
بسم الله الرحمن الرحيم

دانشکده مهندسی برق و رباتیک

گروه رباتیک

عنوان پایان نامه ارشد

# کاهش ویژگی تصاویر دریافتی توسط موبایل ربات با استفاده از الگوریتم سیستم ایمنی مصنوعی

**دانشجو :**

**مریم سادات هاشمی‌پور**

**استاد راهنما :**

**دکتر سید علی سلیمانی**

**پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد**

بهمن ۱۳۹۳

شماره : ۱۲۴۴/آ.ت.ب
تاریخ : ۹۳/۱۱/۰۸
ویرایش : ــــــ

**بسمه تعالی**

مدیریت تحصیلات تکمیلی
فرم شماره (۶)

## فرم صورتجلسه دفاع پایان نامه تحصیلی دوره کارشناسی ارشد

با تأییدات خداوند متعال و با استعانت از حضرت ولی عصر ( عج ) جلسه دفاع از پایان نامه کارشناسی ارشد خانم / آقای :

**مریم سادات هاشمی پور** رشته : **بـــرق** گرایش : **رباتیک**

تحت عنوان : **کاهش ویژگی تصاویر دریافتی توسط موبایل ربات با استفاده از الگوریتم سیستم ایمنی مصنوعی**

که در تاریخ **۹۳/۱۱/۰۸** با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح زیر است :

قبول ( با درجه : بسیار خوب امتیاز ۱۸/۵۷ ) ☑ دفاع مجدد ☐ مردود ☐

۱ـ عالی ( ۲۰ ـ ۱۹ )
۲ـ بسیار خوب ( ۱۸/۹۹ ـ ۱۸ )
۳ـ خوب ( ۱۷/۹۹ ـ ۱۶ )
۴ـ قابل قبول ( ۱۵/۹۹ ـ ۱۴ )
۵- نمره کمتر از ۱۴ غیر قابل قبول

| عضو هیأت داوران | نام ونام خانوادگی | مرتبهٔ علمی | امضـــاء |
|---|---|---|---|
| ۱ـ استاد راهنما | سید علی سید صالحی | دانشیار | |
| ۲ـ استاد مشاور | ــــ | ــــ | ــــ |
| ۳ـ نماینده شورای تحصیلات تکمیلی | محمد مهدی زاده | استادیار | |
| ۴ـ استاد ممتحن | سید رضا سید مرتضی | استادیار | |
| ۵ ـ استاد ممتحن | [illegible] | [illegible] | |

رئیس دانشکده :

تقدیم به؛

خانواده عزیزم که حرمت لحظه لحظه دعایشان تاب و توانم را فزونی بخشید و مهر آسمانیشان،

آرامش را در وجودم جاری ساخت. . .

## سپاس:

از کسانی که به احترام "نون و القلم" ایستادن و اندیشه ورزی را با اشتیاق و قدم به قدم به من و امثال من آموختند.

از اساتید و مربیان خوبم که حوصله کردند و نجوای امید به آینده روشن را دم به دم زمزمه کردند تا گام هایم را محکم و با اعتماد به نفس بردارم و اگرچه کوچک ولی خود را سهیم در آبادانی وطنم بدانم.

کمال تقدیر و تشکر از استاد گرانقدر جناب آقای دکتر سید علی سلیمانی که با حسن خلق همواره راهنما و راه گشای من بودند و این نهال در حال رشد و بالندگی را هنرمندانه پرورانند تا راه کمال را طی کند.

در پایان از تمام کسانی که به "اندیشه" می اندیشند و به من اندیشیدن را آموختند سپاسگزارم.

باشد که یکتای بی همتا افتخار خدمتی سرشار از نشاط و امید همسو با علم و پژوهش جهت شکوفایی ایران زیبا را بر من عنایت فرماید.

# تعهد نامه

اینجانب **مریم سادات هاشمی‌پور** دانشجوی دوره کارشناسی ارشد رشته **مهندسی رباتیک** دانشکده **مهندسی برق و رباتیک** دانشگاه صنعتی شاهرود نویسنده پایان نامه **کاهش ویژگی تصاویر دریافتی توسط موبایل ربات با استفاده از الگوریتم سیستم ایمنی مصنوعی** تحت راهنمائی **دکتر سید علی سلیمانی** متعهد می شوم.

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده ( یا بافتهای آنها ) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

**تاریخ**

**امضای دانشجو**

## مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه های رایانه ای، نرم افزار ها و تجهیزات ساخته شده است ) متعلق به دانشگاه صنعتی شاهرود می باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.

# چکیده

امروزه با توجه به رشد روزافزون جمع‌آوری اطلاعات و قابلیت‌های موجود در ذخیره سازی، توجه بسیار زیادی به مسأله انتخاب، استخراج و کاهش ویژگی شده است. در سال‌های اخیر الگوریتم سیستم ایمنی مصنوعی بر پایه انتخاب کلونال به دلیل خاصیت بهینه‌سازی و تکامل‌گرا بودن بسیار مورد توجه قرار گرفته است. در صورت انتخاب تابع مناسب برای محاسبه میل ترکیبی در الگوریتم سیستم ایمنی مصنوعی و انجام عمل فراجهش و کلونی‌سازی بر مبنای میل ترکیبی در الگوریتم انتخاب کلونال، این روش نتیجه مطلوبی را در زمینه کاهش ویژگی ارائه می‌دهد.

در این تحقیق یک بردار ویژگی به ازای تک تک پیکسل‌های تصویر دریافت شده توسط موبایل ربات استخراج شده است و برای تشخیص پیکسل‌های جاده و غیرجاده از ماشین بردار پشتیبان استفاده شده است. سپس با الگوریتم سیستم ایمنی مصنوعی و ماشین بردار پشتیبان با کرنل RBF به عنوان تابع ارزیابی، عمل کاهش بردارهای ویژگی استخراج شده انجام شده است. جهت افزایش دقت کلاسه‌بند SVM، در کنار عمل کاهش ویژگی به صورت همزمان پارامترهای SVM و کرنل RBF نیز بهینه می‌شوند. با استفاده از الگوریتم انتخاب کلونال تطبیقی در الگوریتم سیستم ایمنی مصنوعی زمان اجرا کاهش یافته است در نهایت روش پیشنهادی روی مجموعه داده‌های UCI و مجموعه داده استخراج شده از تصویر موبایل ربات تست شده است. در این تحقیق به میانگین دقت ۹۴/۲۶ به ازای میانگین ۶۸/۶۲ درصد کاهش ویژگی در مجموعه داده‌های UCI و دقت ۸۳/۹۷ به ازای ۶۴/۹۰ درصد کاهش ویژگی در مجموعه داده استخراج شده از تصاویر موبایل ربات دست یافتیم.

**کلمات کلیدی:** الگوریتم انتخاب کلونال، الگوریتم انتخاب کلونال تطبیقی، الگوریتم سیستم ایمنی مصنوعی، استخراج بافت، کاهش ویژگی، ماتریس هم‌رخداد سطح خاکستری، ماشین بردار پشتیبان.

## فهرست مطالب

**عنوان** **صفحه**

## فهرست اصطلاحات اختصاری

| عنوان | علامت اختصاری |
|---|---|
| Biological Immune System | BIS |
| Artificial Immune System | AIS |
| Artificial Immune System based on Clonal Selection | AISCS |
| Artificial Immune System based on Adaptive Clonal Selection | AISACS |
| Antigen | AG |
| Antibody | AB |
| Genetic Algorithm | GA |
| Support Vector Machine | SVM |
| Support Vector Regression | SVR |
| Radial Basis Function | RBF |
| One Versus One | OVO |
| One Versus All | OVA |
| Artificial Neural Network | ANN |
| Principal Component Analysis | PCA |
| Discrete Wavelet Transform | DWT |
| Discrete Fourier Transform | DFT |
| Hue – Saturation-Value | HSV |
| Gray Level Co-Occurrence Matrix | GLCM |

## فهرست جدول‌ها

**عنوان** **صفحه**

## فهرست شکل‌ها

**عنوان** **صفحه**

# فصل ۱ مقدمه

## ۱-۱ پیشگفتار

با پیشرفت‌های چشمگیر در جمع‌آوری، قابلیت‌های ذخیره سازی و انتقال داده‌ها در طی دهه‌های اخیر شاهد افزایش حجم داده‌ها جهت کاربردهای داده‌کاوی و کشف دانش هستیم. این حجم وسیع داده الزاماً بیانگر ارزش آن‌ها نمی‌باشد چرا که بسیاری از این داده‌ها یا بدون استفاده‌اند یا بار اطلاعاتی خاصی ندارند. در کنار افزایش تعداد نمونه‌های داده جمع‌آوری شده، مسأله افزایش ابعاد داده نیز وجود دارد. در روش‌های یادگیری ماشین، تجزیه و تحلیل داده‌ها و کلاسه‌بندی کار با داده‌ها با ابعاد بالا دشوار است و مشکلاتی را در هر دو روش یادگیری باناظر و یادگیری بدون ناظر ایجاد می‌کند. برای حل این مشکل روش‌های کاهش ویژگی[۱] بسیار مناسب هستند. تعداد ویژگی‌های زیاد حاوی اطلاعات مفید نیستند چرا که ممکن است نامرتبط با کلاس داده‌ها باشند به خصوص اگر تعداد ویژگی‌ها زیاد و تعداد نمونه‌ها کم باشند در نتیجه فضای جستجو کم جمعیت خواهد بود و مدل قادر نیست بین داده-های مرتبط و نویز تمایز قائل شود.

در مسأله کاهش ویژگی همواره به دنبال روشی هستیم که در کنار حذف ویژگی‌های نامرتبط و اضافی، کاهش فضای ذخیره‌سازی و منابع محاسباتی، کارایی و دقت کلاسه‌بند را تا حد ممکن حفظ کند. یکی از راه‌های مناسب در مسأله کاهش ویژگی استفاده از الگوریتم‌های تکاملی[۲] در کنار روش-های مناسب برای استخراج ویژگی‌ها با توجه به کاربرد مورد نظر می‌باشد. از جمله می‌توان به الگوریتم ژنتیک[۳]، الگوریتم تقسیم و محدود کردن، الگوریتم جستجوی ترتیبی[۴]، جستجوی ممنوعه[۵]، کلونی مورچگان[۶]، بهینه‌سازی توده ذرات[۷]، الگوریتم سیستم ایمنی مصنوعی[۸] و ... اشاره نمود.

[1] Feature Reduction
[2] Evolutionary Algorithm
[3] Genetic Algorithm (GA)
[4] Sequential Search
[5] Tabu Search
[6] Ant Colony
[7] Particle Swarm Optimization (PSO)
[8] Artificial Immune System (AIS)

## ۱-۲ انگیزه

روش‌های مختلف استخراج ویژگی[1] داده‌ها را در ابعاد مختلف تولید می‌کنند، پردازش داده‌ها در ابعاد بالا مشکلاتی را در پی دارد. لذا مسأله کاهش ابعاد به یک رویکرد قابل توجه برای محققان تبدیل شده است. تاکنون کارهای متفاوتی برای کاهش ویژگی با استفاده از الگوریتم‌های تکاملی صورت گرفته است و هم اکنون نیز تحقیقات و روش‌های مختلفی جهت بهبود و ارائه راه‌حل‌های نوین در دست اجرا است که هر یک دارای مزایا و محدودیت‌هایی می‌باشند. هدف اصلی در این روش‌ها حفظ دقت و کارایی کلاسه‌بند در کنار کاهش ویژگی می‌باشد. سیستم‌های کاهش ابعاد فعلی، با تبدیلات خطی و غیر خطی سر و کار دارند و اگر چه از دقت بالایی برخوردارند اما ماهیت داده‌ها را تغییر می‌دهند و آن‌ها را به گونه‌ای غیر قابل تفسیر برای کاربر نمایان می‌سازند. از این رو وجود سیستم‌هایی که بتوانند سرعت و دقت را تواماً به همراه داشته و همچنین کاربران را با تفسیر اطلاعات جدید دچار مشکل ننماید احساس می‌شود.

برای این منظور الگوریتم‌های تکاملی بسیاری نظیر الگوریتم ژنتیک، کلونی مورچگان و بهینه‌سازی ذرات استفاده شده‌اند. یکی از روش‌های قابل استفاده الگوریتم سیستم ایمنی مصنوعی است. سیستم های ایمنی مصنوعی روش محاسبه نرم بر مبنای تئوری‌ها، ایده‌ها و اجزای سیستم ایمنی مهره‌داران هستند. در این تحقیق به اثبات می‌رسد که با به کارگیری قابلیت این الگوریتم می‌توان با ارزش‌ترین زیر مجموعه را از بین داده‌های موجود انتخاب کرد، به گونه‌ای که دقت و سرعت را در حد قابل قبولی نگاه داشت. با توجه به نتایج حاصل شده از الگوریتم ایمنی مصنوعی در این زمینه، مشاهده می‌شود این روش عملکرد مطلوبی ارائه می‌دهد یعنی هم تعداد ویژگی‌های کاهش یافته بیشتر شده و هم دقت کلاسه‌بند افزایش یافته است. در الگوریتم‌های مختلف تکاملی برای یافتن زیرمجموعه مناسب از ویژگی‌ها از یک تابع برازندگی[2] استفاده می‌شود. انتخاب نوع، پارامترها و آموزش مناسب این تابع در

[1] Feature Extraction
[2] Fitness Function

انتخاب ویژگی‌های مناسب بسیار مؤثر است.

در این تحقیق از کلاسه‌بند ماشین بردار پشتیبان[1] به عنوان تابع برازندگی استفاده شده است. ماشین بردار پشتیبان روش مناسب و مورد اعتمادی است که کارایی فوق‌العاده بالایی در طبقه‌بندی الگو دارد و نتایج بسیار بهتری نسبت به سایر روش‌های مشابه ارائه می‌دهد.

## ۱-۳ اهداف پایان‌نامه

در این تحقیق با استفاده از روش معرفی شده توسط هارالیک[2] در[۱] و ماشین بردار پشتیبان ویژگی‌های تصویر دریافت شده توسط موبایل ربات را برای تشخیص جاده استخراج کرده سپس از الگوریتم کاهش ویژگی مبتنی بر سیستم ایمنی مصنوعی برای انتخاب زیر مجموعه بهینه ویژگی‌ها استفاده می‌کنیم. برای محاسبه میل ترکیبی[3] در الگوریتم AIS، از ماشین بردار پشتیبان با تابع کرنل پایه شعاعی[4] استفاده شده است. همچنین قصد داریم که در کنار کاهش ویژگی، پارامترهای کرنل RBF در SVM را نیز بهینه کنیم تا کارایی ماشین بردار پشتیبان افزایش یافته و نتیجه مناسب‌تری به دست آوریم. در این تحقیق تلاش شده است هم با استفاده از الگوریتم انتخاب کلونال به صورت تطبیقی در کنار حفظ دقت الگوریتم اصلی همگرایی را در الگوریتم سیستم ایمنی مصنوعی افزایش داده و هم با استفاده از تعریف تابعی متناسب با میل ترکیبی برای فراجهش[5] کارایی الگوریتم AIS را بهبود دهیم.

جهت بررسی صحت الگوریتم پیشنهادی از مجموعه داده‌های پایگاه داده UCI[6] استفاده شده است و برای مقایسه کارایی آن از الگوریتم ژنتیک با پارامترهای مشابه بهره گرفته شده است.

[1] Support Vector Machine (SVM)
[2] Haralick
[3] Affinity
[4] Radial Basis Function Kernel (RBF)
[5] Haypermutation
[6] UC Irvine Machine Learning Repository Databace

## ۱-۴ مروری بر تحقیقات و مطالعات انجام شده

در این بخش به بررسی تعدادی از تحقیقات که در سال‌های اخیر در زمینه انتخاب ویژگی[1] بر روی داده‌های مختلف استاندارد برای کاربردهای مختلف انجام شده است، می‌پردازیم. در بسیاری از این تحقیقات فقط کاهش ویژگی انجام می‌گیرد و تنها تعداد کمی به بهینه‌سازی پارامترهای SVM نیز می‌پردازند.

۱) در سال ۲۰۰۲ در مرجع [۲] از شبکه عصبی مصنوعی[2] و الگوریتم ژنتیک در تشخیص خطا در چرخش ماشین آلات استفاده شده و روش‌های آماری مرتبه دوم برای پیش‌پردازش سیگنال ارتعاش به کار رفته است. در برخی از آزمایش‌ها، نتایج با الگوریتم ژنتیک و ماشین بردار پشتیبان نیز محاسبه شده است در هر دو روش ANN و SVM از کرنل RBF استفاده شده و برای انتخاب پارامتر $\sigma$ در این کرنل انحراف معیار پنج ویژگی ورودی اول محاسبه شده است.

۲) در مرجع معرفی شده در سال ۲۰۰۲ [۳] ، از دقت ماشین بردار پشتیبان با کرنل خطی به عنوان تابع ارزیابی در الگوریتم ژنتیک جهت انتخاب ویژگی و بهینه‌سازی پارامتر جریمه $C$، استفاده شده است. این روش تعداد ویژگی‌هایی که در هر مجموعه داده می‌خواهند انتخاب کند را ثابت در نظر گرفته و به جای استفاده از روش k-Fold Cross Validation ، از SVM با اضافه کردن چند قید تئوری جدید بهره برده است.

۳) در مرجع [۴] که در سال ۲۰۰۶ معرفی شده است، از الگوریتم ژنتیک و ماشین بردار پشتیبان با کرنل RBF به عنوان تابع برازندگی جهت کاهش بردار ویژگی و بهینه سازی پارامترهای کرنل RBF در SVM استفاده شده است. در این روش برای آموزش و تست داده-ها در ماشین بردار پشتیبان کتابخانه LibSVM به کار گرفته شده است. نتایج حاصل از این

[1] Feature Selection
[2] Artificial Neural Network (ANN)

روش با استفاده از مجموعه داده‌های پایگاه داده UCI بررسی شده و با رویکرد جستجوی شبکه[1] در انتخاب پارامترهای پنالتی $\sigma$ و پارامتر ثابت حاشیه نرم $C$ در کرنل RBF مقایسه شده است. نتایج به دست آمده نشان دهنده کارایی مناسب این الگوریتم نسبت به رویکرد جستجوی شبکه در کاهش ویژگی و حفظ دقت کلاسه‌بند می‌باشد. تعداد اعضای اولیه جمعیت ۵۰۰، نرخ Crossover برابر با ۰/۷۵، نرخ جهش ۰/۰۲ و چرخه انتخاب رولت[2] استفاده شده است.

۴) در سال ۲۰۰۸ در مرجع [۵]، عمل کاهش ویژگی و بهینه‌سازی پارامترهای ماشین بردار پشتیبان با کرنل RBF توسط الگوریتم بهینه سازی توده ذرات انجام شده است. نتایج حاصل از این روش با مرجع [۴] مقایسه شده است که در نهایت در برخی از مجموعه داده‌ها روش GA+SVM و در برخی دیگر روش PSO+SVM عملکرد بهتری ارائه کرده است. در این مرجع از روش 10-Fold Cross validation برای ارزیابی استفاده شده است.

۵) در سال ۲۰۰۸، از ترکیب الگوریتم‌های PSO و SVM برای بهبود دقت کلاسه‌بند در انتخاب زیرمجموعه کوچکی از ویژگی‌ها استفاده شده است [۶]. که پارامترهای $\sigma$ و $C$ کرنل RBF را نیز به عنوان ویژگی‌های ورودی در نظر گرفته و این پارامترها را نیز بهینه می‌کند. تمامی مقادیر ورودی نرمال‌سازی شده و در بازه [0,1] قرار دارند. نتایج حاصل از این روش بر روی یک مجموعه داده شامل ۳۰ ویژگی و ۸ کلاس بررسی شده است.

۶) در روشی که در سال ۲۰۰۹ معرفی شده است [۷] یک رویکرد ترکیبی از الگوریتم ژنتیک و ماشین بردار پشتیبان HGASVM برای کلاسه‌بندی مدلاسیون دیجیتال استفاده گردید. در این مرجع تبدیل موجک گسسته[3] و موجک آنتروپی تطبیقی[4] در مرحله استخراج ویژگی به کار گرفته شده‌اند. تابع کرنل بهینه، پارامترهای کرنل، فیلتر موجک و پارامترهای آنتروپی

[1] Grid Search
[2] Roulette Wheel Selection
[3] Discrete Wavelet Transform (DWT)
[4] Adaptive Wavelet Entropy

موجک نیز به صورت بهینه تعیین می‌شوند.

۷) در روشی که در مرجع [۸] ارائه شده است، در کنار کاهش ویژگی با الگوریتم ژنتیک و بهینه-سازی همزمان پارامترهای SVM ، انتقالی با استفاده از روش PCA[1] را نیز در بردار ویژگی انجام می‌دهد. به این صورت که کروموزوم اولیه شامل چهار قسمت ( قسمت اول برای بیت-هایی برای پارامتر $C$ کرنل RBF، قسمت دوم بیت‌هایی برای پارامتر $\sigma$ کرنل RBF، قسمت سوم یک بیت برای بررسی این که روش PCA اعمال شود یا خیر و قسمت چهارم بیت‌هایی برای ویژگی‌ها) می‌باشد. در این روش نرخ Crossover برابر با ۰/۷۵، نرخ جهش برابر با ۰/۰۲ و تعداد اعضای جمعیت ۱۰۰ در نظر گرفته شده است. در نهایت این روش با یک SVM ساده با پارامترهای $C=1$ و $\sigma=0.5$ مقایسه شده است.

۸) مرجع [۹] در سال ۲۰۰۹، نشان دهنده مسأله کاهش ویژگی به کمک الگوریتم کلونال و بهینه‌سازی پارامترهای ماشین بردار پشتیبان می‌باشد. نتایج حاصل از این روش با الگوریتم ژنتیک مقایسه شده است. در این روش از نرخ ثابت ۰/۰۱ برای فراجهش استفاده می‌گردد. تعداد کلون‌هایی که در هر مرحله تولید می‌شوند برابر با تعداد ثابت ۱۰۰ در هر تکرار است. تعداد اعضای جمعیت ۱۰۰ و تعداد تکرارهای الگوریتم ۲۰ تعیین شده است. برای آموزش SVM از روش 10-Fold Cross Validation استفاده شده است. نتایج به دست آمده نشان دهنده قدرت این الگوریتم در کاهش تعداد بیشتری از ویژگی‌ها و افزایش دقت کلاسه‌بند نسبت به روش پیاده‌سازی شده با الگوریتم ژنتیک می‌باشد.

۹) در مرجع [۱۰] که در سال ۲۰۰۸ انجام شده است، یک روش مبتنی بر کلاسه‌بندی برای تجزیه و تحلیل سیگنال‌های ضبط شده معرفی گردید. در این روش چندین مجموعه ویژگی از جمله ویژگی‌های ریخت‌شناسی، ویژگی‌های فرکانس، ویژگی‌های آنتروپی و انرژی موجک استخراج می‌گردد سپس این بردار ویژگی‌ها به کمک سیستم ایمنی مصنوعی تکاملی کاهش

[1] Principal Component Analysis (PCA)

می‌یابند. در این مقاله از ماشین بردار پشتیبان با کرنل RBF به عنوان ابـزاری جهـت انـدازه-گیری میل ترکیبی میان آنتی‌بادی و آنتی‌ژن اسـتفاده شـده اسـت کـه همزمـان پارامترهـای SVM نیز بهینه می‌شوند. نتایج حاصل از این روش با الگوریتم ژنتیک مقایسه شده است.

۱۰)در مرجع [۱۱]، یک روش کاهش ویژگی بر اساس الگوریتم ClonalG معرفی شده اسـت کـه میزان میل ترکیبی بین آنتی‌بادی و آنتی‌ژن با کمک ماشین بردار پشتیبان با استفاده از کرنل خطی اندازه گرفته شده است. نتایج به دست آمده از ایـن تحقیـق نشـان مـی‌دهـد الگـوریتم ژنتیک از یک مرحله به بعد در کمینه محلی می‌افتـد و الگـوریتم ClonalG عملکـرد بهتـری دارد. در این روش نرخ تکثیر کلون‌ها مقدار ثابت ۱۰ و نرخ جهـش مقـدار ثابـت ۰/۱ در نظـر گرفته شده است.

## ۱-۵ ساختار پایان‌نامه

پس از مقدمه و مروری بر مطالعات پیشین، در فصـل دوم ایـن پایـان نامـه مباحـث اسـتخراج و انتخاب ویژگی بررسی می‌شوند.

در فصل سوم مفاهیم مربوط به سیستم ایمنی طبیعـی بـه اختصـار شـرح داده شـده و در ادامـه الگوریتم سیستم ایمنی مصنوعی و کاربردهای آن مورد بررسی قرار خواهد گرفت.

در فصل چهارم به بررسی کلاسه‌بند ماشین بردار پشتیبان خطی و غیرخطی می‌پردازیم و عملکرد SVM را در طبقه‌بندی‌های چند کلاسه تشریح کرده و به بررسی کرنل‌های مختلـف قابـل اسـتفاده در آن می‌پردازیم.

فصل پنجم به تشریح الگوریتم پیشنهادی جهت استخراج و انتخاب ویژگـی اختصـاص داده شـده است.

در فصل ششم نتایج شبیه‌سازی ساختار پیشنهاد شده در فصل پنجم را برای ویژگی‌های اسـتخراج شده از تصاویر موبایل ربات و مجموعه داده‌های پایگاه UCI ارزیابی کرده و در انتهـا بـه بررسـی کلـی

مطالب می‌پردازیم.

فصل هفتم به جمع بندی و نتیجه‌گیری از کار انجام شده به همراه پیشنهاد برای ادامه کاراختصاص یافته است.

