد)	۹۵.۸۳٪	۸۱.۶۶٪	۸۵.۸۳٪
حداقل دقت کلی (درصد)	۸۳.۶۱٪	۶۹.۱۶٪	۷۵.۸۳٪
ضریب کاپا	۰.۸۶	۰.۶۵	۰.۷۱

۵- نتیجه‌گیری

مقاله حاضر به بررسی و پیش‌بینی وضعیت آسم بر مبنای محرک‌های محیطی موثر در تشدید آن پرداخته است. جهت پایش دقیق بیماران آسم در هر مکان و هر زمان و در هر شرایطی، یک روش ترکیبی پیش‌بینی نیمه‌نظارتی بر مبنای اصول گراف، قوانین همسایگی و روش SVM در بستر سیستم‌های اطلاعات مکانی فراگیرتر پیشنهاد شده است. روش پیشنهادی ضمن بهره‌گیری از داده‌های سهل الوصول و کم هزینه‌ی بدون برچسب، دقت پیش‌بینی را نیز بهبود بخشید. این ادعا بر مبنای ارزیابی مدل پیشنهادی در قیاس با روش نظارت شده SVM و روش ACO حاصل گردید. بررسی‌ها حاکی از دقت کلی بالای روش پیشنهادی، در حدود ۹۰ درصد بطور میانگین برای صد داده تست بوده است. همچنین دقت کلی پیش‌بینی در روش پیشنهادی نسبت به روش ACO و SVM به ترتیب ۹.۸۳٪ و ۱۴.۸۳٪ به داده‌های واقعی نزدیک‌تر بوده است. نکته قابل توجه اینکه، این دقت بالا برای پیش‌بینی کلاس وضعیت بیماران بوده است که برای

بررسی‌ها حاکی از آن است که روش ترکیبی گراف مبنا برای الگوریتم نیمه نظارتی SVM با میانگین دقت کلی ۸۹.۹۴٪ در بدترین و بهترین حالت به ترتیب از روش ACO ۵۵٪ و ۱۴.۴۴٪ و از روش SVM ۸.۰۵٪ و ۲۴.۱۶٪ بهتر عمل کرده است. به بیان دیگر، روش پیشنهادی در سناریوهای مختلف با حداکثر ۹۵.۸۳٪ و حداقل ۸۳.۶۱٪ دقت کلی، کلاسهای مقادیر مختلف PEF را پیش‌بینی می‌کند. بررسی انجام شده بر روی دقت کلی و ضریب کاپا روشهای مختلف در جدول ۲ آمده است. در این راستا، ذکر دو نکته حائز اهمیت است. نکته اول اینکه، روش پیشنهادی در سناریوهای مختلف نشان داد که حتی در مواردی که دقت طبقه‌بندی پایین‌تری دارد، اشتباهات مربوط به کلاس اول و دوم بوده است. به بیان دیگر، روش پیشنهادی در پیش‌بینی حالات حاد و خطرناک بیمار با دقت بالاتری (به طور متوسط، در حدود ۹۲٪) عمل کرده است. لذا قابلیت اعتماد به این روش بالا است. نکته دوم اینکه، با افزایش داده‌های ورودی، در روش ACO توانایی تحلیل حجم بالای داده‌ها در زمان کم وجود دارد. لیکن با پیچیده شدن مرز تصمیم، همچنین در نواحی با مرز اریب^۱ روش پیشنهادی دقت بالاتری را ارائه داده است. لذا می‌توان ادعان داشت در این تحقیق، طبقه‌بندی روش پیشنهادی با میانگین دقت کلی ۹.۸۳٪ از روش ACO و ۱۴.۸۳٪ از روش SVM دقیق‌تر بوده است.