# فصل۲ استخراج و انتخاب ویژگی

## ۲-۱ مقدمه

پیشرفت‌های به وجود آمده در جمع‌آوری داده و قابلیت‌های ذخیره‌سازی در طی دهه‌های اخیر باعث شده در بسیاری از علوم با حجم بزرگی از اطلاعات روبه‌رو شویم. محققان در زمینه‌های مختلف مانند مهندسی، ستاره‌شناسی، زیست‌شناسی و اقتصاد هر روز با مشاهدات بیشتر و بیشتری مواجه می‌شوند. در مقایسه با بسترهای داده‌ای قدیمی و کوچکتر، بسترهای داده‌ای امروزی چالش‌های جدیدی در تحلیل داده‌ها به وجود آورده‌اند. روش‌های آماری سنتی به دو دلیل امروزه کارائی خود را از دست داده‌اند. علت اول افزایش تعداد مشاهدات[۱] است، و علت دوم که از اهمیت بالاتری برخوردار است افزایش تعداد متغیرهای مربوط به یک مشاهده می‌باشد. تعداد متغیرهایی که برای هر مشاهده باید اندازه‌گیری شود ابعاد داده نامیده می‌شود. عبارت "متغیر[۲]" بیشتر در آمار استفاده می‌شود در حالی که در علوم کامپیوتر و یادگیری ماشین بیشتر از عبارات "ویژگی[۳]"و یا "صفت[۴]" استفاده می‌گردد [۱۲].

## ۲-۲ مشکلات کار با داده‌ها در ابعاد بالا

۱) در بیشتر مواقع تمام ویژگی‌های داده‌ها برای یافتن دانشی که در داده‌ها نهفته است مهم و حیاتی نیستند.

۲) تعداد زیاد ویژگی‌ها موجب افزایش نویز داده‌ها شده و خطای الگوریتم یادگیری را افزایش می‌دهد.

۳) هزینه محاسباتی برنامه‌های داده‌کاوی با افزایش بعد بالا می‌رود.

[1] Observations
[2] Variable
[3] Feature
[4] Attribute

## ۲-۳ اهداف کلی کاهش ابعاد

۱) ساده‌سازی مدل: تعداد ویژگی‌های زیاد پیچیدگی مدل را افزایش می‌دهد. برخی ویژگی‌ها زیاد ( ویژگی‌های همبسته[1]) یا نامرتبط[2] ( ویژگی‌هایی که هیچ اطلاعاتی در مورد مدل خروجی ندارند ) هستند. حذف این ویژگی‌ها مدل را ساده‌تر می‌کند و باعث درک بهتر مدل می‌گردد.

۲) اصلاح کیفیت مدل: حذف ویژگی‌های زائد نه تنها کیفیت مدل را بدتر نمی‌کند بلکه گاهی آن را اصلاح می‌کند. به عنوان مثال حذف ویژگی‌ها با مقادیر تصادفی که به دلیل نویز ایجاد شده‌اند معمولاً باعث عملکرد بهتر مدل می‌گردد.

۳) بهبود در تعمیم: مدل‌های ساده‌تر با تعداد کمتر پارامتر توانایی تعمیم بالاتری دارند.

۴) کاهش زمان محاسباتی: مدلی که از تعداد ویژگی‌های کمتری استفاده می‌کند سریع‌تر یاد می‌گیرد و عمل می‌کند.

۵) صرفه‌جویی در حافظه سیستم متناسب با ویژگی‌های حذف شده.

۶) کاهش هزینه‌های جمع‌آوری، ارتباطات و نگهداری داده‌های مدل.

۷) تجسم داده‌ها زمانی‌که تعداد ویژگی‌ها کم است مثلاً (۱، ۲ یا ۳ بعد) [۱۳].

روش‌های کاهش ابعاد داده به دو دسته تقسیم می‌شوند:

۱) روش‌های مبتنی بر استخراج ویژگی: این روش‌ها یک فضای چند بعدی را به یک فضا با ابعاد کمتر نگاشت می‌دهند. یعنی با ترکیب مقادیر ویژگی‌های موجود تعداد کمتری ویژگی به‌وجود می‌آورند به‌طوری‌که ویژگی‌های جدید دارای تمام یا بخش اعظمی از اطلاعات موجود در ویژگی‌های اولیه باشند.

۲) روش‌های مبتنی بر انتخاب ویژگی: این روش‌ها با انتخاب زیرمجموعه‌ای از ویژگی‌های اولیه،

[1] Correlated Feature
[2] Irrelevant

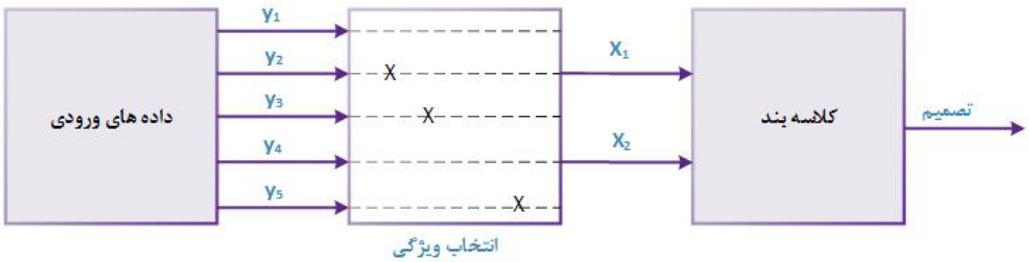
ابعاد داده‌ها را کاهش می‌دهند. گاهی تحلیل‌های داده‌ای نظیر کلاسه‌بندی بر روی فضای کاسته شده نسبت به فضای اصلی عملکرد بهتری را دارند.

## ۲-۴ استخراج ویژگی

در مبحث شناسایی آماری الگو، استخراج و انتخاب ویژگی یافتن یک روش ریاضی برای کاهش ابعاد الگوها است. انتخاب ویژگی‌ها، صفت‌ها و اندازه‌ها تأثیر مهمی بر دقت کلاسه‌بند، زمان مورد نیاز برای کلاسه‌بندی، تعداد نمونه‌ها برای یادگیری و هزینه انجام کلاسه‌بندی دارند. مدلی از سیستم شناسایی الگو که شامل مراحل استخراج و انتخاب ویژگی‌ها می‌باشد در شکل ۲-۱ نمایش داده شده است.


شکل ۲-۱: مدل سیستم شناسایی الگو شامل مراحل استخراج و انتخاب ویژگی‌ها

اطلاعات ورودی در معرض فرآیند انتخاب و استخراج ویژگی برای تعیین بردار ورودی قرار می‌گیرند و به کلاسه‌بند وارد می‌شوند سپس تصمیمی در مورد کلاس بردار الگوی ورودی توسط کلاسه‌بند گرفته می‌شود. کاهش بعد بر اساس استخراج و انتخاب ویژگی انجام می‌گیرد. در شکل ۲-۲ کاهش بعد بر اساس انتخاب ویژگی نمایش داده شده است.

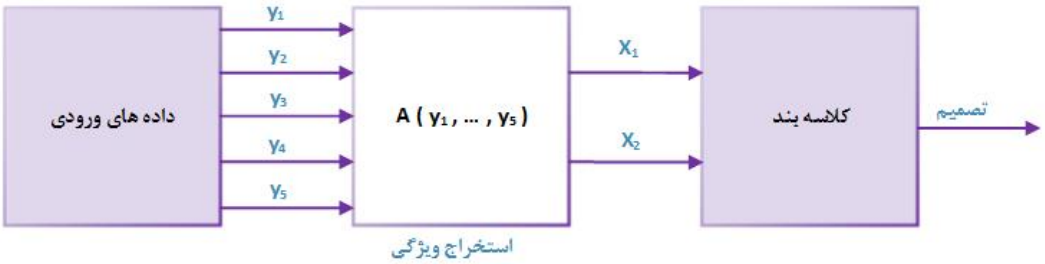



شکل ۲-۲: کاهش بعد بر اساس انتخاب ویژگی

همان‌طور که در شکل ۲-۲ مشاهده می‌شود کاهش بعد بر اساس انتخاب ویژگی بر پایه حذف آن دسته از ویژگی‌هایی است که در تعیین کلاس تأثیرگذار نیستند. به عبارت دیگر ویژگی‌های زائد[1] و نامرتبط نادیده گرفته می‌شود.

[1] Redundant

در استخراج ویژگی مانند شکل ۲-۳، تمامی اطلاعات در نظرگرفته شده و این اطلاعات مفید به یک فضای ویژگی با ابعاد کمتر نگاشت می‌یابد. در مبحث استخراج ویژگی تابع نگاشت ($A$ درشکل ۲-۳) از قبل تعیین می‌گردد [۱۴].




شکل ۲-۳: کاهش بعد بر اساس استخراج ویژگی

## ۲-۴-۱ روش‌های مبتنی بر استخراج ویژگی

در این تحقیق جهت ایجاد یک مجموعه داده برچسب‌گذاری شده نیاز به استخراج ویژگی از تصویر موبایل ربات وجود دارد. بردار ویژگی استخراج شده در این مرحله شامل ویژگی‌های بافت و ویژگی‌های رنگ تصویر در کانال HSV می‌باشد. بنابراین در ادامه به شرح روش‌های استخراج ویژگی-های مرتبط با بافت، از تصویر می‌پردازیم.

## ۲-۴-۲ استخراج ویژگی از تصاویر

امروزه تصاویر نقش بسیار با اهمیتی در زمینه‌های مختلف علمی، پزشکی، روزنامه‌نگاری، تبلیغات، آموزش و سرگرمی دارند بنابراین تجزیه و تحلیل تصاویر با کمک کامپیوتر به یک امر مهم در این زمینه‌ها تبدیل شده است. تجزیه و تحلیل تصاویر شامل قطعه‌بندی، انتقال، کلاسه‌بندی الگو و استخراج ویژگی می‌باشد [۱۵]. فرآیند یافتن اطلاعات سطح بالا مانند رنگ، شکل و بافت و پردازش تصویر به کمک این اطلاعات را استخراج ویژگی گویند.

### ۲-۴-۲-۱ بافت در تصاویر

مفهوم بافت در تصویر به راحتی قابل شناسایی می‌باشد در شکل ۲-۴ مثال‌هایی از ویژگی‌های

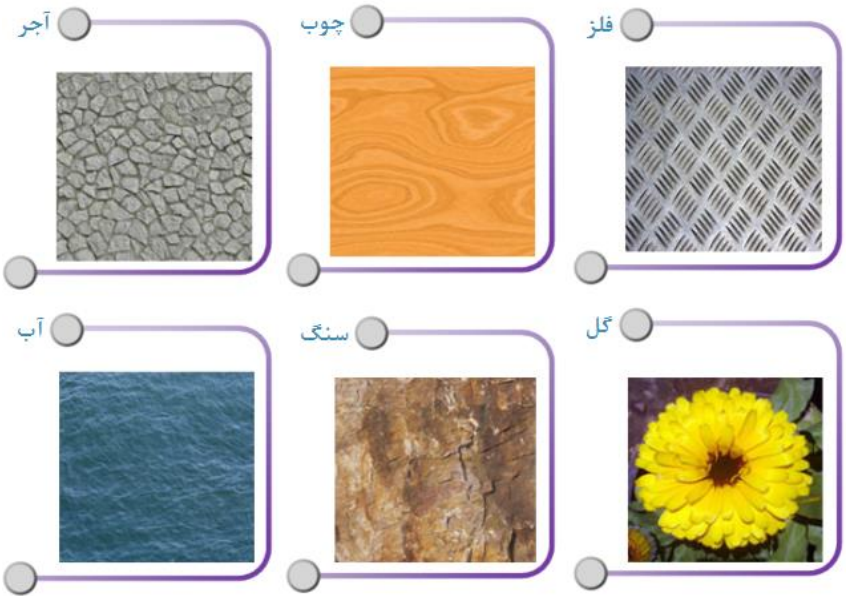
بافت نشان داده شده است اما تعریف دقیق آن کمی دشوار است چرا که محققان بینایی ماشین تعاریف متفاوتی در مورد بافت ارائه داده‌اند:

- بافت، الگوهای بصری با خواص همگن است که از قسمت‌هایی به یک رنگ مثل ابر یا آب بـه دست نمی‌آید [۱۶].
- یک ناحیه در تصویر، بافت ثابتی دارد اگر مجموعه آماری محلی یا سایر خواص محلی تصـویر ثابت باشند یا به آرامی تغییر کنند و یا به صورت متناوب تکرار شوند [۱۷].
- اصطلاح بافت از سه جز تشکیل شده است:

a. برخی از الگوهای منظم محلی که در یک ناحیه تصویر تکرار شده‌اند.

b. این نظم، آرایش غیرتصادفی در بخش‌های اصلی است.

c. بخش‌ها، موجودیت‌های یکنواختی هستند که با ابعاد تقریباً یکسانی در تمامی ناحیه‌هـای تصویر بافت وجود دارند [۱۸].




شکل ۲-۴: مثال هایی از ویژگی های بافت مختلف

### ۲-۴-۲-۲ الگوریتم‌های استخراج ویژگی بافت

تیوسریان[1] و جین[2] [۱۹] روش‌های مختلف استخراج ویژگی بافت را به چهار دسته اصلی به شرح زیر تقسیم کرده‌اند:

#### ۱) روش ساختاری

رویکرد ساختاری [۱] بافت تصویر را با تعریف شکل‌های هندسی اولیه (میکروبافت[3]) و یک سلسله مراتب از نظم فضایی (ماکروبافت[4]) شامل شکل هندسی اولیه نشان می‌دهد. از جمله مزایای روش ساختاری، تشریح نمادین تصویر است که برای تفسیر و تجزیه و تحلیل بسیار مناسب است. این روش به دلیل تغییرپذیری میکروبافت و ماکروبافت و نبود تمایز روشن بین آن‌ها در بافت‌های طبیعی، در زمینه تشخیص بافت‌های طبیعی عملکرد مناسبی ندارند.

#### ۲) روش آماری

روش‌های آماری بافت را به صورت غیرمستقیم به کمک خواص غیرقطعی که روابط بین سطوح خاکستری از یک تصویر را مدیریت می‌کنند، نشان می‌دهند. این روش یکی از اولین تکنیک‌ها در بینایی ماشین است و با محاسبه ویژگی‌های محلی در هر نقطه از تصویر و استخراج یک مجموعه آماری از توزیع ویژگی‌های محلی، در تجزیه و تحلیل مقادیر روشنایی تصویر به کار می‌رود. بر اساس تعداد پیکسل‌هایی که در ویژگی‌های محلی تعریف می‌شوند این روش به سه دسته تقسیم می‌شود: آمار مرتبه اول (یک پیکسل)، آمار مرتبه دوم (یک جفت پیکسل)، آمار مرتبه بالا (سه یا تعداد بیشتر از پیکسل‌ها). در آمار مرتبه اول، خواصی نظیر (میانگین و واریانس) یک پیکسل را با چشم‌پوشی از تعامل فضایی بین پیکسل‌های تصویر تخمین می‌زنند. اما در آمار مرتبه دوم و مرتبه بالا خواص دو یا تعداد بیشتری از پیکسل‌ها بر اساس موقعیت و ارتباطشان با یکدیگر تخمین زده می‌شوند [۲۰].

#### ۳) روش مبتنی بر مدل

[1] Tuceryan
[2] Jain
[3] Micro Texture
[4] Macro Texture

این روش‌ها یک تصویر را به صورت یک مدل احتمالی یا یک ترکیب خطی از مجموعه‌ای از توابع اصلی توصیف می‌کنند. دو گونه متفاوت از روش‌های مبتنی بر مدل شامل: مدل مبتنی بـر پیکسـل[1] و مدل مبتنی بر ناحیه می‌باشد. در مدل مبتنی بر پیکسل، یک تصویر به صورت مجموعه‌ای از پیکسل‌ها و در مدل مبتنی بر ناحیه، یک تصویر به صورت یک مجموعه از زیر الگوها بررسی می‌شود.

**۴) روش انتقال**

این روش با کمک تبدیلاتی نظیر تبدیل فوریه، گابور، تبدیل موجک، تصاویر را در حوزه فرکـانس تجزیه و تحلیل می‌کند. این روش برای پردازش کل تصویر به کار می‌رود و در کاربردهایی که بر اساس تجزیه و تحلیل قسمتی از تصویر ورودی هستند مناسب نمی‌باشد.

## ۲-۴-۳ ماتریس هم‌رخداد[2] و روش استخراج ویژگی بافت هارالیک

امروزه یکی از مهم‌ترین روش‌های تجزیـه و تحلیـل بافـت، روش آمـاری مرتبـه دوم اسـت کـه از ماتریس هم‌رخداد استفاده می‌کند. در سال ۱۹۷۹، آقای هارالیک [۱] روش استخراج ویژگی بافـت بـه کمک ماتریس هم‌رخداد را معرفی‌کرد، در این روش ابتدا GLCM محاسبه شده و سپس ویژگـی‌هـای بافت بر اساس این ماتریس استخراج می‌شوند.

### ۲-۴-۳-۱ ماتریس هم‌رخداد سطح خاکستری[3]

یکی از روش‌های تعیین بافت استفاده از سـطوح خاکسـتری تصـویر اسـت در روش هارالیـک بـا استفاده از ماتریس هم‌رخـداد سـطوح خاکسـتری و در نظرگـرفتن ارتبـاط بـین دو پیکسـل همسـایه ویژگی‌های بافت استخراج می‌شوند. در این روش پیکسل اول به عنوان پیکسل مرجع و دومین پیکسل به عنوان پیکسل همسایه شناخته می‌شود و تصویر با $G$ سـطح خاکسـتری بـه صـورت رابطـه (۲-۱)، نمایش داده می‌شود.

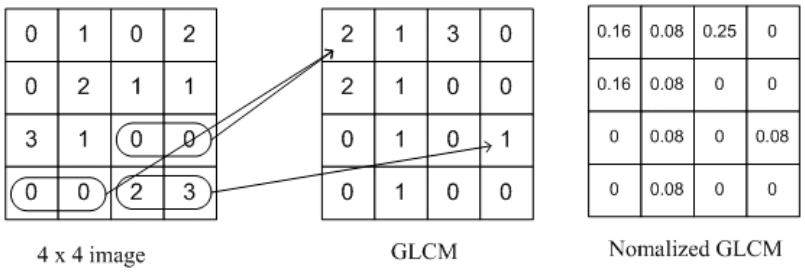
[1] Pixel-Based
[2] Co-Occurrence Matrix
[3] Gray-level Co-occurrence Matrix (GLCM)

$$\{I(x,y)\,, 0 \le x \le N_x - 1\ \,, 0 \le y \le N_y - 1\} \quad (۱-۲)$$

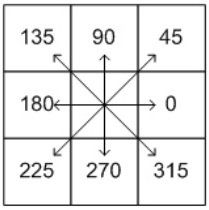
ماتریس هم‌رخداد سطح خاکستری $G \times G$ با عنوان $P_d^\theta$ با بردار جابه‌جایی $d = (dx, dy)$ و جهت $\Theta$، مانند رابطه (۲-۲) تعریف می‌گردد. عنصر $(i,j)$ از $P_d^\theta$ برابر است با: تعداد رخداد جفت سطح خاکستری $(i,j)$، به شرطی که فاصله $i$ از $j$ در جهت $\Theta$ برابر با $d$ باشد.

$$P_d^\theta(i,j) = \#\{((r,s),(t,v)) : I(r,s) = i\,, I(t,v) = j\}$$
$$Where\ \ (r,s),(t,v) \in N_x \times N_y\,; (t,v) = (r + dx\,, s + dy) \quad (۲-۲)$$

شکل ۲-۵، ماتریس هم‌رخداد $P_d^\theta$ با فاصله $d=1$ و جهت افقی $\theta = 0$ را نشان می‌دهد. ماتریس هم‌رخداد برای جهت‌های عمودی $(\theta = 90)$ و قطری $(\theta = 45\,, 135)$ نیز وجود دارد. اگر جهت‌ها از پایین به بالا و از چپ به راست در نظر گرفته شوند. هشت جهت ( 0, 45, 90, 135, 180, 225, 270, 315) مانند شکل ۲-۶ خواهیم داشت.




شکل ۲-۵: GLCM از تصویر 4*4 با فاصله d=1 و جهت Θ=0، [۲۰]




شکل ۲-۶: هشت جهت در GLCM [۲۰]

### ۲-۳-۴-۲ ویژگی‌های بافت هارالیک

هارالیک مشخصه‌های زیر را برای استخراج ویژگی‌های بافت از تصاویر بر اساس GLCM معرفی کرده است [۲۱].

### ۱) گشتاور دوم زاویه‌ای[1]

این مشخصه به عنوان یکنواختی[2] یا انرژی[3] نیز معرفی می‌شود. یکنواختی تصویر با کمک رابطه (۲-۳) اندازه گرفته می‌شود. زمانی‌که پیکسل‌ها خیلی شبیه به هم باشند، مقدار ASM بزرگ خواهد بود.

$$f_1 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P_d^{\theta}(i,j)^2 \qquad (۲-۳)$$

### ۲) تضاد[4]

تضاد یک اندازه از شدت یا تغییرات سطح خاکستری بین پیکسل مرجع و همسایه‌اش است. در ادراک بصری تضاد همان تفاوت در رنگ و روشنایی شی با سایر اشیا در میدان دید یکسان، است.

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P_d^{\theta}(i,j) \right\} \ , where \ n = |i-j| \qquad (۲-۴)$$

زمانی که $i$ و $j$ هم اندازه‌اند، سلول روی قطر قرار گرفته و $i-j=0$ است ( نشان‌دهنده تطابق کامل یک پیکسل با پیکسل همسایه‌اش می‌باشد)، در این حالت به آن پیکسل وزن صفر داده می‌شود. اگر $i$ و $j$ یک واحد با هم اختلاف داشته باشند، تضاد کمی وجود دارد و وزن یک است. اگر $i$ و $j$ به اندازه دو واحد با هم اختلاف داشته باشند، تضاد افزایش یافته و وزن چهار خواهد شد. همان‌طور که در رابطه (۲-۴) مشخص است، وزن به صورت نمایی رشد می‌کند.

### ۳) همبستگی[5]

همبستگی نشان‌دهنده وابستگی خطی مقادیر سطح خاکستری در ماتریس هم‌رخداد است. در حقیقت نشان می‌دهد که پیکسل مرجع چگونه به همسایه‌اش وابسته است. اگر حاصل رابطه (۲-۵) صفر باشد غیرهمبسته و اگر حاصل یک باشد همبستگی کامل وجود دارد. در این رابطه مقادیر

[1] Angular Second Moment (ASM)
[2] Uniformity
[3] Energy
[4] Contrast
[5] Correlation

$\mu_x$،$\mu_y$ و $\sigma_x$، $\sigma_y$ به ترتیب میانگین و مقادیر انحراف معیار $P_x$ و $P_y$ هستند.

$$f_3 = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} P_d^\theta(i,j)\frac{(i-\mu_x)(j-\mu_y)}{\sigma_x\sigma_y} \quad (۵-۲)$$

$$\mu_x = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} i.P_d^\theta(i,j) \quad (۶-۲)$$

$$\mu_y = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} j.P_d^\theta(i,j) \quad (۷-۲)$$

$$\sigma_x = \sqrt{\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} (i-\mu)^2.P_d^\theta(i,j)} \quad (۸-۲)$$

$$\sigma_y = \sqrt{\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} (j-\mu)^2.P_d^\theta(i,j)} \quad (۹-۲)$$

اگر GLCM متقارن باشد، $\mu_x = \mu_y$ و $\sigma_x = \sigma_y$ هستند.

### ۴) مجموع مربعات[1]

این ویژگی به عنوان واریانس[2] نیز معرفی می‌شود، واریانس میزان پراکندگی پیکسل‌های مرجع و همسایه در اطراف میانگین می‌باشد.

$$f_4 = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} (i-\mu)^2.P_d^\theta(i,j) \quad (۱۰-۲)$$

### ۵) معکوس تفاوت لحظه‌ای[3]

این ویژگی که عموماً همگنی[4] نامیده می‌شود، همگنی محلی یک تصویر را اندازه می‌گیرد. ویژگی IDM میزان نزدیکی توزیع عناصر GLCM به GLCM قطری را به دست می‌آورد.

$$f_5 = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} \frac{1}{1+(i-j)^2}.P_d^\theta(i,j) \quad (۱۱-۲)$$

[1] Sum of Squares
[2] Variance
[3] Inverse Difference Moment (IDM)
[4] Homogeneity

با توجه به رابطه (۲-۱۱)، وزن IDM عکس وزن تضاد است، بنابراین با کاهش وزن به صورت نمایی از قطر دور می‌شود.

**۶) مجموع میانگین**[1]

$$f_6 = \sum_{i=0}^{2(N_g-1)} i.P_{x+y}(i) \qquad (۱۲-۲)$$

$$P_{x+y}(k) = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} P_d^{\theta}(i,j)\,, \qquad k = i+j = \{0,1,2,\dots,2(N_g-1)\} \qquad (۱۳-۲)$$

**۷) مجموع واریانس**[2]

$$f_7 = \sum_{i=0}^{2(N_g-1)} (i-f_6)^2.P_{x+y}(i) \qquad (۱۴-۲)$$

**۸) مجموع آنتروپی**[3]

$$f_8 = -\sum_{i=0}^{2(N_g-1)} P_{x+y}(i).\log[P_{x+y}(i)] \qquad (۱۵-۲)$$

**۹) آنتروپی**[4]

آنتروپی اشاره به مقدار انرژی دارد که به صورت حرارت در زمان یک واکنش یا یک تغییر فیزیکی از بین می‌رود. آنتروپی غیرقابل بازیافت است و با استفاده از رابطه (۲-۱۶) محاسبه می‌گردد.

$$f_9 = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} P_d^{\theta}(i,j).\log[P_d^{\theta}(i,j)] \qquad (۱۶-۲)$$

**۱۰) اختلاف واریانس**[5]

$$f_{10} = \sum_{i=0}^{N_g-1} (i-f'_{10})^2.P_{x-y}(i) \qquad (۱۷-۲)$$

$$P_{x-y}(k) = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} P_d^{\theta}(i,j)\,, \qquad k = |i-j| = \{0,1,2,\dots,N_g-1\} \qquad (۱۸-۲)$$

[1] Sum Average
[2] Sum Variance
[3] Sum Entropy
[4] Entropy
[5] Difference Variance

$$f'_{10} = \sum_{i=0}^{N_g-1} i.P_{x-y}(i) \quad (۱۹-۲)$$

**۱۱) اختلاف آنتروپی**[1]

$$f_{11} = -\sum_{i=0}^{N_g-1} P_{x-y}(i).\log\left[P_{x-y}(i)\right] \quad (۲۰-۲)$$

**۱۲) اطلاعات اندازه همبستگی(۱)**[2]

$$f_{12} = \frac{HXY - HXY1}{\max(Hx, Hy)} \quad (۲۱-۲)$$

**۱۳)اطلاعات اندازه همبستگی(۲)**

$$f_{12} = (1 - exp[-2(HXY2 - HXY)])^{1/2} \quad (۲۲-۲)$$

$$P_x(i) = \sum_{j=0}^{N_g-1} P_d^{\theta}(i,j) \quad (۲۳-۲)$$

$$P_y(j) = \sum_{i=0}^{N_g-1} P_d^{\theta}(i,j) \quad (۲۴-۲)$$

$$HX = -\sum_{i=0}^{N_g-1} P_x(i).\log[P_x(i)] \quad (۲۵-۲)$$

$$HY = -\sum_{i=0}^{N_g-1} P_y(i).\log\left[P_y(i)\right] \quad (۲۶-۲)$$

$$HXY = -\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} P_d^{\theta}(i,j).\log\left[P_d^{\theta}(i,j)\right] \quad (۲۷-۲)$$

$$HXY1 = -\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} P_d^{\theta}(i,j).\log\left[P_x(j).P_y(j)\right] \quad (۲۸-۲)$$

$$HXY2 = -\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} P_x(j).P_y(j).\log\left[P_x(j).P_y(j)\right] \quad (۲۹-۲)$$

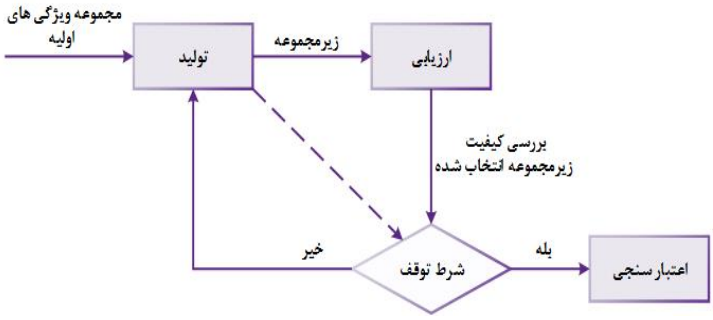
در تمامی روابط (۲-۳) تا (۲-۲۸)، $P_d^{\theta}$ مقادیر نرمال شده دارد.

[1] Difference Entropy
[2] Information Measures of Correlation

## ۲-۵ روش‌های مبتنی بر انتخاب ویژگی

در روش‌های یادگیری ماشین کار با داده‌ها با ابعاد بالا دشوار است برای حل مسأله یادگیری، الگوریتم‌های کاهش ویژگی بسیار مناسب هستند. طبق بررسی محققان، تعداد ویژگی‌های زیاد حاوی اطلاعات مفید نیستند چرا که ممکن است نامرتبط با کلاس داده‌ها باشند. به خصوص اگر تعداد ویژگی‌ها زیاد و تعداد نمونه‌ها کم باشند در نتیجه فضای جستجو کم جمعیت خواهد بود و مدل قادر نیست بین داده‌های مرتبط و نویز تمایز قائل شود. برای حل مشکلات کار با ابعاد بزرگ داده از سال ۱۹۷۰ مفهوم انتخاب ویژگی در زمینه شناسایی آماری الگو [۲۳،۲۲] ، یادگیری ماشین [۲۵،۲۴]، داده کاوی [۲۶] طبقه‌بندی متن [۲۷]، بازیابی تصویر [۲۸] ، مدیریت ارتباط مشتری [۲۹] ،تشخیص نفوذ [۳۰] ، تجزیه و تحلیل ژنوم[1] و ... گسترش یافته است [۳۱].

عمدتاً روش‌های مختلف انتخاب ویژگی در بین همه زیرمجموعه‌های ویژگی‌ها جستجو کرده و بهترین زیرمجموعه را از میان $2^N$ زیر مجموعه کاندید بر اساس برخی معیارهای ارزیابی انتخاب می‌کند. اما با توجه به وسعت جواب‌های ممکن، و اینکه این مجموعه‌های جواب به صورت توانی از $N$ افزایش پیدا می‌کنند، پیدا کردن جواب بهینه مشکل و در $N$ های متوسط و بزرگ بسیار پر هزینه است. روش‌های دیگر بر اساس روش‌های جستجوی اکتشافی یا تصادفی، با در نظر گرفتن یک تعادل در بهینه‌سازی سعی در کاهش هزینه‌های محاسباتی دارند. چهار مرحله در فرآیند انتخاب ویژگی با توجه به شکل ۲-۷، وجود دارد.




شکل ۲-۷: چهار مرحله فرآیند انتخاب ویژگی [۳۲]

[1] Genomic Analysis

۱) تولید زیرمجموعه[1]: تولید زیرمجموعه کاندید برای ارزیابی

۲) ارزیابی زیرمجموعه[2]: ارزیابی زیرمجموعه مورد نظر بر اساس روش داده شده

۳) شرط توقف[3]: تصمیم‌گیری در مورد زمان توقف الگوریتم

۴) اعتبار سنجی[4]: تصمیم گیری در مورد معتبر بودن یا نبودن زیرمجموعه [۳۲]

الگوریتم‌های انتخاب ویژگی با معیارهای ارزیابی مختلفی طراحی شده‌اند و به طور کلی به سه دسته تقسیم می‌شوند:

۱) مدل فیلتر[5][۳۳]

۲) مدل Wrapper [۲۵]

۳) مدل ترکیبی[6] [۳۴]

## ۲-۵-۱ روش‌های عمومی انتخاب ویژگی

در این قسمت چهار بخش فرآیند انتخاب ویژگی تشریح می‌شود.

### ۲-۵-۱-۱ تولید زیر مجموعه

تابع تولیدکننده در واقع یک تابع جستجو است، که زیرمجموعه‌های مختلف که به وسیله تابع ارزیابی، بررسی می‌شوند را به ترتیب تولید می‌کند. ماهیت این فرآیند توسط دو مسأله اساسی تعیین می‌گردد اول تصمیم‌گیری در مورد نقطه یا نقاط شروع و دوم تصمیم‌گیری در مورد روش جستجو است. تابع تولیدکننده از یکی از حالت‌های زیر شروع به کار می‌کند:

۱) بدون ویژگی

۲) با مجموعه تمام ویژگی‌ها

[1] Subset Generation
[2] Subset Evaluation
[3] Stopping Criterion
[4] Result Validation
[5] Filter Model
[6] Hybrid Model

۳) با یک زیرمجموعه تصادفی از ویژگی‌ها

در حالت اول ویژگی‌ها به ترتیب به مجموعه اضافه شده و زیرمجموعه‌های جدید را تولید می‌کنند. این عمل آنقدر تکرار می‌شود تا به زیر مجموعه مورد نظر برسیم. به این گونه روش‌ها، روش‌های پیشرو[1] می‌گویند. در حالت دوم از یک مجموعه شامل تمام ویژگی‌ها، شروع می‌کنیم و به مرور در طی اجرای الگوریتم، ویژگی‌ها را حذف می‌کنیم تا به زیرمجموعه دلخواه برسیم. روش‌هایی که به این صورت عمل می‌کنند، روش‌های عقبگرد[2] نام دارند. جستجو ممکن است با انتخاب تصادفی زیرمجموعه آغاز شود این کار از به دام افتادن در کمینه محلی جلوگیری می‌کند. بر اساس نحوه جستجو در میان این تعداد زیرمجموعه، روش‌های مختلف جستجو در انتخاب ویژگی به سه دسته: جستجوی کامل، جستجوی ترتیبی، جستجوی تصادفی تقسیم می‌شوند [۳۱].

### ۲-۵-۱-۲ ارزیابی زیرمجموعه

معیارهای ارزیابی را می‌توان به صورت گسترده بر اساس وابستگی به الگوریتم‌های استخراج که در نهایت برای انتخاب زیر مجموعه ویژگی‌ها استفاده می‌شوند به دو گروه طبقه‌بندی کرد [۳۱]:

**۱) معیارهای مستقل[3]**

یک معیار مستقل در الگوریتم‌های مدل فیلتر استفاده می‌شود. این نوع معیار سعی در ارزیابی کیفیت یک ویژگی یا زیرمجموعه ویژگی‌ها بر اساس بهره‌برداری از ویژگی‌های ذاتی داده‌ها بدون دخالت هر نوع الگوریتم استخراج را دارد.

**۲) معیارهای وابسته**

روش‌هایی که این نوع از تابع ارزیابی را استفاده می‌کنند، با عنوان Wrapper شناخته می‌شوند که نیاز به الگوریتم استخراج از پیش تعیین شده در انتخاب ویژگی را دارند تا از کارایی آن در تعیین

[1] Forward
[2] Backward
[3] Independent Criterion

ویژگی مناسب برای انتخاب استفاده کنند. دقت عملکرد در این روش‌ها بسیار بالا است، اما هزینه‌های محاسباتی نیز نسبتاً زیاد است [۳۱].

### ۲-۵-۱-۳ معیارهای توقف

بدون داشتن یک شرط خاتمه مناسب، فرآیند انتخاب ویژگی ممکن است برای همیشه درون فضای جستجو برای یافتن جواب سرگردان بماند. برخی معیارهای توقف شامل:

۱) تکمیل شدن جستجو
۲) دستیابی به برخی شرایط مثلاً حداقل تعداد ویژگی یا بیشترین تعداد تکرار
۳) عدم تغییر در نتیجه در اثر حذف یا اضافه کردن ویژگی
۴) انتخاب یک زیرمجموعه‌ای که به اندازه کافی خوب باشد. زیرمجموعه‌ای به اندازه کافی خوب است که خطای کلاسه‌بندی آن کمترین میزان خطای مجاز برای آن روش باشد [۳۱].

### ۲-۵-۱-۴ اعتبارسنجی نتیجه

آسان‌ترین راه برای اعتبارسنجی نتیجه، استفاده از دانش در مورد ویژگی‌های غیرمرتبط و زائد است. اگر از قبل در مورد ویژگی‌های مرتبط اطلاعات وجود داشته باشد، زمانی که از مجموعه داده‌های ساختگی[1] استفاده شود می‌توان از مجموعه دانش قبلی برای مقایسه با ویژگی‌های انتخاب شده استفاده کرد. اما در کاربردهای حقیقی[2] که این دانش قبلی وجود ندارد باید از روش‌های غیر مستقیم استفاده شود، این روش‌ها با نظارت بر تغییر عملکرد الگوریتم در اثر تغییر در تعداد ویژگی‌ها عمل می‌کنند. به عنوان مثال اگر از خطای کلاسه‌بندی برای بررسی کارایی یک روش استفاده کنیم به راحتی می‌توان این خطای کلاسه‌بندی را قبل و بعد از عملیات انتخاب ویژگی محاسبه کرد و با توجه به نتیجه به دست آمده میزان اعتبار روش به کار برده شده را تخمین زد [۳۱].

[1] Synthetic Data
[2] Real-World Applications

## ۲-۵-۲ مدل‌های الگوریتم‌های انتخاب ویژگی با توجه به معیارهای ارزیابی

در ادامه خلاصه‌ای از مدل‌های الگوریتم‌های انتخاب ویژگی با توجه به معیارهای ارزیابی بررسی شده‌اند.

### ۲-۵-۲-۱ مدل فیلتر

در این مدل برای مجموعه داده مورد نظر، الگوریتم از مجموعه آغازی (یک مجموعه کامل یا هر مجموعه تصادفی انتخاب شده با استفاده از یک روش جستجوی خاص) شروع به جستجو می‌کند. هر زیر مجموعه تولید شده با یک معیار مستقل ارزیابی می‌گردد و با بهترین زیرمجموعه‌های قبلی‌اش مقایسه می‌شود اگر این مجموعه جدید بهتر باشد جایگزین بهترین زیرمجموعه قبلی می‌شود. الگوریتم تا رسیدن به شرط توقف ادامه می‌یابد، خروجی الگوریتم بهترین زیرمجموعه جاری به عنوان نتیجه نهایی است. با تغییر روش‌های جستجو و معیارهای ارزیابی که در الگوریتم استفاده می‌شوند می‌توان الگوریتم منفرد متفاوتی را در مدل فیلتر طراحی کرد. مدل فیلتر معیارهای ارزیابی مستقل را بدون به کارگیری هر الگوریتم استخراجی اعمال می‌کند بنابراین از لحاظ محاسباتی بسیار کارآمد است.

### ۲-۵-۲-۲ مدل Wrapper

این الگوریتم بسیار شبیه الگوریتم فیلتر است با این تفاوت که از یک الگوریتم استخراج از پیش تعریف شده به جای معیار اندازه‌گیری برای ارزیابی زیرمجموعه بهره می‌برد. برای هر زیرمجموعه تولید شده، این روش تأثیرش را با اعمال الگوریتم استخراج به داده‌ها، بررسی می‌کند و کیفیت نتیجه استخراج را ارزیابی می‌کند. بنابراین الگوریتم‌های استخراج متفاوت نتایج مختلفی برای انتخاب ویژگی ارائه می‌دهند. تنوع در روش‌های جستجو از طریق تابع تولید و الگوریتم‌های استخراج منجر به الگوریتم‌های متفاوت Wrapper می‌شود. از آنجا که الگوریتم‌های استخراج برای کنترل انتخاب زیرمجموعه ویژگی‌ها به کار می‌روند مدل Wrapper کارایی بیشتری در انتخاب زیرمجموعه‌ای از ویژگی‌ها دارد ولی پیچیدگی محاسباتی آن نسبت به مدل فیلتر بیشتر است [۳۵].

### ۲-۵-۲-۳ مدل ترکیبی

برای دستیابی به مزایای هر دو روش فیلتر و Wrapper و جلوگیری از تعریف یک معیار توقف از پیش تعریف شده، یک مدل ترکیبی برای اداره کردن مجموعه داده‌های بزرگ معرفی شده است. این روش از هر دو شرط، معیار اندازه‌گیری مستقل برای تصمیم‌گیری جهت دستیابی به بهترین زیرمجموعه و از الگوریتم استخراج برای ارزیابی زیرمجموعه ویژگی‌ها و انتخاب نهایی بهترین زیرمجموعه از میان همه بهترین زیرمجموعه‌ها استفاده می‌کند. عموماً این روش از یک زیرمجموعه آغازین خالی شروع می‌کند و تا یافتن بهترین زیرمجموعه ادامه می‌دهد. می‌توان کیفیت نتیجه الگوریتم استخراج را یک معیار توقف طبیعی برای مدل ترکیبی در نظر گرفت.

## ۲-۵-۳ کاربردهای انتخاب ویژگی

داده‌ها به دلایل بسیاری غیر از داده کاوی از جمله بایگانی، نیاز قانون و ... جمع‌آوری می‌شوند. زمانی‌که با تعداد متغیر و زیاد صفات روبه‌رو هستیم مثلاً در مواردی نظیر مرور وب تجزیه و تحلیل میکرو آرایه‌ها[1] تجزیه و تحلیل اسناد متنی و... می‌توان قسمت‌های مختلف یک مسأله را با کاهش ابعاد آن به ابعاد کمتر درک کرد. در ادامه برخی کاربردهای انتخاب ویژگی را تشریح می‌کنیم.

### ۲-۵-۳-۱ طبقه‌بندی اسناد

مسأله طبقه‌بندی اسناد [۳۶] در اصل اختصاص یک کلاس از پیش تعریف شده به اسناد متنی است. ضرورت این مسأله در مواردی که با حجم وسیع متن‌های آنلاین در شبکه‌های جهانی، ایمیل‌ها و کتابخانه‌های دیجیتال روبه‌رو هستیم، قابل لمس‌تر است. در این مسایل یک فضای ویژگی معمولی شامل تعداد زیادی کلمه یا عبارت است که در اسناد رخ می‌دهد، این تعداد برای یک سند در اندازه متوسط به صدها هزار عبارت می‌رسد. بنابراین مسأله کاهش ویژگی بدون از بین رفتن دقت دسته-بندی در اینجا اهمیت بسیار می‌یابد.

[1] Microarray Analysis

### ۲-۵-۳-۲ بازیابی تصویر

در سال‌های اخیر افزایش سریع اندازه کیفیت تصاویر در هر دو حوزه نظامی و غیرنظامی وجود داشته است. بازیابی تصویر مبتنی بر محتوا برای کنترل مؤثر تصاویر در مقیاس بزرگ معرفی شده است. در این روش برخلاف اختصاص یک کلاس از پیش تعریف شده به متن، تصاویر با استفاده از ویژگی‌های خود تصویر مثلاً رنگ، شکل، بافت و ... شاخص‌گذاری می‌شوند. بزرگترین مشکل در بازیابی تصاویر بر اساس محتوا بزرگ بودن ابعاد است [۲۸].

### ۲-۵-۳-۳ مدیریت ارتباط با مشتری

در این روش مشتری به عنوان یک سود بزرگ است و از دست دادن یکی از آن‌ها عاملی برای از دست دادن یک بخش می‌شود، بنابراین نیازمند یک تیم با تجربه هستیم که به مقاصد مشتریان نظارت داشته باشند. یک مجموعه شاخص توسط تیم استفاده می‌شوند و اثبات می‌شود که این شاخص‌ها برای پیش‌بینی مشتری که از دست خواهد رفت مناسب‌اند. مشکل یافتن شاخص جدید برای توصیف محیط پویا و متغیر کسب وکار در میان بسیاری از شاخص‌های دیگر (ویژگی‌ها) است [۳۷].

### ۲-۵-۳-۴ تشخیص نفوذ

شبکه‌های کامپیوتری در پیشرفت جوامع نقش حیاتی دارند، لذا به هدفی برای نفوذ دشمنان تبدیل شده‌اند. امنیت سیستم کامپیوتری زمانی که نفوذ اتفاق می‌افتد دچار اختلال می‌شود بنابراین تشخیص نفوذ یک راه حفظ امنیت سیستم‌های کامپیوتری است. در [۳۰] لی[1]، استولفو[2] و ماک[3] پیشنهاد یک چارچوب داده کاوی سیستماتیک را برای تجزیه و تحلیل داده‌ها و ساخت مدل‌های تشخیص نفوذ دادند. در این چارچوب مقدار زیادی از داده‌ها برای اولین بار توسط الگوریتم‌های داده-

[1] Lee
[2] Stolfo
[3] Mok

کاوی به منظور دستیابی به الگوهای تکراری[1] استفاده می‌شوند، سپس این الگوها به عنوان راهنما برای انتخاب ویژگی استفاده می‌شوند.

### ۲-۵-۳-۵ تجزیه و تحلیل ژنومی

اطلاعات ساختاری و کارکردی از تجزیه و تحلیل ژنوم انسان در سال‌های اخیر در زمینه‌های زیادی گسترش یافته است. یکی از مشکلات این کار محدود بودن تعداد نمونه‌ها است. در [۳۸] مشاهده می‌کنیم مجموعه نمونه‌ها شامل ۳۸ داده برای آموزش است که این داده‌ها ۷۱۳۰ بعد دارند. برای حل مشکل افراطی کم بودن مشاهدات و زیاد بودن تعداد ویژگی‌ها زینگ[2]، جوردن[3] و کارپ[4] راه حل کاهش ویژگی را ارائه دادند.

[1] Frequent Pattern
[2] Xing
[3] Jordan
[4] Karp

# فصل۳ الگوریتم سیستم ایمنی مصنوعی

## ۳-۱ سیستم ایمنی[1]

انسان در محیطی زندگی می‌کند که عوامل و میکروارگانیسم‌های بیماری‌زای متعددی سلامتی وی را به طور دائم تهدید می‌کند بدن برای مقابله با عوامل بیماری‌زا مجهز به سیستم دفاعی و ایمنی است. در پزشکی عبارت ایمنی اشاره به شرایطی دارد که یک عضو در برابر بیماری‌ها بخصوص بیماری‌های عفونی مقاومت می‌کند. در یک تعریف گسترده‌تر ایمنی شامل واکنش بدن به مواد خارجی و عوامل بیماری‌زا است که شامل پاسخ‌های ایمنی اولیه[2] و ثانویه[3] می‌باشد. عملکرد فیزیولوژیک سیستم ایمنی به دفاع از یک عضو در برابر انواع مواد مضر مانند قارچ‌ها، باکتری‌ها، انگل-ها، ویروس‌ها می‌پردازد. حتی مواد خارجی غیرآلوده نیز می‌توانند پاسخ ایمنی تولید کنند. یکی از توانایی‌های مهم سیستم ایمنی تمایز قائل شدن بین سلول‌های بدن از سلول‌های خارجی است که به آن تمایز خودی/ غیرخودی[4] می‌گویند. به طور کلی سیستم ایمنی قادر به شناسایی عوامل خطرناک و تصمیم‌گیری برای پاسخ مناسب در برابر آن‌ها و همچنین تحمل در برابر مولکول‌های خودی و نادیده گرفتن بسیاری از مواد بی‌ضرر است [۳۹].

### ۳-۱-۱ آنتی‌ژن[5]

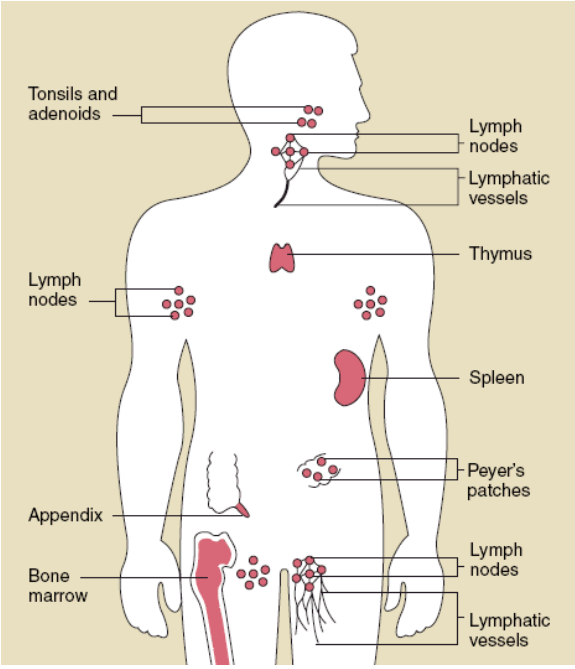
آنتی‌ژن‌ها موادی هستند که در نتیجه ورود به بدن قادرند سیستم ایمنی را علیه خودشان به طور اختصاصی تحریک کنند. اصطلاح آنتی‌ژنیسیتی به معنی توانایی بالقوه یک آنتی‌ژن در ایجاد عکس العمل می‌باشد. کلمه آنتی‌ژن ریشه یونانی دارد و از پیشوند آنتی به معنای در برابر و پسوند ژن به معنی مولد درست شده است.

[1] Biological Immune System (BIS)
[2] Primary Response
[3] Secondary Response
[4] Self / Non Self
[5] Antigen (AG)

## ۳-۱-۲ عناصر سیستم ایمنی

### ۳-۱-۲-۱ اعضا

اندام‌هایی که سیستم ایمنی را تشکیل می‌دهند در شکل ۳-۱ نمایش داده شده‌اند. این اندام‌ها همان‌طور که در شکل ۳-۲ مشاهده می‌شود به دو دسته تقسیم می‌شوند که در ادامه آن‌ها را شرح خواهیم داد.



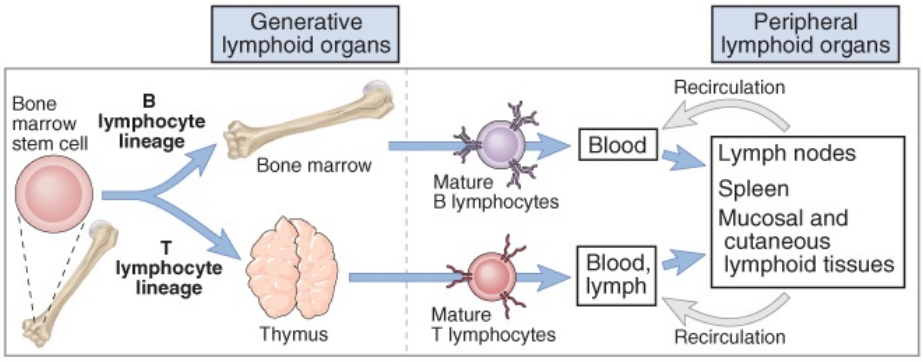

شکل ۳-۱: اندام‌های لنفاوی [۴۰]

#### ۱) اندام‌های لنفاوی مرکزی[1]

وظیفه این اندام‌ها تولید و کمک به سلول‌های ایمنی بالغ همان لنفوسیت‌ها[2] است که شامل مغز استخوان و تیموس می‌باشند.

- مغز استخوان: بافت نرم درون استخوان‌ها و منشأ تولید سلول‌های خونی می‌باشد [۳۹].
- تیموس: اندامی که پشت استخوان قفسه سینه قرار گرفته است و وظیفه آن تولید سلول‌های $T$ بالغ می‌باشد، این سلول‌ها قادر به تولید پاسخ ایمنی هستند [۴۱].

[1] Central Lymphoid or Generative Lymphoid
[2] Lymphocyte




شکل ۳-۲: اندام‌های لنفاوی مرکزی و ثانویه [۴۲]

### ۲) اندام‌های لنفاوی ثانویه[1]

این اندام‌ها جهت تسهیل تعامل بین لنفوست‌ها و آنتی‌ژن‌ها به کار می‌روند و شامل:

- لوزه‌ها و آدنوئید: در قسمت حلق و عقب ناحیه دهانی قرار گرفته‌اند.
- گره‌های لنفاوی: اندام بیضی شکل از سیستم ایمنی هستند که به صورت گسترده‌ای در سراسر بدن پخش شده‌اند. غدد لنفاوی شامل سلول‌های $T$، سلول‌های $B$ و سایر سلول‌های ایمنی می‌باشند و به عنوان فیلتر یا تله ذرات بیگانه مانند برخی سلول‌های سرطانی عمل می‌کنند [۴۳].
- طحال: در قسمت بالا و چپ شکم قرار دارد و فعالیت‌های بسیاری از اجرای سیستم ایمنی را هدایت می‌کند. طحال یک تصفیه کننده مؤثر خون است که سلول‌های سفید و قرمز فرسوده را حذف کرده و به آنتی‌ژن‌های موجود در خون به صورت فعال پاسخ می‌دهد [۴۳].
- آپاندیس: زائده‌ای کوچک است که از روده بزرگ و در ناحیه پایین و راست شکم منشعب می‌شود بافت‌های لنفاوی جداره آپاندیس با شناسایی میکروب‌های موجود در مواد زائد در حال دفع از بدن به سیستم ایمنی بدن اجازه می‌دهند تا در مقابله و دفع میکروب‌های مضر برنامه‌ریزی کند [۴۴].