^۱ Biased

پیش‌بینی نظیر الگوریتم بهینه‌سازی انبوه ذرات و سایر مدل‌های ارائه شده برای حل نیمه نظارتی روش SVM نظیر جستجوی محلی ترکیبی یا گرادیان نزولی SVM با نتایج مدل ارائه شده در این تحقیق مقایسه گردد. همچنین در این مقاله سه اصل مهم سیستم‌های فراگستر یعنی ارائه خدمات برای هر کاربر در هر زمان و هر مکان مورد توجه قرار گرفت. توجه به سایر ویژگی‌های بسترهای محاسباتی فراگستر، به ویژه تسهیل قابلیت ارائه هر سرویس، تحت هر شبکه و با هر ابزاری توصیه می‌گردد.

پایش وضعیت بیماران و اعلام هشدارهای لازم کافی است. پیش‌بینی مقادیر PEF با این دقت امکان‌پذیر نیست؛ چراکه بیماری آسم تحت تاثیر پارامترهای فراوان دیگری (نظیر استرس، تنش‌های عصبی، ژنتیک و غیره) نیز می‌باشد.

صرف نظر از نوآوری و مزایای این پژوهش، تحقیق حاضر نظیر تمامی مطالعات علمی با محدودیت‌هایی همراه بوده است. در دسترس نبودن تمام داده‌های موثر نظیر ترافیک و محدوده آلرژن‌ها و عدم دسترسی به حسگر سایر آلاینده‌ها از جمله این موارد هستند. همچنین پیشنهاد می‌شود برای ادامه کار در این حیطه، سایر روشهای توانمند

مراجع

- [1] Ki-Joune, L. (2007). "Ubiquitous GIS, Part I: Basic Concepts of Ubiquitous GIS". Lecture Slides, Pusan National University.
- [2] Malek, M. (2013). "Context-aware geoinformation and ubiquitous computing", K.N.Toosi University of Technology Press, First Edition, (in Persian).
- [3] Weiser, M. (1991). "The computer for the 21st century". Scientific American. Vol.265, No.3, PP. 94-104.
- [4] Krumm, J. (2010). "Ubiquitous Computing Fundamentals". CRC Press, Taylor & Francis Group.
- [5] Weiser, M. (1993). "Hot topics: Ubiquitous computing", IEEE Computer. Vol.26, No.10, PP. 71-72.
- [6] Sheng, Q. Z., Yu, J. and Dustdar, S. (2010). "Enabling Context-Aware Web Services Methods, Architectures, and Technologies". CRC Press, Taylor & Francis Group.
- [7] Chapelle, O., Sindhwani, V. and Keerthi, S. S. (2008). "Optimization Techniques for Semi-Supervised Support Vector Machines". Journal of Machine Learning Research. Vol.9, PP.203-233.
- [8] Keshtegar, E. (2015). "Design and Implementation of Ubiquitous Health System for asthma patients monitoring". Master Thesis, Faculty of Geodesy and Geomatics Engineering, K.N.Toosi University of Technology, Tehran, Iran, (in Persian).
- [9] Yuan, B. and Herbert, J. (2012). "Fuzzy CARA- A Fuzzy-Based Context Reasoning System For Pervasive Healthcare". The 3rd International Conference on Ambient Systems, Networks and Technologies, Procedia Computer Science, Vol.10, PP. 357- 365.
- [10] Choi, J.H., Choi, S.Y., Shin, D. and Shin, D. (2005). "Research and Implementation of the Context-Aware Middleware Based on Neural Network". Springer-Verlag Berlin Heidelberg, LNAI 3397. PP. 295-303.
- [11] Feki, M.A., Lee, S.W., Bien, Z. and Mokhtari, M. (2008). "Combined Fuzzy State Q-learning Algorithm to predict Context Aware User Activity under uncertainty in Assistive Environment". IEEE Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing. PP. 57-62.
- [12] Lo, Ch.Ch., Chen, Ch.H., Cheng, D.Y. and Kung, H.Y. (2011). "Ubiquitous Healthcare Service System with Context-awareness Capability: Design and Implementation". Expert Systems with Applications. Vol.38, PP. 4416-4436.
- [13] Lee, K., Lunney, T., Curran, K. and Santos, J. (2009). "A Context-Aware Platform for Computer Assisted Healthcare". PGNet, ISBN: 978-1-902560-22-9.
- [14] Elgazzar, Kh., Aboelfotoh, M., Martin, P. and Hassanein, H.S. (2012). "Ubiquitous Health Monitoring Using Mobile Web Services". The 3rd International Conference on Ambient Systems, Networks and Technologies, Procedia Computer Science. Vol.10, PP. 332 – 339.
- [15] Chu, H.T., Huang, Ch. Ch., Lian, Z. H. and Tsai, J. J. P. (2006). "A ubiquitous warning system for asthma-inducement". IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing. Taichung, PP. 186-191.
- [16] Lee, Ch., Chen, J., and Tseng, V.S. (2011). "A novel data mining mechanism considering bio-signal and environmental data with applications on asthma monitoring". Computer Methods and Programs in Biomedicine, Vol.101, No.1, PP.44-61.