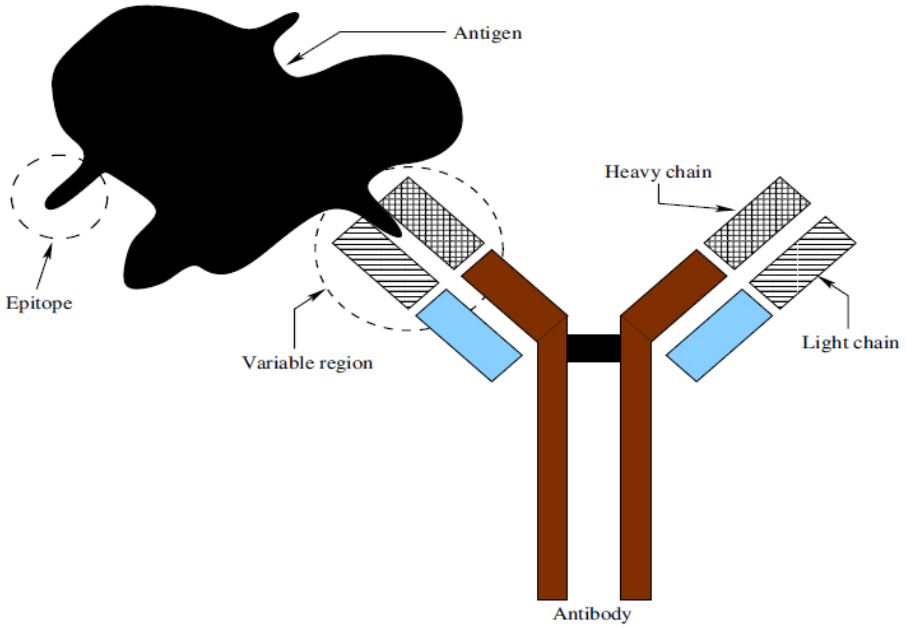
[1] Peripheral Lymphoid

### ۳-۱-۲-۲ سلول‌های ایمنی و مولکول‌ها

سلول‌های سفید که لنفوسیت نامیده می‌شوند در مغز استخوان تولید شده و در حال گردش درخون هستند. جمعیت اصلی لنفوسیت‌ها را سلول‌های $B$ و سلول‌های $T$ تشکیل می‌دهند. در سطح سلول‌های $B$ و $T$ مولکول‌های گیرنده‌ای وجود دارند که به سایر سلول‌ها می‌چسبند.

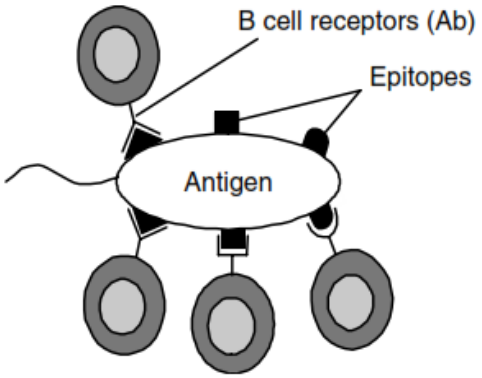
- سلول $B$: این نوع سلول مغز استخوان را به عنوان یک لنفوسیت بالغ ترک می‌کند و عمدتاً در طحال و لوزه‌ها حضور دارد. در این اندام‌ها سلول‌های $B$ به سلول‌های پلاسما توسعه یافته و برای تماس با آنتی‌ژن‌ها آماده می‌شوند. میلیاردها از این نوع سلول در حال گردش در خون هستند که موجب تشخیص ناهنجاری‌ها و تولید پاسخ ایمنی خواهند شد. سلول‌های $B$ وقتی علامت خطر را دریافت می‌کنند تغییرات مهمی در آن‌ها صورت می‌گیرد و مواد شیمیایی پرقدرتی را در خون تولید می‌کنند این مواد شیمیایی به نام آنتی بادی‌ها[1] یا پادتن‌ها نیز نامیده می‌شوند [۴۰].
- آنتی‌بادی: پروتئین‌های شیمیایی شبیه $Y$، مانند شکل ۳-۳ هستند و از دو بخش تشکیل شده‌اند: یک بخش به سلول $B$ متصل است و به آن ناحیه ثابت گفته می‌شود و بخش دیگر آن سر بازوها یا مناطق متغیر (متفاوت در آنتی‌بادی‌های مختلف) می‌باشد. مناطق متغیر پاراتوپ‌ها[2] هستند که آنتی‌بادی‌ها را برای مطابقت با آنتی‌ژن و اتصال به اپی‌توپ[3] آنتی‌ژن فعال می‌کنند [۴۰]. هنگامی که شکل بخش متغیر مکمل یک آنتی‌ژن باشد می‌گوییم آنتی-بادی با آنتی‌ژن میل ترکیبی دارد. برخلاف آنتی‌ژن‌ها، آنتی‌بادی‌ها زمانی که لنفوسیت‌ها با آنتی‌ژن‌ها در تماس قرار می‌گیرند تولید می‌شوند.

[1] Antibody (AB)
[2] Paratopes
[3] Epitopes




شکل ۳-۳: نمایش اتصال آنتی‌بادی و آنتی‌ژن [۴۰].

هرلنفوسیت از زیر گروه لنفوسیت $B$ برای ساختن تنها و تنها یک آنتی‌بادی با ویژگی خاص برنامه ریزی شده است، به نحوی که بعد از تهیه، آن‌ها را در سطح خود قرار می‌دهد تا به عنوان پذیرنده برای آنتی‌ژن مربوطه وارد عمل شود. در شکل ۳-۴، برقراری اتصال سلول $B$ با آنتی‌بادی‌های متفاوت و آنتی‌ژن با اپی‌توپ مکمل آن نمایش داده شده است [۴۳].

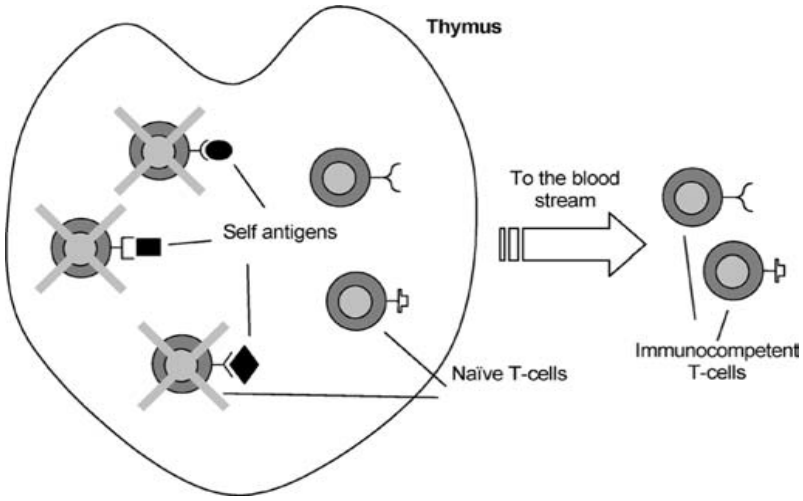



شکل ۳-۴: سلول $B$ با گیرنده های متفاوت (آنتی‌بادی) در حال برقراری اتصال با آنتی‌ژنی که اپی‌توپ مکمل پاراتوپ آنتی‌بادی را دارد [۳۹]

- سلول$T$: این نوع سلول به صورت نابالغ مغز استخوان را ترک می‌کند و از طریق جریان خون وارد تیموس می‌شود. در تیموس بخش محافظت شده‌ای وجود دارد که در آن فقط سلول‌های

خودی حضور دارند و هیچ عامل بیماری‌زایی نمی‌تواند به آن بخش وارد شود سپس سلول‌ها در تیموس بالغ می‌شوند. همان‌طور که در شکل ۳-۵ مشاهده می‌شود، سلول‌های بالغ شده در تیموس تنها زمانی اجازه ورود به جریان خون را دارند که گیرنده‌شان توانایی شناسایی سلول-های خودی را نداشته باشد. بنابراین تمایز قائل شدن بین سلول‌های خودی و غیرخودی برای سلول‌های $T$ بسیار بااهمیت است.

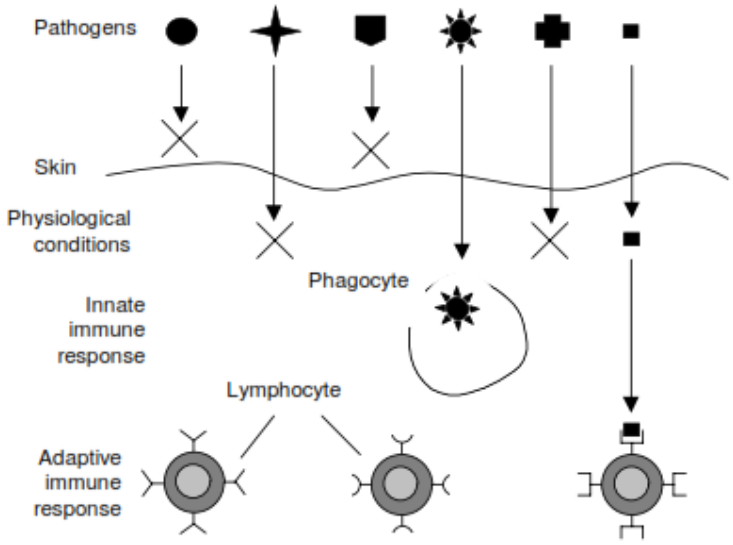
همان‌طور که قبلاً گفته شد آنتی‌بادی به سلول‌های $B$ متصل است اما سلول‌های $B$ نمی‌توانند آنتی‌ژن‌های خودی را تشخیص دهند، بنابراین برای تشخیص یک آنتی‌ژن غیرخودی به کمک یک سلول $T$ نیاز دارند. به عبارت دیگر پاسخ ایمنی سلول‌های $B$ هنگامی تولید می‌شود که یک سلول $B$ و یک سلول $T$ کمکی به طور هم زمان یک آنتی ژن را شناسایی کنند [۴۰].




شکل ۳-۵: بلوغ سلول های T در تیموس و ورود آن ها به جریان خون [۴۱]

### ۳-۱-۳ لایه‌های سیستم ایمنی

سیستم ایمنی یک سیستم چند لایه است که هر لایه شامل انواع متفاوتی از مکانیزم‌های دفاعی است. در شکل ۳-۶، این لایه‌ها نمایش داده شده‌اند.




شکل ۳-۶: لایه‌های مختلف سیستم ایمنی [۴۱]

### ۳-۱-۳-۱ سدهای خارجی

پوست از خطوط دفاعی اصلی به شمار می‌آید زیرا پوست سالم نسبت به اکثر عوامل عفونی نفوذناپذیر است. ترشحات مخاطی تولید شده توسط غشاهای پوشاننده سطوح داخلی نیز مانند سدی در برابر نفوذ میکروب‌ها عمل می‌کنند، عوامل میکروبی که درون مخاط به دام می‌افتند توسط عوامل مکانیکی نظیر مژک‌ها، شستشوی اشک، بزاق و ... دفع می‌شوند. بسیاری از مایعات ترشح شده از بدن مانند اسید معده، لیزوزم در اشک، ترشحات بینی و بزاق نیز خاصیت ضد میکروبی دارند [۴۳].

### ۳-۱-۳-۲ ایمنی ذاتی[1]

ایمنی ذاتی تمام مکانیزم‌های دفاعی که از ابتدای تولد در بدن هر فرد وجود دارد را شامل می‌شود. پاسخی که توسط ایمنی ذاتی تولید می‌شود اختصاصی و پایدار نیست و سبب مصونیت نمی‌شود [۳۹]. در این مرحله تفاوتی ندارد عامل خارجی که بدن با آن مواجه شده است چه نوعی باشد چراکه ایمنی ذاتی به همه این عوامل پاسخ یکسان می‌دهد.

[1] Innate Immunity

### ۳-۱-۳-۳ ایمنی اکتسابی[1]

بروز جهش فرصت زیادی را در اختیار مهاجمان میکروبی قرار می‌دهد تا بتوانند از سد دفاع ایمنی ذاتی بدن بگریزند. بنابراین بدن نیاز به ساز وکارهای دفاعی دارد تا هرکدام از آن‌ها به طور جداگانه بتوانند متناسب با عوامل مهاجم و بدون توجه به نوع آن‌ها پاسخ ایمنی ایجاد کنند. لنفوسیت‌های بالغ توانایی شناسایی مولکول‌ها و سلول‌های خودی را از بیگانه، و نیز مقابله با عوامل بیگانه را به دست می‌آورند و وارد جریان خون می‌شوند. آن‌ها بر سطح خود دارای گیرنده‌هایی هستند که از لحاظ شکل هندسی مکمل نوع خاصی از آنتی‌ژن که بر سطح عوامل بیگانه قرار دارد، می‌باشند. به این ترتیب هر لنفوسیت، با داشتن نوع خاصی گیرنده، آنتی‌ژن خاصی را شناسایی کرده و از بین می‌برد. به همین علت می‌گوییم که لنفوسیت‌ها به طور اختصاصی عمل می‌کنند [۴۳].

### ۳-۱-۴ شناسایی ایمنی

نه تنها بدن باید تفاوت موجود بین یک آنتی‌ژن را نسبت به آنتی‌ژن دیگر تشخیص دهد، بلکه باید عوامل بیگانه یعنی غیر خود را نیز بشناسد، زیرا عدم شناخت عوامل بیگانه و تفاوت قائل نشدن بین خود و غیرخود می‌تواند منجر به تولید آنتی‌بادی علیه اجزای بدن خود فرد (اتوآنتی‌بادی) و ایجاد بیماری‌های خودایمنی[2] شود [۴۵]. آن دسته از ترکیبات و اجزای در حال گردش در بدن که بتوانند در دوره جنینی خود را به اعضای لنفاوی در حال رشد برسانند، به عنوان اجزای خودی به حساب می‌آیند. لذا نوعی بی‌پاسخی دائمی یا تحمل ایمنی[3] نسبت به آنتی‌ژن‌های موجود در این بافت‌ها ایجاد می‌شود. به این سبب هنگام بلوغ و تکامل سیستم ایمنی، فعالیت لنفوسیت‌های خود واکنش‌گر[4] مهار و دچار تحمل می‌شوند. مفاهیم انتخاب منفی و انتخاب مثبت با توجه به موارد گفته شده در این قسمت مطرح می‌شوند [۴۳].

[1] Adaptive Immunity
[2] Self Tolerance
[3] Tolerance
[4] Self-reacting lymphocyte

### ۳-۱-۴-۱ انتخاب منفی[1]

هدف از انتخاب منفی بررسی برای تحمل در برابر سلول‌های خودی است. سلول‌های $T$ که قادر به شناسایی سلول‌های خودی هستند در این تست ناموفق‌اند و فقط آن سلول‌هایی که قادر به شناسایی سلول‌های غیر خودی‌اند نگه‌داشته می‌شوند.

### ۳-۱-۴-۲ انتخاب مثبت[2]

عکس انتخاب منفی است. در این انتخاب سلول‌های $T$ که قادر به شناسایی سلول‌های خودی نباشند از بین رفته و بقیه نگه‌داشته می شوند.

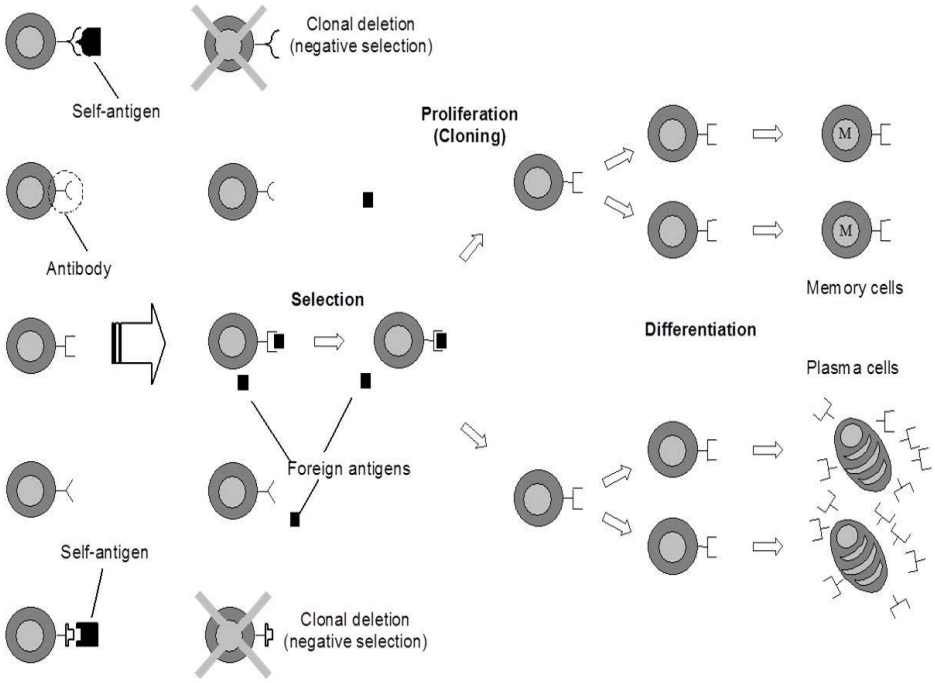
### ۳-۱-۴-۳ انتخاب کلونال[3]

بدن می‌تواند صدها تا هزاران و احتمالاً میلیون‌ها نوع مولکول آنتی‌بادی متفاوت بسازد. اما در عمل امکان داشتن تعداد بسیار زیاد لنفوسیت برای تولید یک به یک آنتی‌بادی‌ها وجود ندارد. از طرفی جای کافی در بدن برای اسکان آن‌ها پیدا نخواهد شد. برای جبران این مسأله، تنها لنفوسیت‌هایی که با آنتی ژن تماس داشته‌اند به طور موفقیت آمیز مراحل تکثیر را طی می‌کنند و کلون‌های بزرگی از پلاسماسل‌ها[4] را به وجود می آورند. پلاسماسل‌های مذکور آنتی‌بادی‌هایی را می‌سازند که لنفوسیت‌های مادر برنامه‌ریزی کرده‌اند. به این سیستم که در طی آن مقادیر زیاد و کافی از آنتی‌بادی برای از پای درآوردن عوامل عفونی تولید می‌شود انتخاب کلون می گویند. در شکل ۳-۷، سلول انتخاب شده توسط آنتی‌ژن در طول تکثیر کلون به مقدار زیاد تقسیم شده و سلول‌های حاصل، بالغ شده و تعداد زیادی سلول تولیدکننده آنتی‌بادی را ایجاد می‌کنند.

بخشی از سلول‌های حاصل از لنفوسیت‌های اولیه واکنش دهنده با آنتی‌ژن به سلول‌های خاطره [5]

[1] Negative Selection
[2] Positive Selection
[3] Clonal Selection
[4] Pelasma Cell
[5] Memory Cell

(حافظه) و مابقی پس از بلوغ به سلول‌های کارگزار[1] تبدیل می‌گردند. سلول‌های حافظه مدت بیشتری عمر می‌کنند و اگر بار دیگر بدن با عامل بیماری‌زای پیشین مواجه شود، تعداد بیشتری آنتی‌بادی و تعداد کمتری سلول حافظه تولید می‌کند [۴۳].




شکل ۳-۷: فرآیند انتخاب کلونال [۴۵]

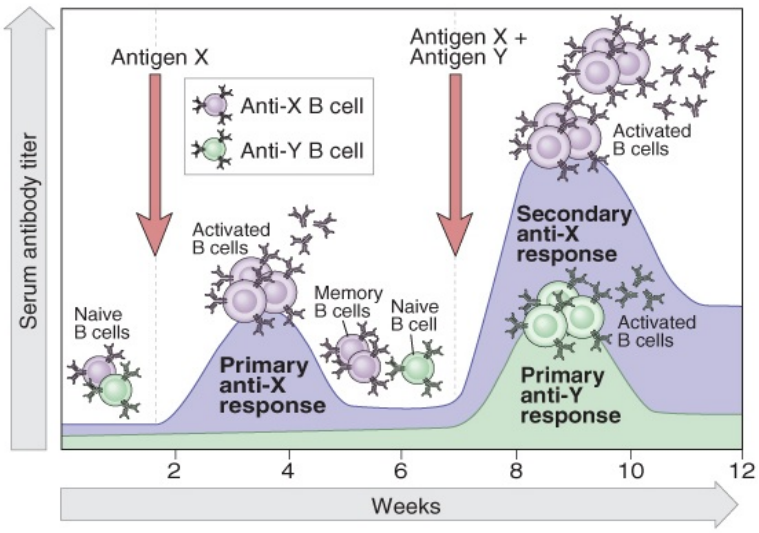
سلول‌های خاطره برای اینکه به سلول‌های کارگزار تبدیل شوند به چرخه‌های کمتری نیاز دارند که خود موجب کوتاه شدن زمان واکنش در پاسخ ثانویه می شود. سلول‌های حاصل از تکثیر کلون که دارای خاطره برای آنتی‌ژن اصلی هستند، علت اصلی قدرت بیشتر پاسخ ایمنی ثانویه نسبت به پاسخ ایمنی اولیه می‌باشند. در شکل ۳-۸، سرعت بیشتر پاسخ ایمنی ثانویه نسبت به پاسخ ایمنی اولیه مشاهده می‌گردد.

### ۳-۱-۴-۴ میل ترکیبی

در انتخاب کلونال آنتی‌بادی‌هایی برای تکثیر و گسترش به عنوان پلاسماسل‌ها انتخاب می‌شوند که میل ترکیبی بیشتری با آنتی‌ژن مورد نظر را داشته باشد. یعنی گیرنده آنتی‌بادی انتخاب شده باید

[1] Effector Cell

بیشترین شباهت را با گیرنده آنتی‌ژن مورد نظر داشته باشد تا پیوند قوی‌تری برقرار گردد.




شکل ۳-۸: مقایسه پاسخ ایمنی اولیه و ثانویه [۴۲]

## ۳-۲ الگوریتم سیستم ایمنی مصنوعی

واژه سیستم‌های ایمنی مصنوعی نخستین بار درسال ۱۹۸۶ توسط فارمر[1] مطرح شده است. شاخه محاسبه ایمنی[2] و یا الگوریتم AIS در بیست سال گذشته در حال گسترش است [۴۵] و یکی از منابع الهام‌بخش برای سیستم‌های مهندسی است که دارای خواصی نظیر مقاوم بودن، سازگاری، یادگیری، حافظه، شناسایی، استخراج ویژگی، تنوع و مقیاس‌پذیری می‌باشد [۴۶, ۴۷].

ریاضیدانان و محققین ایمونولوژی مصنوعی هر کدام تعاریف متفاوت ولی از نظر مفاهیم بنیادین مشابهی ارائه داده‌اند که چند نمونه از آن‌ها عبارتند از:

تعریف اول از تیمیس: سیستم های ایمنی مصنوعی، شاخه ای از علوم کامپیوتر است که از سیستم ایمنی طبیعی نشأت گرفته شده است[۴۷].

تعریف دوم از دسگوپتا[3]: سیستم های ایمنی مصنوعی، متدولوژی هوشمندی است که از سیستم ایمنی دنیای واقعی برای حل مسایل استفاده می‌کند [۴۶].

[1] Farmer
[2] Immunological Computation (IC)
[3] Dasgupta

در نهایت سیستم های ایمنی مصنوعی، سیستم‌های تطبیق پذیری هستند کـه از تئـوری ایمنـی شناسی الهام گرفته شده‌اند و از مشاهدات به عمل آمده از توابع، مدل ها و قاعـده کلـی آن‌هـا در حـل مسایل استفاده می‌شود.

جدول ۳-۱، مدل‌های محاسباتی از BIS که بیشتر بررسی شده‌اند را نشان می‌دهد و اصـطلاحات رایج در الگوریتم‌های ایمنی و معادل آن‌ها در یادگیری ماشین در جدول ۳-۲ شرح داده شده‌اند [۳۹].

جدول ۳-۱: مدل‌های محاسباتی از BIS که بیشتر بررسی شده‌اند

| مفاهیم ایمونولوژیک | اساس مدل ایمنی | کاربرد در مسایل محاسباتی |
|---|---|---|
| شناسایی خودی و غیرخودی توسط سلول‌های T | الگوریتم انتخاب منفی [۴۸] | شناسایی خطا، تغییرات و ناهنجاری |
| شبکه آیدیوتایپ[1] | نظریه شبکه ایمنی [۴۹] | یادگیری با ناظر و بدون ناظر |
| حافظه ایمنی، توسعه کلونال سلول-های B، بلوغ میل ترکیبی | الگوریتم انتخاب کلونال [۲۹] | جستجو و بهینه‌سازی |
| ایمنی ذاتی | DT [50] | استراتژی دفاع |

جدول ۳-۲: اصطلاحات رایج در الگوریتم‌های ایمنی و معادل آن ها در یادگیری ماشین

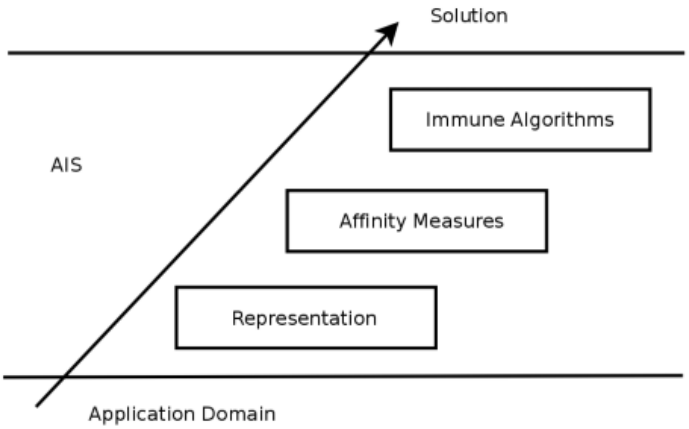
| مدل ایمنی | یادگیری ماشین |
|---|---|
| سلول‌های $B$ و $T$ و آنتی‌بادی | آشکارسازها، خوشه‌ها، کلاسه‌بندها و رشته‌ها |
| سلول‌های خودی، مولکول‌های خودی و سلول‌های ایمنی | نمونه‌های مثبت، داده‌های آموزشی و الگوها |
| آنتی‌ژن، پاتوژن و اپی‌توپ | داده‌های ورودی و داده‌های تست |
| محاسبه میل ترکیبی در فضای شکل- فضا[2]، قواعد مکمل[3] و سایر قوانین | فاصله و اندازه‌گیری شباهت، قواعد تطبیق عبارات |

یک چارچوب کلی برای AIS توسط [۴۵] معرفی شده است که در شکل ۳-۹، مشاهده می‌شـود. این چار چوب دارای سه لایه به شرح زیر می‌باشد.

- نمایش اجزا[4]: چگونگی نمایش اجزای سیستم را مشخص می‌کند.

[1] Idiotypic (آرایش ملکولی اسیدهای آمینه منحصر به مفرد در محل اتصال آنتی‌ژن به یک آنتی‌بادی خاص)
[2] Shape - Space
[3] Complementary Rule
[4] Component Representations

- اندازه میل ترکیبی[1]: مشخص می‌کند تعامل بین اجزای سیستم چگونه اندازه‌گیری می‌شود.
- الگوریتم ایمنی: چگونه اجزای سیستم برای تعیین پویایی سیستم در تعامل با یکدیگر قرار می‌گیرند.



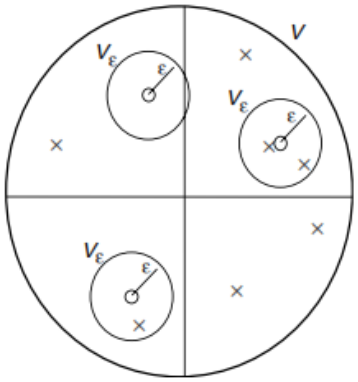

شکل ۳-۹ : چارچوب کلی برای AIS [۴۵]

### ۳-۲-۱ شکل- فضا و میل ترکیبی

در سال ۱۹۷۹ پیرسون[2] و اوستر[3] [۵۱] مفهوم شکل - فضا یا نمایش فضایی را برای نمایش اتصال آنتی‌بادی به آنتی‌ژن معرفی کردند.

آنتی‌بادی و آنتی‌ژن با توجه به خواص فیزیکی اتصالشان توصیف می‌شوند. در مدل‌های محاسباتی عبارت میل ترکیبی بین آنتی‌بادی و آنتی‌ژن بر اساس فاصله بین نقاط در مدل شکل – فضا تعریف می‌شود. فاصله‌های کوچک بین آنتی‌بادی و آنتی‌ژن نشان‌دهنده میل ترکیبی زیاد بین آن دو است. در شکل ۳-۱۰، دایره بزرگ خارجی $V$، نقاط با علامت $x$ و دایره‌های کوچک داخلی به ترتیب نشان‌دهنده شکل- فضا، آنتی‌ژن‌ها و میل ترکیبی آنتی‌بادی‌ها است و $\varepsilon$ آستانه شناسایی می‌باشد. اگر میل ترکیبی بین آنتی‌بادی و آنتی‌ژن از $\varepsilon$ کمتر باشد یعنی آنتی‌ژن داخل ناحیه میل ترکیبی آنتی-بادی قرار گرفته است، لذا گفته می‌شوند آنتی‌بادی با آنتی‌ژن پیوند برقرار کرده است [۳۹].

[1] Affinity Measures
[2] Perelson
[3] Oster




شکل ۳-۱۰: نمایش شکل – فضا ؛ آنتی‌بادی و آنتی‌ژن‌ها به صورت نقاطی در فضای N بعدی نمایش داده شده‌اند [۳۹].

### ۳-۲-۱-۱ روش‌های نمایش

موجودیت‌ها در الگوریتم ایمنی عمدتاً شامل سلول‌های $B$، $T$، آنتی‌بادی‌ها و آنتی‌ژن‌ها هستند. نحوه نمایشی که در بیشتر الگوریتم‌های ایمنی به کار می‌رود به صورت زیر است:

- رشته‌های دودویی
- رشته‌هایی محدود از حروف الفبا
- بردارهایی با مقادیر حقیقی
- نمایش ترکیبی (زمانی‌که یک موجودیت شامل چندین ویژگی است و هر ویژگی متفاوت از سایر ویژگی‌هاست از نمایش ترکیبی استفاده می‌شود.)

بسیاری از مدل‌ها از نمایش دودویی استفاده می‌کنند، این شیوه نمایش فوایدی دارد از جمله: هر نوع داده‌ای قابل نمایش به صورت دودویی می‌باشد، تجزیه و تحلیل آن آسان و مناسب برای نمایش داده‌های قطعی[1] است. محدودیت‌های نمایش فرم دودویی شامل: تفسیر آن در فضا مشکل است، استفاده مستقیم از این شیوه نمایش در روش‌هایی که در فضای پیوسته هستند دشوار است [۵۲].

## ۳-۲-۲ میل ترکیبی AIS

برای تعریف مفهوم میل ترکیبی بین آنتی‌بادی و آنتی‌ژن از انواع مختلف روش‌های محاسبه

[1] Categorical Data

شباهت و یا اندازه‌گیری فاصله استفاده شده است. به عنوان مثال اگر نمایش رشته‌ای باشد فاصله همینگ مناسب است و یا اگر از بردار حقیقی برای نمایش استفاده شود فاصله اقلیدسی استفاده می‌گردد.

### ۳-۲-۲-۱ قوانین تطبیق رشته[1]

انتخاب یک قانون تطبیق که "مطابقت[2]" یا "شناسایی" نامیده می‌شود به روش نمایش و نوع داده وابسته است. انواع روش‌هایی که استفاده می‌شوند در ادامه معرفی خواهند شد [۳۹].

۱) فاصله همینگ

فاصله همینگ تعریف شده بین دو رشته، تعداد کاراکترها متفاوت بین این دو رشته را مشخص می‌کند.

$$h(x,y)=\sum_{i=1}^{N}\left(\overline{X_i \oplus Y_i}\right) \qquad (۱-۳)$$

در رابطه (۳-۱)، طول رشته برابر $N$ ، $X_i$ و $Y_i$ به ترتیب $i$ امین بیت از رشته‌های $X$ و $Y$ ، عملگر "$\oplus$ یا انحصاری[3]" می‌باشند.

۲) فاصله دودویی

در این روش اگر $X_i$ و $Y_i$ به ترتیب iامین بیت از رشته‌های $X$ و $Y$ باشند، مقادیر $a$، $b$، $c$ و $d$ به صورت زیر تعریف می‌شوند:

$a$: تعداد بیت‌های متناظر با مقادیر یک در هر دو رشته $X$ و $Y$.

$b$: تعداد بیت‌ها با مقادیر یک در رشته $X$ که بیت متناظر با آن در رشته $Y$ صفر است.

$c$: تعداد بیت‌ها با مقادیر صفر در رشته $X$ که بیت متناظر با آن در رشته $Y$ یک است.

$d$: تعداد بیت‌های متناظر با مقادیر صفر در هر دو رشته $X$ و $Y$.

[1] String-Matching Rules
[2] Matching
[3] XOR

$$a = \sum_{i=1}^{N} \xi_i \ , \qquad \xi_i = \begin{cases} 1 \ , & if \ \ X_i = Y_i = 1 \\ 0 \ , & otherwise \end{cases} \tag{۲-۳}$$

$$b = \sum_{i=1}^{N} \xi_i \ , \qquad \xi_i = \begin{cases} 1 \ , & if \ \ X_i = 1 \ , Y_i = 0 \\ 0 \ , & otherwise \end{cases} \tag{۳-۳}$$

$$c = \sum_{i=1}^{N} \xi_i \ , \qquad \xi_i = \begin{cases} 1 \ , & if \ \ X_i = 0 \ , Y_i = 1 \\ 0 \ , & otherwise \end{cases} \tag{۴-۳}$$

$$d = \sum_{i=1}^{N} \xi_i \ , \qquad \xi_i = \begin{cases} 1 \ , & if \ \ X_i = Y_i = 0 \\ 0 \ , & otherwise \end{cases} \tag{۵-۳}$$

این مقادیر با هم ترکیب شده و توابع متفاوتی همانند روابط (۳-۶) تا (۳-۹) ایجاد می‌کنند.

- روش راسل[۱] و روآ[۲]

$$f = \frac{a}{a+b+c+d} \tag{۶-۳}$$

- روش جاکارد[۳] و نیدهام[۴]

$$f = \frac{a}{a+b+c} \tag{۷-۳}$$

- روش سوکال[۵] و میچنر[۶]

$$f = \frac{a+d}{a+b+c+d} \tag{۸-۳}$$

- روش یول[۷]

$$f = \frac{ad-bc}{ad+bc} \tag{۹-۳}$$

### ۳-۲-۲-۲ تطابق R بیت پیوسته[۸]

در روش rbc که توسط پرکاس[۹] در سال ۱۹۹۳ معرفی شد [۵۳]، $X$ و $Y$ دو رشته به طول یکسان و تعریف شده روی مجموعه الفبای محدود هستند. حاصل عبارت $match(X,Y)$ زمانی صحیح است که $X$ و $Y$ در حداقل $r$ بیت پیوسته مطابق هم باشند. به عنوان مثال اگر:

X = ABADCBAB

[1] Russel
[2] Rao
[3] Jacard
[4] Needham
[5] Sokal
[6] Michener
[7] Yule
[8] R-Contiguous Bits Matching (rbc)
[9] Percus

Y = CAGDCBBA

اگر $r \leq 3$ باشد *match(X,Y)* صحیح است و اگر $r > 3$ باشد این مقدار نادرست است.

### ۳-۲-۲-۳ قوانین تطابق بردار مقادیر حقیقی[1]

#### ۱) فاصله اقلیدسی

$$d(x,y) = \sqrt{\sum_i (X_i, Y_i)^2} = \|x - y\| \qquad (۱۰-۳)$$

فاصله اقلیدسی زمانی که ابعاد، وزن یکسان ندارند با ضرب هر جز از بردارها در یک وزن خاص اصلاح می‌شود.

#### ۲) فاصله Minkowski

این روش با عنوان فاصله نُرم $\lambda$ نیز شناخته شده است.

$$d(x,y) = \left(\sum |X_i - Y_i|^{\lambda}\right)^{1/\lambda} \qquad (۱۱-۳)$$

در رابطه (۳-۱۱)، اگر $\lambda=1$ باشد، تبدیل به فاصله منهتن می‌شود و اگر $\lambda = 2$ برابر با فاصله اقلیدسی است.

#### ۳) فاصله Chebyshev

این نوع فاصله به نام فاصله نُرم بینهایت نیز شناخته می‌شود و به صورت $D_\infty$ نشان داده می‌شود و به عنوان بیشترین تفاضل برای همه ویژگی‌ها تعریف می‌گردد.

$$d(x,y) = max\{|X_i - Y_i| \ , \ for \quad i = 1, \dots, n\} \qquad (۱۲-۳)$$

### ۳-۲-۳ الگوریتم انتخاب منفی

اولین الگوریتم انتخاب منفی توسط فارست[2] در سال ۱۹۹۴ معرفی شده است [۴۸]. این الگوریتم الهام گرفته شده از فرآیند بلوغ سلول‌های $T$ و انتخاب مجموعه‌ای از این سلول‌های بالغ که قادرند تنها با آنتی‌ژن‌های غیرخودی پیوند برقرار کنند، الهام گرفته شده است.

[1] Real-Valued Vector
[2] Forrest

همان‌طور که در الگوریتم شکل ۳-۱۱ و فلوچارت شکل ۳-۱۲، نمایش داده شده است، نقطه آغازین مشترک در تمامی الگوریتم‌های انتخاب منفی، تولید یک مجموعه از سلول‌های خودی $S$ است. در مرحله بعد یک مجموعه آشکارساز به نام $P$ که توانایی شناسایی سلول‌های خودی $S$ را دارند به صورت تصادفی، تعریف می‌گردد. سپس میل ترکیبی بین تک تک اعضای مجموعه $S$ با این آشکارسازها اندازه‌گیری می‌شود اگر این میل ترکیبی اندازه‌گیری شده از آستانه تعریف شده بیشتر باشد یعنی آشکارساز توانایی شناسایی سلول خودی را دارد لذا این آشکارساز حذف می‌شود، در غیر این صورت آشکارساز به مجموعه $D$ اضافه می‌گردد و در نهایت این الگوریتم به یک مجموعه داده جدیدی برای دسته‌بندی آن‌ها به سلول‌های خودی و غیرخودی اعمال می‌شوند.

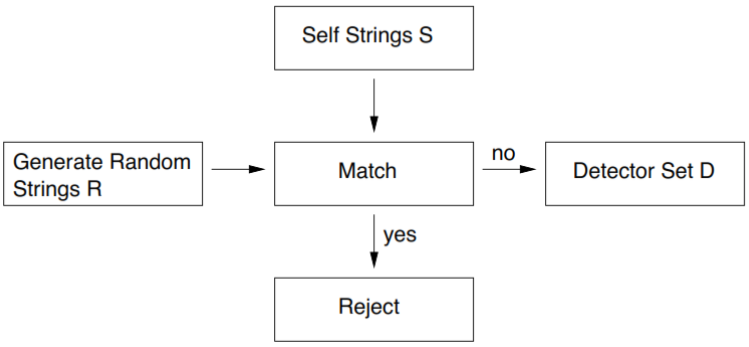
**ورودی: S** {مجموعه عناصرخودی دیده شده}
**خروجی: D** {مجموعه آشکارسازهای تولید شده}
**شروع**
**تکرار**
تولید آشکارسازهای بالقوه و اضافه کردن به مجموعه P
اندازه‌گیری میل ترکیبی میان عناصر مجموعه P و هر عضو مجموعه خودی S
**اگر** حداقل یک عضو از S یک آشکارساز از مجموعه P را با توجه به آستانه شناسایی تعیین شده، تشخیص دهد:
آشکارساز حذف می‌شود
**در غیر این صورت**
به مجموعه آشکارسازهای D اضافه می‌شود
**پایان**
**تا** (رسیدن به شرط توقف)
**پایان**

شکل ۳-۱۱: الگوریتم انتخاب منفی

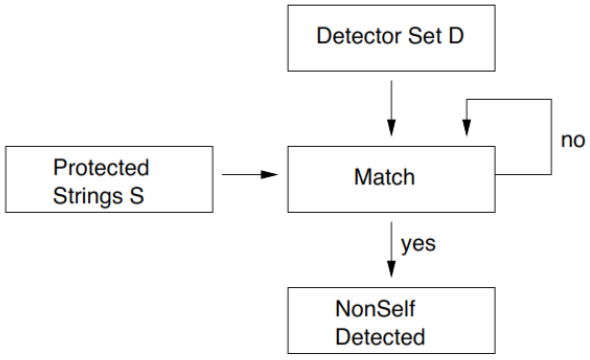
الگوریتم انتخاب منفی از رشته‌های دودویی برای نمایش سلول‌های خودی و آشکارسازهای مکملشان و برای اندازه‌گیری میل ترکیبی بین آن‌ها از روش تطابق $R$ بیت پیوسته استفاده می‌کند. الگوریتم انتخاب منفی را در حالتی که مقادیر به صورت حقیقی هستند نیز می‌توان به کار برد [۵۴].

دو مورد از محدودیت‌های ذکر شده برای الگوریتم انتخاب منفی در مراجع مختلف شامل:

- کارایی ضعیف‌تر این الگوریتم نسبت به روش‌های آماری تشخیص ناهنجاری [۵۵]
- دشواری ایجاد آشکارسازها در فضاها با ابعاد بزرگ [۵۶]




شکل (الف)




شکل (ب)

شکل ۳-۱۲: شناسایی الگو با الگوریتم انتخاب منفی، شکل (الف): تولید آشکارسازها، شکل (ب): بررسی و انتخاب الگوهای خودی و غیرخودی [۵۷]

### ۳-۲-۳-۱ کاربردهای الگوریتم انتخاب منفی

از جمله کاربردهایی که این الگوریتم دارد شامل [۵۸]:

- امنیت کامپیوتر برای تشخیص دستکاری داده‌ها ناشی از حملات ویروس‌ها
- امنیت شبکه
- شناسایی خطا و تشخیص

- کامپیوترهای DNA

### ۳-۲-۴ نظریه انتخاب کلونال

الگوریتم کلونال الهام گرفته شده از فرآیند بلوغ میل ترکیبی سلول‌های $B$ و فرآیند فراجهش می‌باشد. الگوریتم AIS از ایده سلول‌های حافظه برای نگه‌داری راه حل‌های خوب مسأله بهره می‌گیرد. دیکاسترو و تیمیس [۴۵] دو ویژگی مهم برای بلوغ میل ترکیبی در سلول‌های $B$ را بیان کرده‌اند. اولین ویژگی تکثیر سلول‌های $B$ وابسته به شدت میل ترکیبی آن‌ها با آنتی‌ژنی که به آن متصل شده‌اند و دومین ویژگی میزان جهش آنتی‌بادی سلول $B$ است که این میزان متناسب با عکس میزان میل-ترکیبی آن آنتی‌بادی با آنتی‌ژن می‌باشد. یعنی فرآیندهای انتخاب و جهش وابسته به میل ترکیبی آنتی‌بادی و آنتی‌ژن است. در مرجع [۵۹] الگوریتم CLONALG برای کاربردهایی نظیر تطابق الگو، یادگیری ماشین و بهینه سازی توابع چند هدفه[1] معرفی شده است. این الگوریتم در شکل ۳-۱۳، نمایش داده شده است.

ورودی‌های این الگوریتم یک مجموعه از الگوها به نام $S$ به عنوان آنتی‌ژن و تعداد بدترین عنصر-هایی که برای جایگزینی انتخاب می‌شوند $n$، می‌باشند. خروجی الگوریتم یک مجموعه حافظه آنتی-بادی $M$ است که با اعضای مجموعه $S$ مطابقت دارند. ابتدا یک مجموعه تصادفی آنتی‌بادی ایجاد می-گردد، سپس به ازای تمامی آنتی‌ژن‌های مجموعه $S$، میل ترکیبی آن‌ها را با اعضای مجموعه $A$ محاسبه می‌کنیم و متناسب با اندازه میل ترکیبی محاسبه شده برای هر آنتی‌بادی، عمل کلونی‌سازی از آن آنتی‌بادی انجام می‌گردد و مطابق با عکس میل ترکیبی عمل جهش صورت می‌گیرد سپس یک کپی از آنتی‌بادی با بیشترین میل ترکیبی به عنوان سلول حافظه در مجموعه $M$ قرار می‌گیرد. در نهایت $n$ آنتی‌بادی با کمترین میل ترکیبی با آنتی‌بادی‌های تصادفی جایگزین می‌شوند.

[1] Multi Modal

**ورودی: S** = مجموعه الگوها (آنتی‌ژن‌ها)، **n** = تعداد بدترین عناصری کـه بـرای جـایگزینی انتخاب می‌شوند.

**خروجی: M** = سلول‌های حافظه که توانایی کلاسه‌بندی الگوهای دیده نشده را دارند.

**شروع**

تولید مجموعه تصادفی ازآنتی‌بادی (A)

**تکرار** (برای همه الگوهای مجموعه S):

محاسبه میل ترکیبی آن‌ها با هر آنتی‌بادی مجموعه A

تولید کلون از آنتی‌بادی‌هایی که میل ترکیبی بالاتری دارند. (تعداد کلون‌هـا بـرای هر آنتی‌بادی متناسب با میل ترکیبی آن‌ها است.)

جهش کلون‌های تولید شده متناسب با عکس میل ترکیبـی آن‌هـا و اضـافه کـردن آن‌ها به مجموعه A . ذخیره یک کپی از آنتی‌بادی با میل ترکیبی بالاتر از مجموعه A در مجموعه M

جایگزینی تا از آنتی‌بادی‌ها با کمترین میل ترکیبی از مجموعه A با آنتی‌بادی‌هـای تصادفی

**پایان**

**پایان**

شکل ۳-۱۳: الگوریتم GlONALG

### ۳-۲-۴-۱ کاربردهای انتخاب کلونال

از جمله کاربردهایی که این الگوریتم دارد شامل [۵۸]:

- بهینه‌سازی توابع یک هدفه، چند هدفه، ترکیبی و متغیر
- مقداردهی اولیه مراکز توابع پایه ای شعاعی
- انواع مختلف شناسایی الگو
- مسایل رنگ‌آمیزی گراف
- برنامه‌ریزی خودکار[1]
- مسایل شناسایی کاراکتر چندگانه

[1] Automated Scheduling

- کلاسه‌بندی اسناد
- پیش‌بینی سری‌های زمانی[1]

### ۳-۲-۵ خلاصه‌ای از کاربردهای مختلف الگوریتم AIS

کاربردهای الگوریتم سیستم ایمنی مصنوعی به صورت زیر دسته‌بندی می‌شوند [۶۰]:

۱) یادگیری: خوشه‌بندی، کلاسه‌بندی، شناسایی الگو، زمینه‌های کنترل و رباتیک

۲) تشخیص ناهنجاری: عیب یابی، برنامه‌های امنیت کامپیوتر و شبکه

۳) بهینه‌سازی: پیوسته و ترکیبی

[1] Time Series Forecasting

# فصل۴ کلاسه‌بند ماشین بردار پشتیبان

## ۴-۱ مقدمه

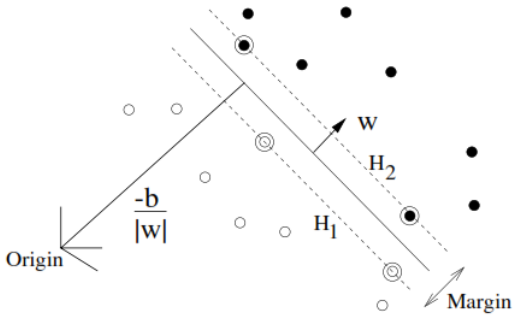
در حال حاضر ماشین بردار پشتیبان یکی از موضوعات مهم در یادگیری ماشین و یکی از محبوب‌ترین روش‌های طبقه‌بندی باناظر می‌باشد. این روش توانایی زیادی در کلاسه‌بندی غیرخطی[۱]، رگرسیون و تشخیص داده‌های پرت[۲] دارد. یک مسأله کلاسه‌بندی عموماً شامل داده‌های مجزا برای آموزش و تست است. هر نمونه در داده‌های آموزش یک مقدار هدف[۳] (برچسب کلاس) و چندین صفت (ویژگی‌ها یا مقادیر مشاهده شده) دارد. هدف SVM تولید یک مدل بر اساس داده‌های آموزش است که مقادیر هدف را در داده‌های تست پیش‌بینی کنند.

ماشین بردار پشتیبان در سال ۱۹۹۵ توسط کرتس[۴] و وپنیک[۵] [۶۱] برای کلاسه‌بندی دودویی معرفی شد و در زمینه‌هایی نظیر شناسایی آماری الگو [۶۲]، دسته‌بندی متن [۳۶] و کلاسه‌بندی تصاویر ابر طیفی [۶۳] کاربرد دارد. در این روش مراحلی به شرح زیر وجود دارند [۶۴]:

- تفکیک کلاس: اساساً به دنبال یک ابر صفحه جداکننده[۶] بهینه با بیشترین حاشیه[۷] بین دو کلاس هستیم، مانند ( شکل ۴-۱).
- کلاس‌های دارای همپوشانی: تأثیر داده‌هایی که به اشتباه به سمت مخالف حاشیه جداساز رفته‌اند با کاهش وزنشان کمتر می‌شود ( در حالت حاشیه نرم[۸]).
- غیرخطی بودن: زمانی‌که یک مرز خطی نمی‌یابیم داده‌ها را به یک فضا با ابعاد بزرگتر نگاشت داده در این فضای جدید قادر به جداسازی خطی داده‌ها هستیم.
- راه حل مسأله: تمام این کارها را می‌توان به عنوان یک مسأله بهینه‌سازی کوآدراتیک[۹] از

[1] Non-Linear Classification
[2] Outlier Detection
[3] Target Value
[4] Cortes
[5] Vapnik
[6] Separating Hyperplane
[7] Margin
[8] Soft Margin
[9] Quadratic Optimization Problem

طریق روش‌های شناخته شده فرمول‌بندی و حل کرد.




شکل ۴-۱: ابرصفحه جداساز برای داده‌های جداپذیر خطی در SVM [۶۲]

به طور کلی SVM یک طبقه‌بند دودویی خطی است که با توسعه آن و استفاده از توابع کرنل به عنوان یک طبقه‌بند چند کلاسه خطی و غیرخطی به کار می‌رود. در ادامه به معرفی ساده‌ترین مسأله یعنی SVM خطی روی داده‌های جداپذیر آموزشی می‌پردازیم [۶۲].