- [17] Cheng, Sh., Huang, Q., Liu, J. and Tang, X. (2013). "A Novel Inductive Semi-supervised SVM with Graph-Based Self-training". Springer, Intelligent Science and Intelligent Data Engineering. Vol.7751, PP.82–89.
- [18] Cortes, C., and Vapnik, V. (1995). "Support-vector networks". Machine Learning. Vol.20, PP.273–297.
- [19] Vapnik, V., and Lerner, A. (1963). "Pattern recognition using generalized portrait method". Automation and Remote Control. Vol.24, PP.774–780.
- [20] Fletcher, T. (2008). "Support vector machines explained". Tutorial, University college London.
- [21] Luts, J., Fabian Ojeda, F., Plas, R., Moor, B., Huffel, S. and Suykens, J. A. K. (2010). "A tutorial on support vector machine-based methods for classification problems in chemometrics". AnalyticaChimicaActa. Vol.665, PP.129–145.
- [22] Cervantes, J., Lia, X., Yu, W. and Li, K. (2008). "Support vector machine classification for large data sets via minimum enclosing ball clustering". Neurocomputing. Vol.71, PP.611–619.
- [23] Dorigo, M. (1992). "Optimization, learning and natural algorithms", Ph.D Thesis, Dipartimento di Elettronica, Politecnico di Milano, Italy.
- [24] Bonabeau, E., Dorigo, M. and Theraulaz, G. (1999). "Swarm Intelligence: from natural to artificial systems", Oxford University Press, Inc., New York.
- [25] Mohamed Jafar, O. A. and Sivakumar, R. (2010). "Ant-based Clustering Algorithms: A Brief Survey". International Journal of Computer Theory and Engineering. Vol.2, No.5, PP.787-796.
- [26] Deneubourg, J.-L., Gross, S., Franks, N., Sendova-Franks, A., Detrain, C. and Chretien, L. (1991). "The dynamics of collective sorting: Robot-like ants and ant-like robots". In Proceedings of the First International Conference on Simulation of Adaptive Behavior: From Animals to Animats, Cambridge, MA, MIT Press. PP. 356-363.
- [27] Subramanya, A. and Talukdar, P. P. (2012). "A Tutorial on Graph-based Semi-Supervised Learning Algorithms for NLP". Tutorial, ACL, South Korea.
- [28] Akdis, A. C. and Agache, I. (2013). "Global atlas of asthma". Published by the European Academy of Allergy and Clinical Immunology.
- [29] Gregg, I., and Nunn, A. J. (1973). "Peak Expiratory Flow in Normal Subjects". British Medical Journal. Vol.3, PP.282-284.
- [30] Afify, H. A. (2011). "Evaluation of change detection techniques for monitoring land-cover changes: A case study in new Burg El-Arab area". Alexandria Engineering Journal. Vol.50, PP.187–195.