## ۴-۲ ماشین بردار پشتیبان خطی

داده‌های آموزشی به صورت $\{x_i, y_i\}, i = 1, \dots, l, y_i \in \{-1,1\}, x_i \in R^d$ برچسب گذاری شده‌اند. فرض کنیم یک ابرصفحه نمونه‌های مثبت را از نمونه‌های منفی جدا می‌کند، نقاط $x$ که روی ابرصفحه قرار دارند شرط $wx + b = 0$ را ارضا می‌کنند، $w$ بردار نرمال ابرصفحه، $\frac{b}{\|w\|}$ فاصله عمودی مبدأ از ابرصفحه و $\|w\|$ بردار نرمال اقلیدسی $w$ می‌باشد. اگر $d_+$ $(d_-)$ کمترین فاصله نقاط مثبت (منفی) از ابرصفحه باشند حاشیه یک ابرصفحه جداکننده به صورت $d_+ + d_-$ تعریف می‌شود. برای مورد جدایی‌پذیر خطی، الگوریتم بردار پشتیبان ابرصفحه، بزرگترین حاشیه را پیدا می‌کند. این مسأله می‌تواند به راحتی به صورت زیر فرموله شود، اگر کلیه داده‌های آموزشی قیود زیر را ارضا کنند:

$$x_i.w + b \geq +1 \ , \quad for \ , \ y_i = +1 \qquad (۱-۴)$$

$$x_i.w + b \leq -1 \ , \quad for \ , \ y_i = -1 \qquad (۲-۴)$$

این دو قید به صورت رابطه (۴-۳) خلاصه می‌شوند:

$$y_i(x_i.w + b) - 1 \geq 0 \ , \quad \forall \, i \qquad (۳-۴)$$

حال نقاطی را در نظر می‌گیریم که در نامساوی (۴-۱) صدق می‌کنند. این نقاط روی ابرصفحه $H_1 = x_i.w + b = 1$ با بردار نرمال $w$ و فاصله عمودی $\frac{|1-b|}{\|w\|}$ از مبدأ قرار دارند. به طور مشابه نقاطی که در نامساوی (۴-۲) صدق می‌کنند در ابر صفحه $H_2 = x_i.w + b = -1$ با بردار نرمال $w$ و فاصله $\frac{|-1-b|}{\|w\|}$ از مبدأ قرار دارند. بنابراین $d_+ = d_- = \frac{1}{\|w\|}$ و مقدار حاشیه برابر با $\frac{2}{\|w\|}$ خواهد بود. دو ابرصفحه $H_1$ و $H_2$ موازی هستند (دارای بردارهای نرمال برابر می‌باشند). بنابراین می‌توان زوج ابرصفحه‌ای را پیدا کرد که بیشترین مقدار حاشیه را داشته باشند و این کار از طریق مینیمم کردن $\|w\|^2$ با توجه به قیود (۴-۳) صورت می‌گیرد.

آن دسته از نقاط آموزشی که در رابطه (۴-۳) صدق می‌کنند یعنی بر روی یکی از ابر صفحه‌های $H_1$ یا $H_2$ قرار می‌گیرند بردارهای پشتیبان[1] نامیده می‌شوند. این نقاط در شکل ۴-۱، با یک دایره اضافی مشخص شده‌اند.

حال به فرموله‌سازی لاگرانژ مسأله می‌پردازیم. به دو دلیل این کار انجام می‌گردد: اول اینکه قیود نشان داده شده در رابطه (۴-۳) با خود ضرایب لاگرانژ جایگزین خواهند شد، این جایگزینی ادامه کار را بسیار راحت‌تر خواهد کرد. دوم اینکه در این گونه فرموله نمودن مسأله، داده‌های آموزشی فقط به صورت ضرب نقطه‌ای بردارها پدیدار می‌شوند، این یک ویژگی حیاتی است که اجازه تعمیم فرآیند حل مسأله را به حالت غیر خطی می‌دهد.

بنابراین ضرایب لاگرانژ $\alpha_i, i = 1, \dots, l$ معرفی می‌شوند که هر یک برای یکی از قیود نامساوی (۴-۳) می‌باشند. برای قیود به فرم $c_i \geq 0$ ، معادلات محدودیت در ضرایب مثبت لاگرانژ ضرب شده و از تابع هدف کم می‌شوند. برای محدودیت‌های تساوی، ضرایب لاگرانژ محدود نمی‌شوند. لذا رابطه (۴-۴) را خواهیم داشت:

$$L_P \equiv \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} \alpha_i y_i (x_i.w + b) + \sum_{i}^{l} \alpha_i \qquad (۴-۴)$$

برای به دست آوردن $\alpha_i$های مناسب، تابع $L_P$ باید بر حسب $w$ و $b$ کمینه و نسبت به متغیرهای

[1] Support Vector

لاگرانژ غیرمنفی $\alpha_i$ بیشینه گردد.

$$w = \sum_i \alpha_i y_i x_i \quad for \quad \frac{\partial L_P}{\partial w} = 0 \tag{۴-۵}$$

$$\sum_i \alpha_i y_i \quad for \quad \frac{\partial L_P}{\partial b} = 0 \tag{۴-۶}$$

با جایگزینی دو رابطه (۴-۵) و (۴-۶) در رابطه (۴-۴):

$$L_D = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j x_i . x_j \tag{۴-۷}$$

دو فرمول $L_P$ و $L_D$ از یک تابع هدف ولی با محدودیت‌های متفاوت نتیجه گرفته می‌شوند و راه‌حل از طریق مینیمم کردن $L_P$ یا ماکزیمم کردن $L_D$ به دست می‌آید.

آموزش ماشین بردار پشتیبان (برای موارد خطی جدایی پذیر) همراه با ماکزیمم کردن $L_D$ بر اساس $\alpha_i$ که دارای محدودیت (۴-۶) و مقدار مثبت است و با توجه به راه حل (۴-۵) انجام می‌پذیرد. در راه حل، نقاطی که برای آن‌ها $\alpha_i > 0$، بردار پشتیبان نامیده می‌شوند و روی یکی از ابر صفحه‌های $H_1$ یا $H_2$ قرار می‌گیرند. بقیه نقاط آموزشی دارای مقدار $\alpha_i = 0$ می‌باشند. در این ماشین‌ها، بردارهای پشتیبان عناصر کلیدی مجموعه آموزش می‌باشند که نزدیک‌ترین نقاط به مرز تصمیم‌گیری هستند و اگر بقیه نقاط آموزشی همگی حذف شوند و آموزش دوباره تکرار شود، همان ابرصفحه جداکننده به دست می‌آید.

### ۴-۲-۱ روش Karush-Kuhn-Tucker (KKT)

این روش نقش مهمی هم در تئوری و هم در عمل در مسایل مربوط به بهینه‌سازی مقید دارد. برای مسأله گفته شده پیشین، شرایط KKT به صورت زیر توصیف می‌شوند [۶۵]:

$$\frac{\partial}{\partial \omega_v} L_P = \omega_v - \sum_i \alpha_i y_i x_{iv} = 0 \ , \qquad v = 1, \dots, d \tag{۴-۸}$$

$$\frac{\partial}{\partial b} L_P = -\sum_i \alpha_i y_i = 0 \tag{۴-۹}$$

$$y_i(x_i . w + b) - 1 \geq 0 \ , \qquad i = 1, \dots, l \tag{۴-۱۰}$$

$$\alpha_i \geq 0 \ , \quad \forall i \tag{۴-۱۱}$$

$$\alpha_i(y_i(x_i.w+b)-1)=0\,,\qquad \forall i \tag{۱۲-۴}$$

شرایط KKT به راحتی برای حل مسأله مربوط به بردارهای پشتیبان می‌تواند به کار رود، زیرا در این مسأله محدودیت‌ها همواره خطی هستند. علاوه بر این مسأله SVM محدب می‌باشد، و برای مسایل محدب، شرایط KKT لازم و کافی‌اند. بنابراین، حل SVM معادل با پیدا کردن راه حل برای شرایط KKT می‌باشد.

پارامتر $w$ به طور صریح توسط رویه آموزش[1] مشخص شده است، پارامتر $b$ به سادگی با به کارگیری رابطه (۴-۱۲) از شرایط KKT به دست می‌آید این کار با انتخاب $i$ هایی که به ازای آن‌ها $\alpha_i \neq 0$ صورت می‌گیرد.

### ۴-۲-۲ مرحله تست

برای تست یک بردار پشتیبان آموزش داده شده، بررسی می‌شود الگوی تست ورودی $x$ در کدام سمت مرز تصمیم‌گیری یعنی ابر صفحه‌ای که بین $H_1$ و $H_2$ قرار دارد و موازی با آن‌ها است، قرار می‌گیرد و به آن برچسب کلاس مربوطه اختصاص داده می‌شود یعنی کلاس الگوی ورودی $x$ برابر با $sgn(x_i.w+b)$ است.

### ۴-۲-۳ شرایط جدایی ناپذیر

الگوریتم شرح داده شده برای داده‌های جداپذیر کاربرد دارد و زمانی‌که داده‌ها جدا ناپذیر باشند با این روش هیچ مجموعه جواب محتملی وجود ندارد. در نهایت هدف مناسب کردن قیود (۴-۱) و (۴-۲) برای حل این نوع مسایل است. این کار زمانی اتفاق می‌افتد که یک مقدار جریمه برای داده‌ای که به اشتباه درسوی دیگری از مرز جداساز قرار دارد در نظر گرفته شود، بنابراین مقدار مثبت $\xi_i\,, i=1,\dots,l$ در محدودیت‌ها به صورت زیر اضافه می‌شود.

$$x_i.w+b\geq +1-\xi_i\ ,\qquad for\ ,\ y_i=+1 \tag{۱۳-۴}$$

[1] Training Procedure

$$x_i.w + b \geq -1 + \xi_i \ , \qquad for \ , \ y_i = -1 \tag{۱۴-۴}$$

$$\xi_i \geq 0 \ , \quad \forall i \tag{۱۵-۴}$$

اگر مقدار $\xi_i$ از مقدار واحد بیشتر شود، خطا رخ می‌دهد لذا $\sum_i \xi_i$ حد بالایی برای تعداد خطاهای آموزش است. بنابراین تابع هدف به کمینه کردن عبارت $\frac{\|w\|^2}{2} + C(\sum_i \xi_i)^k$ تغییر می‌یابد، $C$ مقدار ثابت حاشیه نرم است که توسط کاربر انتخاب می‌شود مقادیر بزرگتر برای $C$، متناظر با تخصیص مقدار بیشتر جریمه برای خطا است. این یک مسأله برنامه‌نویسی محدب برای مقادیر صحیح $k$ است. به ازای $k = 2$ و $k = 1$، این مسأله یک مسأله برنامه‌نویسی کوآدراتیک می‌باشد.

هدف بیشینه‌سازی رابطه (۴-۱۶) است:

$$L_D \equiv \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j x_i.x_j \tag{۱۶-۴}$$

به شرطی که:

$$0 \leq \alpha_i \leq C \tag{۱۷-۴}$$

$$\sum_i \alpha_i y_i = 0 \tag{۱۸-۴}$$

و در نهایت به پاسخ زیر می‌رسیم:

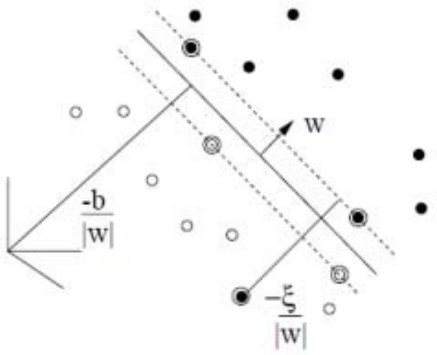
$$w = \sum_{i=1}^{NS} \alpha_i y_i x_i \tag{۱۹-۴}$$

در رابطه (۴-۱۹)، $Ns$ تعداد بردارهای پشتیبان را مشخص می‌کند. بنابراین، تنها تفاوت نسبت به حالت ابرصفحه بهینه این است که اکنون $\alpha_i$ دارای حد بالای $C$ می‌باشد. این وضعیت در شکل ۴-۲ نشان داده شده است.

برای حل این مسأله نیاز به شرایط KKT داریم:

$$L_P = \frac{1}{2}\|w\|^2 + C\left(\sum_i \xi_i\right) - \sum_{i=1}^{l} \alpha_i\{y_i(x_i.w + b) - 1 + \xi_i\} - \sum_i \mu_i \xi_i \tag{۲۰-۴}$$

در رابطه (۴-۲۰)، $\mu_i$ ضرایب لاگرانژ است که موجب می‌شود مقدار $\xi_i$ همواره مثبت باشد.




شکل ۴-۲: ابرصفحه جداساز خطی برای داده‌های جداناپذیر خطی [۶۲]

$$\frac{\partial}{\partial \omega_v} L_P = \omega_v - \sum_i \alpha_i y_i x_{iv} = 0 \quad (۲۱-۴)$$

$$\frac{\partial}{\partial b} L_P = -\sum_i \alpha_i y_i = 0 \quad (۲۲-۴)$$

$$\frac{\partial}{\partial \xi_i} L_P = C - \alpha_i - \mu_i = 0 \quad (۲۳-۴)$$

$$y_i(x_i.w + b) - 1 + \xi_i \geq 0 \quad (۲۴-۴)$$

$$\xi_i \geq 0 \quad (۲۵-۴)$$

$$\alpha_i \geq 0 \quad (۲۶-۴)$$

$$\mu_i \geq 0 \quad (۲۷-۴)$$

$$\alpha_i[y_i(x_i.w + b) - 1 + \xi_i] = 0 \quad (۲۸-۴)$$

$$\mu_i \xi_i = 0 \quad (۲۹-۴)$$

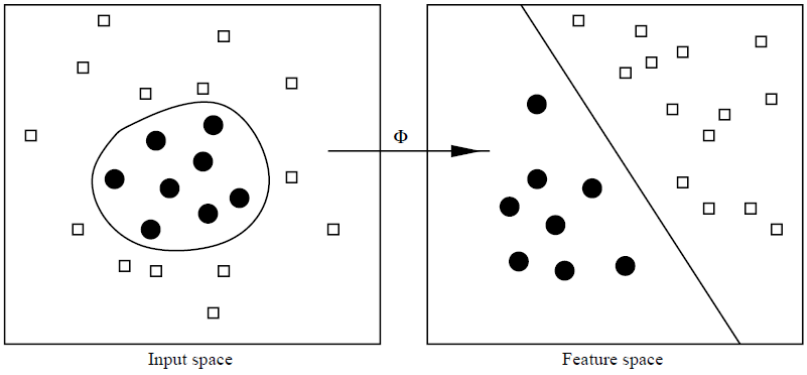
در این روابط $i$ مقداری بین یک تا تعداد نقاط آموزش دارد، و مقادیر $v$ بین یک تا ابعاد داده می‌باشد. از معادلات (۴-۲۸) و (۴-۲۹) برای محاسبه $b$ استفاده می‌شود.

## ۴-۳ ماشین بردار پشتیبان غیرخطی

در این قسمت در مواردی که تابع تصمیم‌گیری، تابعی خطی از داده‌ها نباشد روش‌های قبل برای رسیدن به جواب تعمیم داده می‌شوند. داده‌ها در قسمت آموزش (روابط(۴-۱۶)-(۴-۱۸)) به صورت ضرب نقطه‌ای $x_i, y_i$ می‌باشند. حال ابتدا داده‌ها به یک فضای اقلیدسی دیگر (در برخی موارد با ابعاد نامتناهی) به نام $H$، با استفاده از $\varphi$، همانند (شکل ۴-۳)، نگاشت داده می‌شوند:

$$\varphi : R^d \to H \qquad (۳۰-۴)$$

در این حالت الگوریتم آموزش تنها به ضرب نقطه‌ای داده‌ها در $H$، یعنی $\varphi(x_i).\varphi(x_j)$ وابسته است. یک تابع کرنل[1] $K$ به شرط $K(x_i, y_i) = \varphi(x_i).\varphi(y_i)$ داریم، آن‌گاه تنها لازم است از این تابع کرنل به جای ضرب نقطه‌ای در الگوریتم یادگیری استفاده شود و حتی نیاز نیست دقیقاً بدانیم تابع نگاشت $\varphi$ چگونه تابعی بوده است. سپس در همه قسمت‌های الگوریتم آموزش، $x_i, y_i$ را با $K(x_i, y_i)$ جایگزین می‌شود. برای آموزش بردار پشتیبانی که در فضای با ابعاد نامتناهی قرار دارد همان مقدار زمان لازم است که برای آموزش داده‌های اولیه (نگاشت نشده) زمان لازم بود. کلیه قواعد و شرایط قسمت‌های گذشته برقرار است، زیرا همچنان در حال انجام یک جداسازی خطی، اما در یک فضای متفاوت هستیم.




شکل ۴-۳: نگاشت غیرخطی از فضای ورودی به فضای ویژگی با ابعاد بالاتر [۲]

در فاز تست، SVM از طریق محاسبه ضرب نقطه‌ای یک نقطه تست $x$ با $w$ به کار برده می‌شود، یا به طور دقیق‌تر با محاسبه تابع زیر:

$$f(x) = \sum_{i}^{Ns} \alpha_i\, y_i \varphi(s_i).\varphi(x) + b = \sum_{i}^{Ns} \alpha_i\, y_i K(s_i, x) + b \qquad (۳۱-۴)$$

در رابطه (۴-۳۱)، $S_i$ها بردارهای پشتیبان هستند. بنابراین می‌توان از محاسبه $\varphi(x)$ به طور مستقیم اجتناب کرده و به جای آن $K(s_i, x) = \varphi(s_i).\varphi(x)$ را به کار برد.

[1] Kernel Function

## ۴-۴ معرفی انواع کرنل‌های SVM

همان‌طور که قبلاً گفته شد توابع کرنل ضرب داخلی دو نقطه در فضای مناسب ویژگی‌ها را برمی‌گردانند. بنابراین سعی در تعریف شباهت‌ها با استفاده از کرنل‌ها داریم به گونه‌ای که هزینه محاسباتی در فضای ویژگی‌ها با ابعاد بزرگ نیز کم باشد. انواع کرنل‌ها شامل [۶۶]:

### ۴-۴-۱ کرنل خطی[1]

$$K(x, x') = \langle x.x' \rangle \quad (۳۲-۴)$$

### ۴-۴-۲ کرنل گوسی تابع پایه شعاعی[2]

$$K(x, x') = \exp\left(-\sigma \|x - x'\|^2\right) \quad (۳۳-۴)$$

### ۴-۴-۳ کرنل چندجمله‌ای[3]

$$K(x, x') = (scale.\langle x.x' \rangle + offset)^{degree} \quad (۳۴-۴)$$

### ۴-۴-۴ کرنل سیگموئید[4]

$$K(x, x') = tanh(scale.\langle x.x' \rangle + offset) \quad (۳۵-۴)$$

### ۴-۴-۵ کرنل تابع بسل[5]

$$K(x, x') = \frac{Bessel^{n}_{(v+1)} - (\sigma \|x - x'\|)}{(\|x - x'\|)^{-n(v+1)}} \quad (۳۶-۴)$$

### ۴-۴-۶ کرنل لاپلاس تابع پایه شعاعی[6]

$$K(x, x') = \exp\left(-\sigma \|x - x'\|\right) \quad (۳۷-۴)$$

[1] Linear Kernel
[2] Gaussian Radial Basis Function (RBF)
[3] Polynomial
[4] Sigmoid Kernel
[5] Bessel Function
[6] Laplace Radial Basis Function (RBF)

### ۴-۴-۷ کرنل پایه شعاعی ANOVA[1]

$$K(x,x') = \left(\sum_{k=1}^{n} exp(-\sigma(x^k - x'^k)^2)\right)^d \quad (۳۸-۴)$$

کرنل‌های لاپلاس و گوسین پایه شعاعی و کرنل بسل، کرنل‌هایی همه منظوره هستند و زمانی‌که دانش قبلی در مورد داده‌ها وجود ندارد استفاده می‌شوند. کرنل خطی ساده‌ترین نوع کرنل‌ها است، زمانی که بردارهای داده بزرگ و پراکنده هستند مفید است و معمولاً در طبقه‌بندی متن موفق‌تر عمل می‌کند. کرنل چندجمله‌ای کرنل مشهور در پردازش تصویر است. از کرنل سیگموئید به عنوان پروکسی برای شبکه عصبی استفاده می‌شود و کرنل پایه شعاعی ANOVA در مسایل رگرسیون عملکرد مناسبی دارد.

### ۴-۴-۸ نتیجه بررسی انواع کرنل‌ها در SVM

در استفاده از SVM سه مسأله مهم وجود دارد:

۱) انتخاب تابع کرنل مناسب

۲) انتخاب مقادیر بهینه برای پارامترهای کرنل و پارامتر پنالتی $C$

۳) انتخاب زیرمجموعه مناسب برای آموزش و تست در SVM

اعمال سه مورد بالا در هنگام استفاده از SVM باعث افزایش دقت خواهد شد [۷]. با توجه به کاربردهای مختلفی که در آن‌ها از SVM استفاده می‌شود انواع کرنل‌های مختلف نتایج متفاوتی ارائه خواهند داد. اگر پارامترهای کرنل انتخابی و پارامتر پنالتی $C$ به صورت بهینه انتخاب نشوند مثلاً پارامتر پنالتی $C$ خیلی بزرگ یا خیلی کوچک باشد خطای کلاسه‌بندی SVM افزایش می‌یابد. بنابراین در مسایل مختلفی که ماشین بردار پشتیبان به عنوان ابزاری مناسب جهت کلاسه‌بندی و تخمین به کار می‌رود همواره به دنبال روشی جهت بهینه‌سازی پارامترهای آن نیز هستیم. با توجه به نتایج به دست آمده در مرجع [۶۷]، کرنل RBF با پارامترهای بهینه بهترین نتیجه را در کاربردهای

[1] ANOVA Radial Basis

مختلف نسبت به سایر کرنل‌ها دارد. این کرنل برخلاف کرنل خطی توانایی کلاسه‌بندی داده‌های چند بعدی را داشته و نسبت به کرنلی نظیر چندجمله‌ای نیاز به پارامترهای ورودی بهینه کمتری دارد.

## ۴-۵ روش‌های طبقه‌بندی چند کلاسه با استفاده از SVM

تعدادی از روش‌ها برای تولید SVM چندکلاسه از SVM دودویی توسط محققان مطرح شده-است و هنوز به عنوان یک موضوع در حال تحقیق ادامه دارد.

### ۴-۵-۱ رویکرد یکی در برابر همه[1]

این روش در سال ۱۹۹۵ توسط وپنیک معرفی شد که در آن یک کلاس را با سایر کلاس‌ها مقایسه می‌کند. این روش به نام کلاسه‌بند Winner-Take-All هم نامیده می‌شود. فرض کنیم یک مجموعه داده به $M$ کلاس طبقه‌بندی می‌شود بنابراین $M$ تا کلاسه‌بند SVM دودویی ایجاد می‌گردد که هر کلاسه‌بند یک کلاس را از بقیه $M\text{-}1$ کلاس متمایز می‌کند. خروجی نهایی کلاس متعلق به SVM با بیشترین حاشیه است. مزیت این روش این است که: تعداد کلاسه‌بند دودویی مورد نیاز برابر با تعداد کلاس‌ها است یعنی اگر تعداد کلاس‌ها $n$ تا باشد تعداد کلاسه‌بندهای دودویی نیز $n$ تا خواهد بود و نیاز به حل $n$ مسأله بهینه‌سازی کوآدراتیک وجود دارد. اولین اشکال این روش، نیاز به حافظه بسیار زیاد در طول مرحله آموزش است. این مقدار حافظه برابر با مربع تعداد کل داده‌های آموزشی می‌باشد که در صورت حجیم بودن داده‌های آموزشی مشکل ساز خواهد بود. دومین اشکال این روش این است: اگر فرض شود $M$ کلاس وجود دارد و هر کلاس تعداد داده‌های آموزشی یکسان دارند در طول فرآیند آموزش نسبت نمونه‌های آموزشی از یک کلاس به بقیه کلاس‌ها $\frac{1}{M-1}$ است که نشان-دهنده اندازه ناموزون نمونه‌های آموزشی است [۶۸].

[1] One Versus All (OVA)

### ۴-۵-۲ رویکرد یکی در برابر یکی[1]

در این روش، کلاسه‌بند SVM برای همه جفت کلاس‌های ممکن ایجاد می‌شود [۶۹]. خروجی برای هر کلاسه‌بند برچسب کلاس مشاهده شده است یعنی برچسب کلاسی که بیشتر به نقاط یک بردار داده اختصاص داده شده است، انتخاب می‌شود. در صورت تساوی بین برچسب‌های مختلف، از یک روش شکستن تساوی استفاده می‌شود. از جمله روش‌های رایج برای این کار انتخاب تصادفی یکی از برچسب‌های کلاس‌های مساوی است.

تعداد کلاسه‌بندهایی که در این روش تولید می‌شوند نسبت به روش OVA بیشتر است و تعداد داده‌های آموزشی و حافظه مورد نیاز نسبت به OVA بسیار کمتر است، لذا این روش نسبت به OVA مناسب‌تر خواهد بود. اشکال اصلی این روش افزایش تعداد کلاسه‌بندها با افزایش تعداد کلاس‌ها است به عنوان مثال تمام کلاسه‌بندهای دودویی ممکن از $n$ تا کلاس برابر با $\frac{n(n-1)}{2}$ است. اعمال هر کلاسه‌بند به داده‌های تست یک رأی را به کلاس برنده می‌دهد [۷۰].

### ۴-۵-۳ رویکرد گراف تصمیم بدون دور مستقیم[2]

پلت[3] [۷۱] یک روش کلاسه‌بند چند کلاسه را بر اساس DDAG با ساختاری شبیه ساختار درخت، به نام ماشین بردار پشتیبان گراف بدون دور مستقیم[4] پیشنهاد داده است. روش DDAG در اصل مشابه کلاسه‌بند دودویی عمل می‌کند، یعنی برای $M$ کلاس به تعداد $\frac{M(M-1)}{2}$ کلاس‌بندهای دودویی ایجاد می‌کند. هر کلاسه‌بند SVM دودویی به عنوان یک گره در ساختار گراف قرار می‌گیرد. در این گراف گره‌ها در یک ساختار مثلث‌وار با یک گره ریشه در نوک مثلث و افزایش یک به یک گره-ها در هر لایه تا رسیدن به لایه آخر با $M$ گره، قرار می‌گیرند. در این روش در مرحله آموزش مانند روش OVO به تعداد $\frac{M(M-1)}{2}$ ، SVM دودویی و در مرحله تست از یک گراف دودویی بدون دور

[1] One Versus One(OVO)
[2] Decision Directed Acyclic Graph (DDAG)
[3] Platt
[4] Directed Acyclic Graph SVM (DAGSVM)

مستقیم که $\frac{M(M-1)}{2}$ گره داخلی و $M$ برگ دارد، استفاده می‌شود. یک مزیت DAG نسبت به OVO کوتاه‌تر بودن زمان تست آن است [۷۰].

### ۴-۵-۴ رویکرد تصحیح خطا بر اساس کد خروجی[1]

مفهوم ECOC مسایل چندکلاسه همان اعمال کلاسه‌بند باینری است. این رویکرد با تبدیل مسأله کلاسه‌بندی $M$ کلاسه به $L$ تا مسأله کلاسه‌بندی دو کلاسه عمل می‌کند. این روش یک کد کلمه‌ای[2] یکتا را به جای برچسب کلاس به هر کلاس نسبت می‌دهد. کد تصحیح خطا $(L, M, d)$ یک کد به طول $L$ بیت با $C$ کلمه یکتا با فاصله همینگ $d$ است. فاصله همینگ بین دو کلمه برابر با تعداد بیت‌های متناظر به مقادیر متفاوت در دو کلمه است و $M$ تعداد کلاس‌ها است [۶۸].

### ۴-۵-۵ نتیجه بررسی انواع روش‌های چند کلاسه SVM

با توجه به نتایج حاصل در [۶۸]، روش‌های OVO و DAG دقت قابل مقایسه‌ای را ارائه می‌دهند و به منابع محاسباتی یکسانی نیاز دارند. زمان آموزش روش‌های DAG و OVO از OVA کمتر است. بالاترین دقت توسط ECOC جامع به دست آمده است ولی این روش به زمان آموزش خیلی زیادی نیاز دارد. مشکل اساسی با OVA تولید داده‌های کلاسه‌بندی نشده است که باعث کاهش دقت کلاسه‌بندی می‌گردد. در نهایت نتایج نشان دهنده مناسب بودن OVO از لحاظ دقت کلاسه‌بندی و هزینه محاسباتی است. بنابراین در الگوریتم پیشنهادی از این روش استفاده شده است.

## ۴-۶ کتابخانه LibSVM

کتابخانه LibSVM یک کتابخانه متن باز برای ماشین بردار پشتیبان می‌باشد که در سال ۲۰۰۰ در دانشگاه تایوان ایجاد شده است. هدف این کتابخانه تسهیل و بهینه‌سازی استفاده از SVM در پروژه‌های مختلف است.

[1] Error-Correcting Output Code (ECOC)
[2] Code Word

استفاده از LibSVM شامل دو مرحله است: ابتدا مجموعه داده را آموزش داده و یک مدل به دست می‌آورد و سپس از مدل به دست آمده برای پیش‌بینی کلاس داده‌های تست بهره می‌گیرد. این کتابخانه از الگوریتم بهینه‌سازی کمینه ترتیبی[1] برای حل مسأله کودراتیک در مرحله آموزش استفاده می‌کند. این کتابخانه در حوزه‌های یادگیری زیر استفاده می‌شود:

- کلاسه‌بندی بردارهای پشتیبان[2] (برای حالت دو کلاسه و چندکلاسه)
- رگرسیون بردارهای پشتیبان[3]
- ماشین بردار یک کلاسه[4]

کتابخانه LibSVM با استفاده از زبان Java نیز پیاده‌سازی شده است و رابط‌های گوناگونی برای استفاده از آن در محیط‌ها و زبان‌های مختلف از جمله متلب وجود دارد. کتابخانه LibSVM قابلیت استفاده در حالت چندکلاسه را دارد و به دلایلی که در قسمت قبل ذکر شد در این تحقیق از رویکرد یکی در برابر یکی (OVO) آن در حالت چندکلاسه استفاده می‌شود [۷۲].

[1] Sequential Minimal Optimization (SMO)
[2] Support Vector Classification (SVC)
[3] Support Vector Regression (SVR)
[4] One Class SVM ( Distribution Estimation)

# فصل۵ الگوریتم AIS در کاهش ویژگی

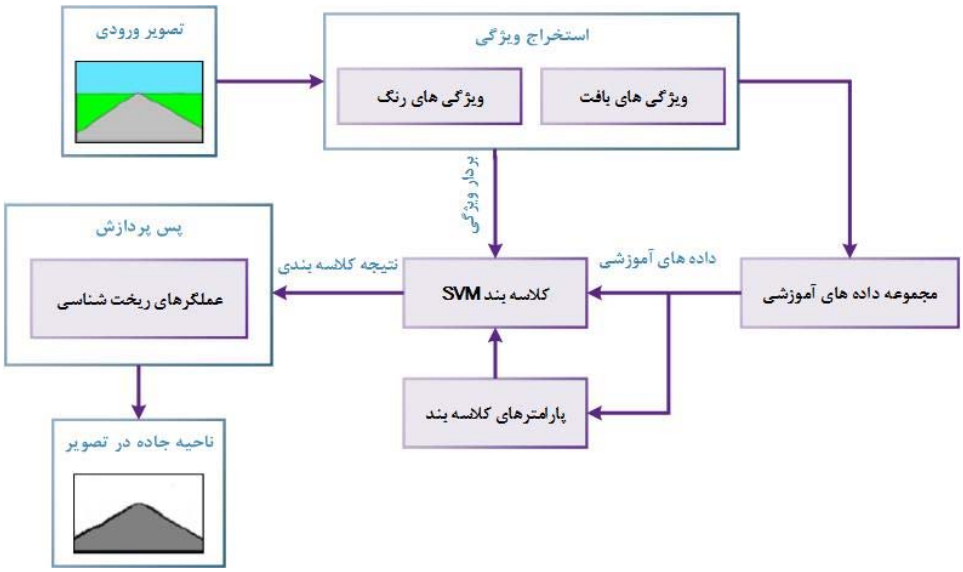
## ۵-۱ الگوریتم پیشنهادی جهت استخراج ویژگی از تصاویر

### ۵-۱-۱ مقدمه

شناسایی جاده در تصاویر برای توسعه پروژه‌های خودروهای هوشمند بسیار با اهمیت است و در سال‌های اخیر متدهای بسیاری در حوزه بینایی ماشین برای حل این مسأله معرفی شده است. الگوریتم‌های شناسایی جاده به سه دسته تقسیم می‌شوند: روش‌های مبتنی بر ویژگی، مبتنی بر مدل و مبتنی بر ناحیه. متدهای مبتنی بر ناحیه به یک مسأله در حوزه یادگیری ماشین تبدیل شده‌اند که به کامپیوترها اجازه می‌دهند بر اساس مجموعه‌ای که برای آموزش در نظر گرفته شده است رفتار خود را تغییر دهند [۷۳].

### ۵-۱-۲ گام‌های الگوریتم پیشنهادی جهت استخراج ویژگی از تصاویر

الگوریتم پیشنهادی در این قسمت از پنج گام تشکیل شده و در شکل ۵-۱ نشان داده شده است:




شکل ۵-۱: فلوچارت الگوریتم استخراج ویژگی از تصویر

**گام ۱:** استخراج یک بردار ویژگی به ازای هر پیکسل از تصویر

**گام ۲:** تولید یک مجموعه داده برای آموزش که با یک مجموعه برچسب‌گذاری شده توسط یک

ناظر تولید شده است

**گام ۳:** تخمین پارامترهای کلاسه‌بند ماشین بردار پشتیبان با روش جستجوی شبکه

**گام ۴:** آموزش کلاسه‌بند ماشین بردار پشتیبان با مجموعه داده آموزشی و استفاده از این مجموعه داده برای کلاسه‌بندی بردارهای ویژگی کل تصویر که به عنوان داده‌های تست هستند و تشخیص پیکسل‌های جاده و غیرجاده

**گام ۵:** استفاده از عملگرهای ریخت‌شناسی[1]

## ۵-۱-۳ شرح الگوریتم تشخیص جاده

### ۵-۱-۳-۱ استخراج ویژگی

در این روش از ویژگی‌های بافت و رنگ در بردار ویژگی استفاده شده است. اطلاعات رنگ به طور مستقیم توسط دوربین به صورت یک تصویر RGB[2] به دست می‌آید. شدت نور در سه مقدار قرمز، سبز و آبی در نمایش اولیه یک تصویر RGB نتایج طبقه‌بندی ضعیفی را ارائه می‌دهند. بنابراین برای کاهش این ضعف از فضای رنگ HSV[3] استفاده کردیم. به عنوان مثال اگر یک تصویر تک رنگ داشته باشیم که در قسمتی از این تصویر سایه وجود دارد، در فضای رنگ RGB ویژگی‌های رنگ قسمت سایه‌دار کاملاً متفاوت با قسمت‌های بدون سایه خواهد بود. اما در فضای رنگ HSV جز فام[4] در هر دو قسمت سایه‌دار و بدون سایه از تصویر تا حدود زیادی یکسان خواهد بود.

جهت استخراج ویژگی‌های بافت تصویر از روش هارالیک [۱]، بررسی شده در فصل دوم که بر مبنای استخراج ویژگی آماری مرتبه دوم است استفاده شده است. در این روش به ازای هر پیکسل یک همسایگی در نظر گرفته و به ازای آن همسایگی ماتریس هم‌رخداد سطح خاکستری را تشکیل می‌دهیم، سپس با استفاده از روابط معرفی شده در فصل دوم سیزده ویژگی را استخراج کرده و یک بردار

[1] Morphological
[2] Red-Green-Blue
[3] Hue-Saturation-Value
[4] Hue

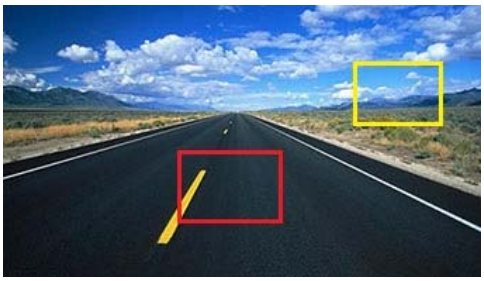
ویژگی شامل این سیزده ویژگی بافت و سه ویژگی رنگ (فام، اشباع[1] و مقدار[2] ) با طول شانزده تولید می‌کنیم.

$$F_{i,j} = \left[f_{t_{1(i,j)}}, \dots, f_{t_{n(i,j)}}, f_{c_{1(i,j)}}, \dots, f_{c_{n(i,j)}}\right] \ , \ i = 1, \dots, H \ , \ j = 1, \dots, w \qquad (۱-۵)$$

در این بردار ویژگی، $f_{t_{n(i,j)}}$، n امین ویژگی آماری استخراج شده برای بافت در نقطه $(i,j)$ و $f_{c_{n(i,j)}}$، n امین ویژگی استخراج شده برای رنگ در نقطه $(i,j)$ در فضای رنگ HSV است. مقادیر $H$ و $W$ به ترتیب طول و عرض تصویر ورودی هستند.

### ۵-۱-۳-۲ مقداردهی اولیه به مجموعه داده آموزشی

داده‌های آموزشی توسط ناظر انسانی برچسب‌گذاری می‌شوند. پنجره‌هایی همانند شکل ۵-۲ توسط ناظر روی تصویر برای انتخاب داده‌های آموزشی در نظر گرفته شده‌اند سپس بردارهای ویژگی پیکسل‌ها در این پنجره‌ها استخراج و برچسب‌گذاری می‌شوند. پنجره قرمز برای مجموعه داده‌های آموزشی مثبت برچسب (+1) و پنجره زرد برای مجموعه داده‌های آموزشی منفی با برچسب (-1) در ماشین بردار پشتیبان خواهند بود.



شکل ۵-۲: پنجره‌های درنظر گرفته شده توسط ناظر جهت داده‌های آموزشی

### ۵-۱-۳-۳ محاسبه پارامترهای کلاسه‌بند ماشین بردار پشتیبان

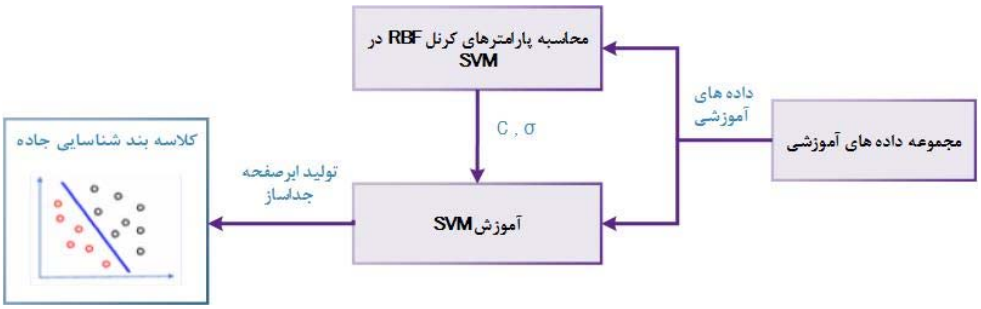
ارتباط بین کلاس‌های جاده و غیرجاده و ویژگی‌های آن‌ها یک ارتباط غیرخطی است لذا برای جداسازی این دو کلاس در کلاسه‌بند SVM از کرنل RBF استفاده گردید. طبق نتایج فصل چهارم، کرنل RBF نیاز به تنظیم دو پارامتر $C$ و $\sigma$ دارد که تنظیم صحیح این دو پارامتر دقت کلاسه‌بندی را

[1] Saturation
[2] Value

افزایش خواهد داد. بنابراین با استفاده از روش 10-Fold Cross Validation، کتابخانه LibSVM جهت آموزش و الگوریتم جستجوی شبکه، این دو پارامتر را تخمین می‌زنیم. در این مرحله بردار ویژگی‌های آموزش و تست در بازه [0,1] نرمال‌سازی شده‌اند.

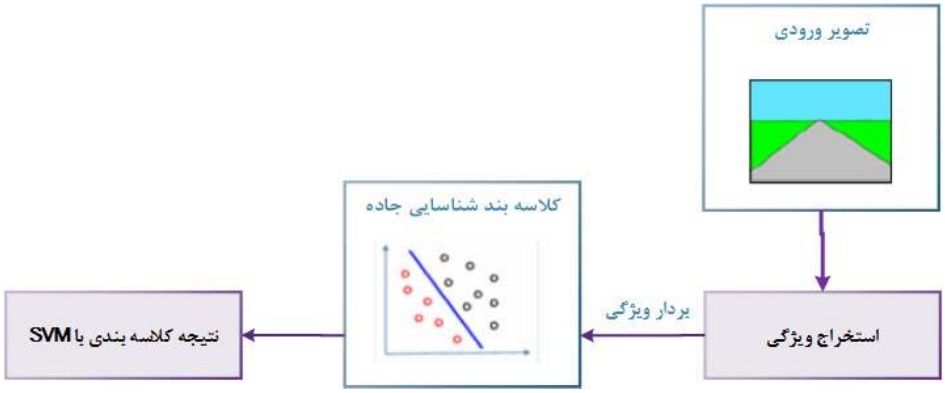
### ۵-۱-۳-۴ آموزش و تست کلاسه‌بند ماشین بردار پشتیبان

نخست مرحله آموزش داده‌های آموزشی با استفاده از SVM مانند (شکل ۵-۳) انجام خواهد شد.




شکل ۵-۳: آموزش داده‌های آموزشی انتخاب شده توسط ناظر با استفاده از کلاسه‌بند SVM

در مرحله کلاسه‌بندی به ازای هر پیکسل از تصویر ورودی یک بردار ویژگی با شانزده ویژگی طبق روش گفته شده در قسمت قبل استخراج کرده پس از نرمال‌سازی این مجموعه بردار ورودی در بازه [0,1]، کلاس هر بردار را با کلاسه‌بند آموزش دیده تعیین می‌کنیم. در این مرحله مشخص می‌گردد هر پیکسل که با کمک یک بردار با شانزده ویژگی نمایش داده شده، متعلق به ناحیه جاده یا غیرجاده است.



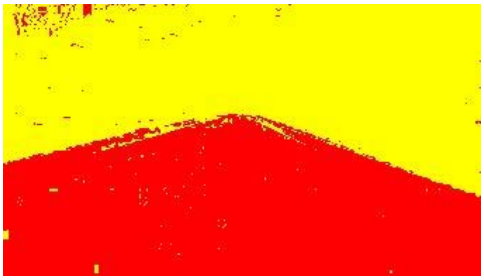

شکل ۵-۴: کلاسه‌بندی بردارهای ویژگی تصویر ورودی با کلاسه‌بند SVM

### ۵-۱-۳-۵ عملگرهای ریخت‌شناسی

عملگرهای ریخت‌شناسی جهت درک فرم تصویر ( نواحی مرزها، چارچوب‌ها، پوسته‌های محدب و ...) به کار می‌روند و نقش بسیار مهمی در کاربردهایی نظیر بینایی ماشین و تشخیص خودکار اشیا دارند [۷۴].

عملگر ریخت‌شناسی جهت یافتن بزرگترین ناحیه جاده متصل، با پوشش حفره‌ها در این ناحیه متصل به کار می‌رود. همان‌طور که در شکل ۵-۵ مشاهده می‌شود حفره‌های کوچکی در قسمت‌های جاده و غیر جاده وجود دارند که با اعمال عملگر ریخت‌شناسی پرسازی[1] بر این تصویر، حفره‌ها برطرف خواهند شد و دو ناحیه متصل جاده و غیرجاده همانند شکل ۵-۶ به دست می‌آید. در نهایت مانند شکل ۵-۷ پیکسل‌های بزرگترین ناحیه شناسایی شده، به عنوان جاده برچسب‌گذاری می‌شوند.

در پایان تصویری خواهیم داشت که پیکسل‌های ناحیه جاده در آن مشخص است. بردارهای ویژگی استخراج شده از پیکسل‌های ناحیه جاده در تصویر نمونه‌ها با کلاس مثبت و سایر پیکسل‌ها نمونه‌ها با کلاس منفی خواهند بود. بنابراین یک مجموعه داده برچسب‌گذاری شده با تعدادی نمونه به اندازه کل پیکسل‌های تصویر با شانزده ویژگی داریم که از آن در مرحله بعد به عنوان مجموعه داده ورودی جهت کاهش ویژگی با استفاده از الگوریتم سیستم ایمنی مصنوعی استفاده می‌کنیم. مراحل کارهای انجام شده در قسمت استخراج ویژگی از تصویر (۱) در شکل ۵-۸ نمایش داده شده است.



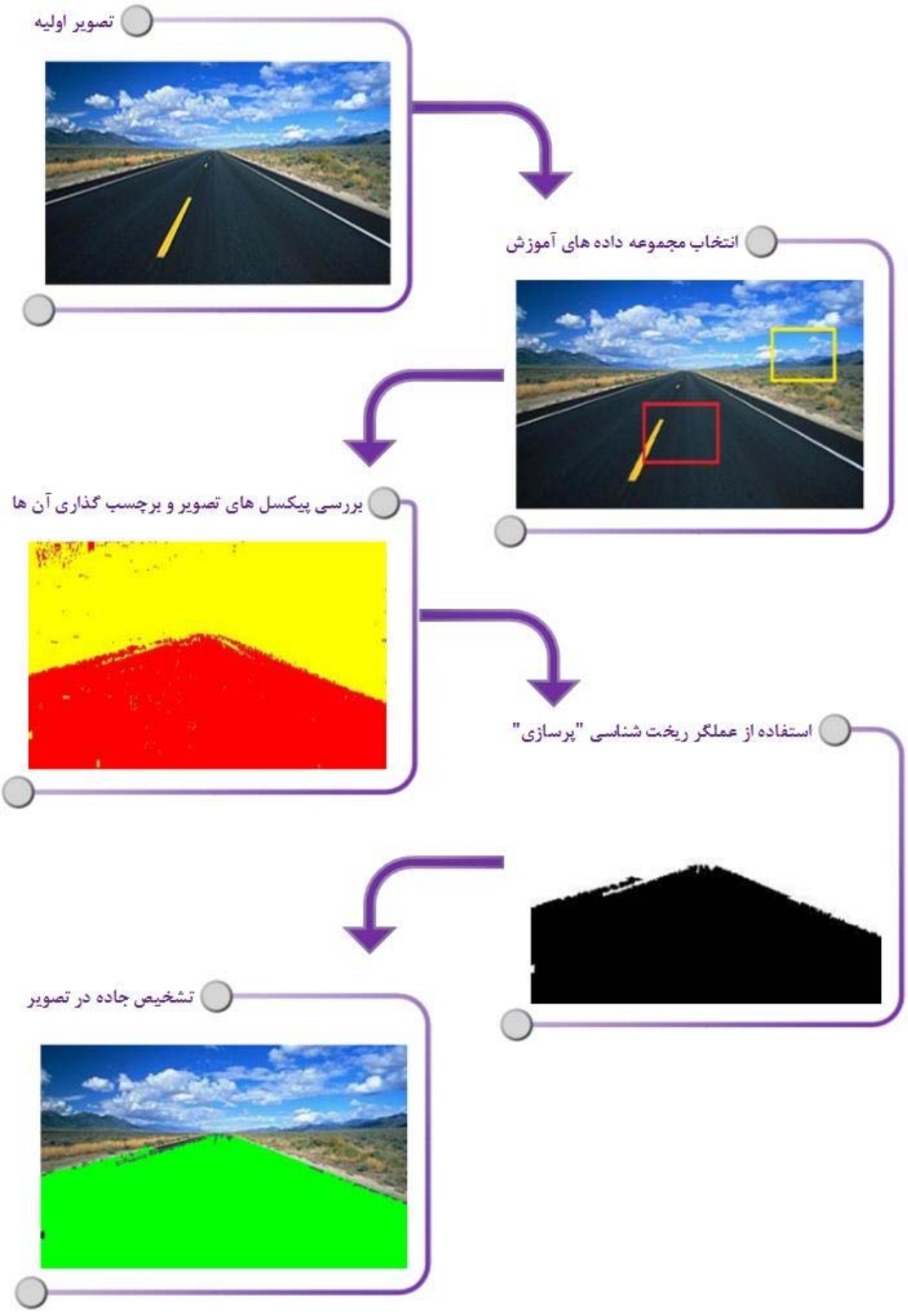
شکل ۵-۵: تصویر به دست آمده پس از کلاسه‌بندی با SVM

[1] Filling

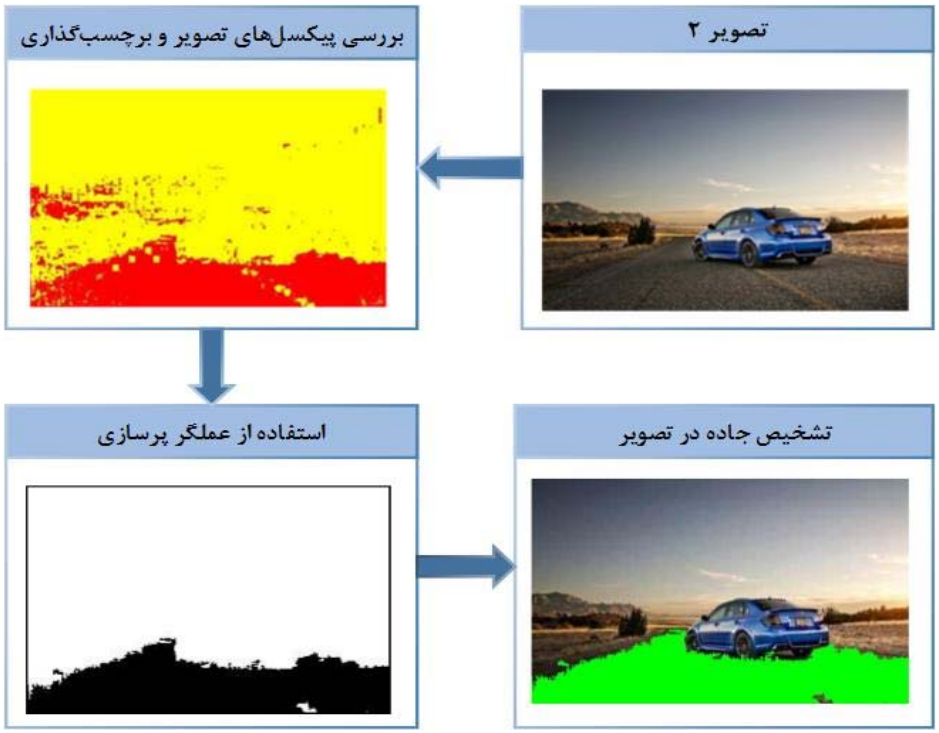
شکل ۵-۶: نتیجه اعمال عملگر ریخت شناسی (پرسازی) بر تصویر نتیجه کلاسه‌بند SVM

شکل ۵-۷: تصویر نهایی با مشخص شدن ناحیه جاده در تصویر

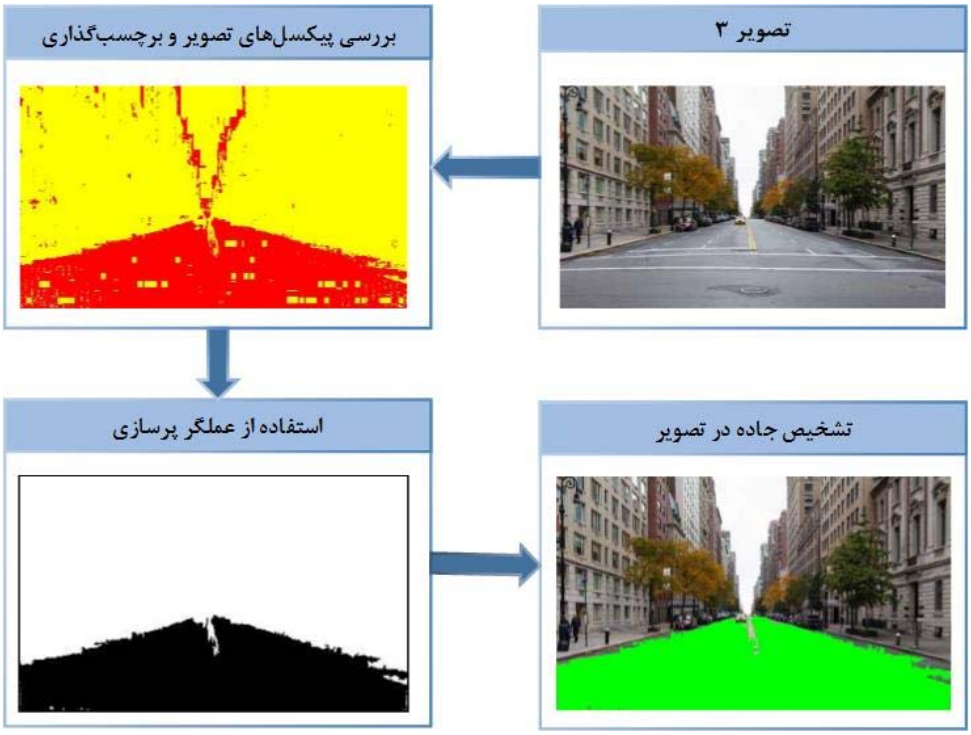
در این بخش ویژگی‌های هشت تصویر دریافت شده توسط موبایل ربات به کمک روش معرفی شده استخراج گردید که در فصل ششم نتایج کاهش ویژگی روی مجموعه داده استخراج شده از هر تصویر ارائه شده است. این تصاویر در( شکل ۵-۹) تا (شکل ۵-۱۵) نمایش داده شده‌اند.



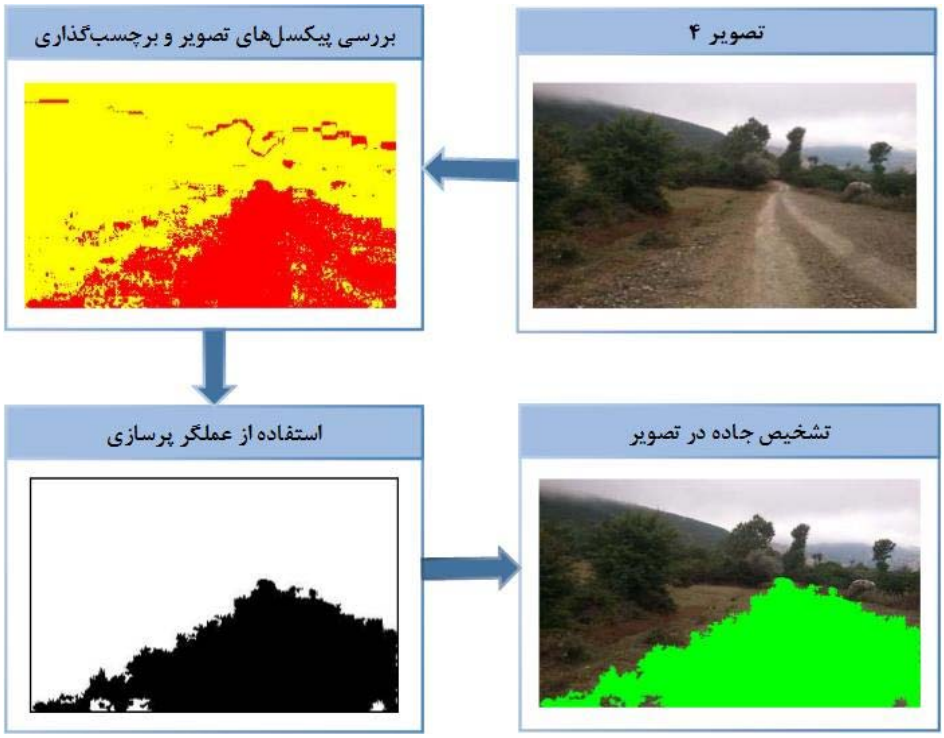

شکل ۵-۸: مراحل استخراج ویژگی از تصویر (۱)



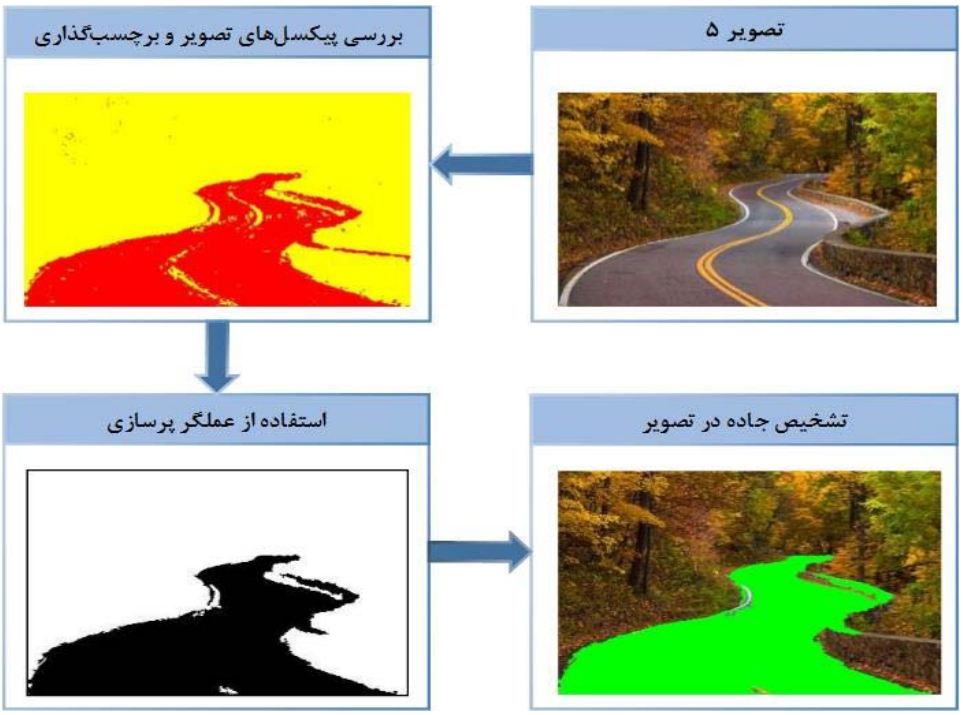

شکل ۵-۹: مراحل استخراج ویژگی از تصویر (۲)



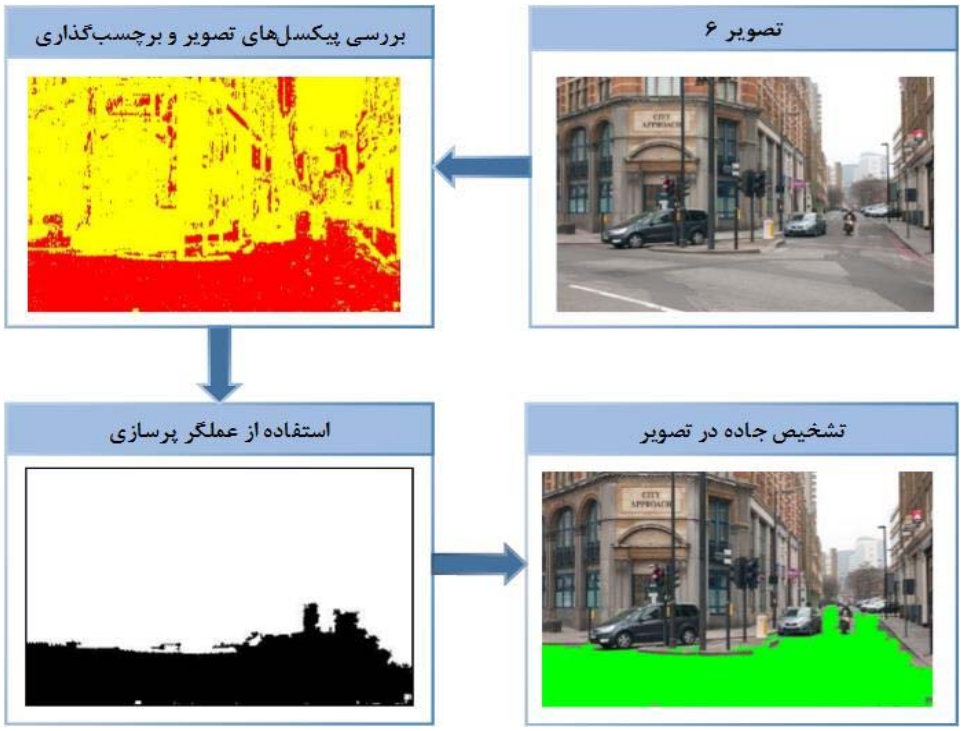

شکل ۵-۱۰: مراحل استخراج ویژگی از تصویر (۳)



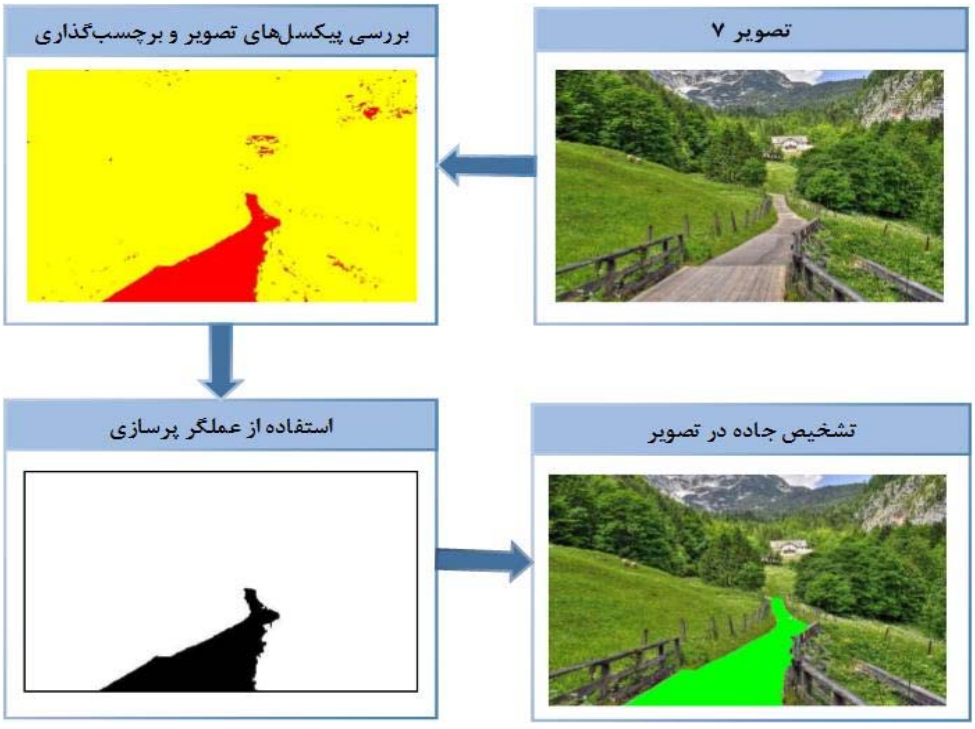

شکل ۵-۱۱: مراحل استخراج ویژگی از تصویر (۴)



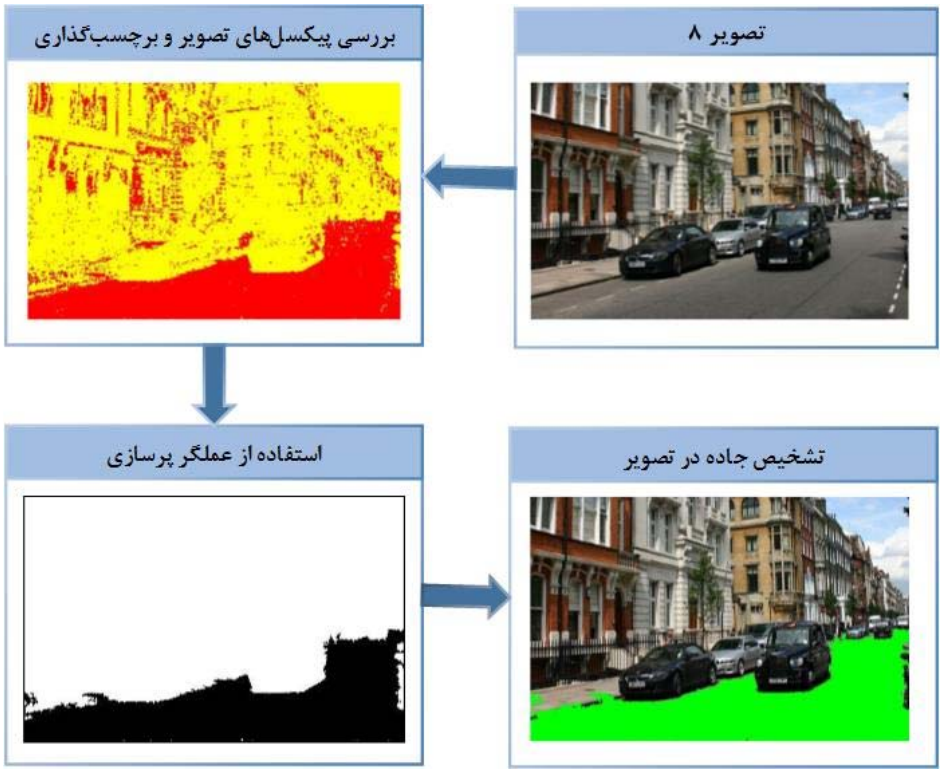

شکل ۵-۱۲: مراحل استخراج ویژگی از تصویر (۵)




شکل ۵-۱۳: مراحل استخراج ویژگی از تصویر (۶)




شکل ۵-۱۴: مراحل استخراج ویژگی از تصویر (۷)




شکل ۵-۱۵: مراحل استخراج ویژگی از تصویر (۸)

## ۵-۲ الگوریتم پیشنهادی جهت انتخاب ویژگی

سیستم‌های فعلی کاهش بعد با مشکلاتی نظیر دقت پایین دسته‌بندی، غیرقابل تفسیر بودن داده‌ها و کم شدن ارزش آن‌ها روبه‌رو هستند. به منظور رفع این مشکلات الگوریتم انتخاب کلونال با استفاده از مهم‌ترین ویژگی خود یعنی بهینه‌سازی به عنوان روش نوین مطرح می‌گردد. لذا در روش پیشنهادی از الگوریتم سیستم ایمنی مصنوعی با بهره‌گیری از الگوریتم انتخاب کلونال در مرحله کلونی‌سازی استفاده می‌گردد.

به دلیل اهمیت سرعت در الگوریتم‌های کاهش ویژگی، در این تحقیق از الگوریتم انتخاب کلونال تطبیقی هم جهت مقایسه استفاده شده است. این روش در کنار کاهش بردار ویژگی، زمان اجرا را نیز به گونه قابل توجهی کاهش می‌دهد. در فصل بعد نتایج حاصل از این روش‌ها مقایسه می‌شوند.

در بسیاری از تحقیقات گذشته از شبکه عصبی مصنوعی به عنوان یک کلاسه‌بند و تابع تخمین استفاده شده است. چون وزن‌دهی اولیه در روش ANN به صورت تصادفی انجام می‌گیرد، این روش

نمی‌تواند یک کمینه سراسری داشته باشد. از سوی دیگر ماشین بردار پشتیبان به عنوان یک مسأله بهینه‌سازی محدب آموزش می‌بیند که می‌تواند یک راه‌حل سراسری را تخمین بزند. بنابراین مشکل ANN قرار گرفتن در کمینه محلی است که در SVM این ضعف وجود ندارد، مشکل Over Fitting نیز نسبت به ANN به صورت موفقیت‌آمیزی در SVM ضعیف شده است. آموزش در SVM نسبت به ANN سریع‌تر انجام می‌گیرد [۷]، با توجه به این دلایل برای پیاده‌سازی روش پیشنهادی از ماشین بردار پشتیبان که یک الگوریتم یادگیری باناظر است استفاده گردیده است.

انواع کرنل‌ها در SVM نتایج متفاوتی بر مجموعه داده‌های مختلف دارند. روابط بین نمونه‌ها از کلاس‌های مختلف در انواع مجموعه داده‌ها، روابطی غیرخطی هستند لذا باید یک کرنل غیرخطی را انتخاب کنیم. نتایج به دست آمده از بررسی ماشین بردار پشتیبان در فصل چهارم نشان می‌دهد کرنل RBF با پارامترهای بهینه نتیجه بسیار مناسبی را ارائه داده است بنابراین در این قسمت از تحقیق نیز از این کرنل استفاده می‌کنیم.

در هنگام تعیین پارامتر پنالتی $C$، اگر مقدار این پارامتر خیلی کوچک انتخاب شود نتیجه حاصل از کلاسه‌بند SVM راضی کننده نخواهد بود و چون تعداد بردارهای پشتیبان متناسب با خطای کلاسه‌بند SVM هستند، تعدادشان افزایش می‌یابد. اگر این پارامتر خیلی بزرگ انتخاب شود، دقت کلاسه‌بند در مرحله آموزش بسیار بالا و در مرحله تست بسیار پایین خواهد بود چراکه بزرگ بودن بیش از حد $C$ جریمه زیادی را برای نقاط جداناپذیر در نظر گرفته بنابراین تعداد بردارهای پشتیبان ذخیره شده افزایش یافته و مشکل Overfitting خواهیم داشت. از سوی دیگر کوچک شدن زیاد پارامتر $C$ مشکل Underfitting را ایجاد می‌کند. در اصل پارامتر $C$ برقرار کننده تعادل بین خطای SVM روی داده‌های آموزش و بیشینه سازی حاشیه است.

پس از انتخاب کرنل RBF، پارامترهای $C$ و $\sigma$ و ویژگی‌های موجود در مجموعه داده استفاده شده به عنوان مقادیر ورودی الگوریتم هستند.

### ۵-۲-۱ نمایش آنتی‌ژن

جمعیت آنتی‌ژن، مجموعه داده متشکل از تعدادی نمونه و ویژگی است. در این قسمت از مجموعه داده‌های موجود در پایگاه داده UCI استفاده شده است در انتخاب این مجموعه‌ها سعی شده، آن‌هایی مورد استفاده قرار گیرند که تنها دارای ویژگی‌های عددی باشند. همچنین از ویژگی‌های استخراج شده از تصاویر موبایل ربات که در قسمت قبل تشریح شد، برای کاهش بردار ویژگی‌ها استفاده می‌شود.

### ۵-۲-۲ نمایش آنتی‌بادی

آنتی‌بادی ورودی‌های مسأله هستند که در نهایت جواب بهینه از بین آن‌ها انتخاب می‌شود. آنتی-بادی‌ها یک ماتریس دودویی با ابعاد $M\times N$ هستند که کلیه ویژگی‌های آنتی‌ژن یا داده‌های آموزشی را ماسک می‌کنند.

هر آنتی‌بادی به صورت یک بردار دودویی درنظر گرفته می‌شود که از سه قسمت تشکیل شده است. مانند شکل ۵-۱۶، بیت‌های قسمت اول مربوط به ویژگی‌ها، بیت‌های قسمت دوم مربوط به پارامتر جریمه $C$ و بیت‌های قسمت سوم مربوط به پارامتر$\sigma$ از کرنل RBF می‌باشند. بیت‌های $ab_1^f \sim ab_N^f$ به عنوان ماسکی برای ویژگی‌های موجود در مجموعه داده ورودی، بیت‌های $ab_1^C \sim ab_{NC}^C$ و $ab_1^\sigma \sim ab_{N\sigma}^\sigma$ به ترتیب برای نمایش پارامترهای $C$ و $\sigma$ هستند. در روش پیشنهادی تعداد بیت‌های ماسک ویژگی‌ها $N_f$ متناسب با تعداد ویژگی‌های مجموعه داده ورودی، و $N_C = N_\sigma = 22$ در نظر گرفت شده‌اند. در نمایش دودویی بیت با مقدار"1" به معنی انتخاب ویژگی متناظر و بیت با مقدار "0" به معنی عدم انتخاب ویژگی متناظر با آن بیت است.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $ab_1^f$ | … | $ab_{Nf}^f$ | $ab_1^c$ | … | $ab_{Nc}^c$ | $ab_1^\sigma$ | … | $ab_{N\sigma}^\sigma$ |

شکل ۵-۱۶: بردار سه قسمتی آنتی‌بادی

بردار آنتی‌بادی به صورت تصادفی مقداردهی می‌شود. به این صورت که یک عدد تصادفی در بازه یک و تعداد ویژگی‌های ورودی مجموعه داده، تولید شده و به اندازه عدد تصادفی تولید شده، بیت‌های

قسمت ماسک بردار آنتی‌بادی به صورت تصادفی یک خواهند شد. با کمک این روش تنوع بیت‌ها با مقادیر یک قسمت ماسک در هر مرحله، بین بازه یک و تعداد ویژگی‌های ورودی است که باعث شده جمعیت اولیه تولید شده تمام فضای جستجو را پوشش دهد. دو قسمت مربوط به پارامترهای $C$ و $\sigma$ به صورت تصادفی به صورت صفر و یک مقدار می‌گیرند. در ادامه باید مقادیر پارامترهای $C$ و $\sigma$ را با کمک رابطه (۵-۲) از فرم دودویی به فرم دهدهی تبدیل شوند.

$$P = min_p + \frac{max_p - min_p}{2^l - 1} \times d \qquad (۵-۲)$$

در رابطه (۵-۲)، $min_p$ و $max_p$ به ترتیب حد بالا و پایین برای پارامترهای $C$ و $\sigma$ هستند که توسط کاربر تعیین می‌شوند، پارامتر $d$ مقدار دهدهی رشته دودویی و $l$ طول بیت‌ها هستند.

### ۵-۲-۳ محاسبه میل ترکیبی آنتی‌بادی و آنتی‌ژن

به میزانی که یک آنتی‌بادی به یک آنتی‌ژن شبیه‌تر باشد، میل ترکیبی بین آن‌ها بیشتر شده و دقت حاصل از ماشین بردار پشتیبان افزایش می‌یابد. لذا برای محاسبه میل ترکیبی میان آنتی‌بادی و آنتی‌ژن از میزان دقت ماشین بردار پشتیبان با کرنل RBF محاسبه شده با کمک کتابخانه LibSVM و روش 10-Fold Cross Validation استفاده شده است.

$$Affinity = SVM\_Accuracy \qquad (۵-۳)$$

### ۵-۲-۴ شرح مراحل الگوریتم

در شکل ۵-۱۷، فلوچارت الگوریتم پیشنهادی نمایش داده شده است که در ادامه به شرح گام-های این الگوریتم می‌پردازیم:

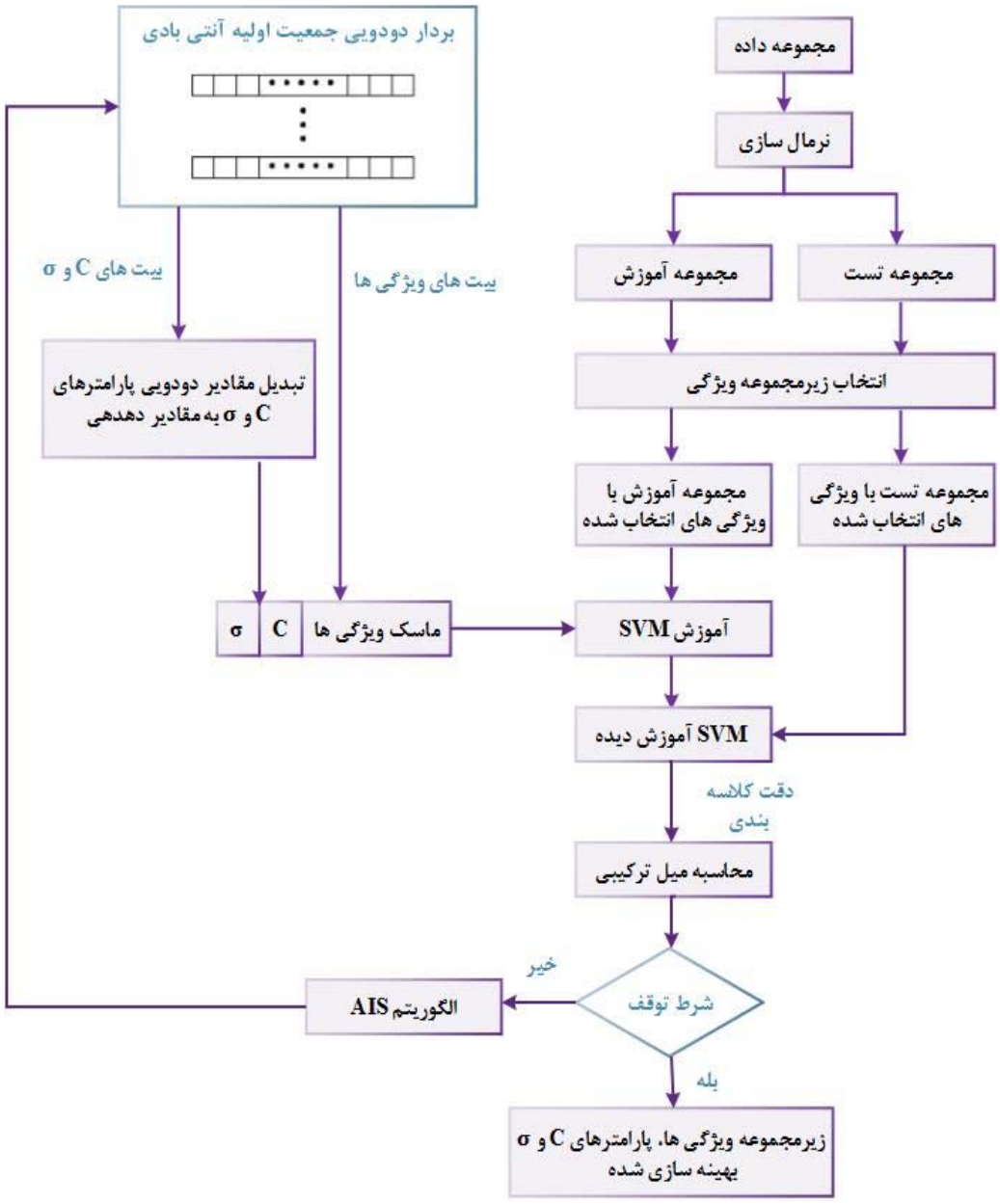
**گام ۱؛ مقیاس‌گذاری**[1]: برای استفاده از مجموعه داده‌ها (آنتی‌ژن‌ها) ابتدا باید آن‌ها را نرمال-سازی کرد. مزیت این روش جلوگیری از غلبه ویژگی‌ها با مقادیر زیاد بر ویژگی‌ها با مقادیر کم است. لذا برای افزایش دقت SVM مجموعه داده‌ها (بردارهای آنتی‌ژن) را با کمک رابطه (۵-۴) نرمال‌سازی

[1] Scaling

می‌کنیم.

$$v' = \frac{v - min_a}{max_a - min_a} \quad (۴-۵)$$

در رابطه (۵-۴)، $v$ مجموعه داده ورودی، $v'$ داده‌های نرمال شده خروجی، $min_a$ و $max_a$ به ترتیب مقادیر کمینه و بیشینه ویژگی‌ها در تمامی نمونه‌ها می‌باشند.




شکل ۵-۱۷: فلوچارت الگوریتم انتخاب ویژگی با استفاده از AIS

**گام ۲؛ تولید جمعیت اولیه آنتی‌بادی‌ها**: در این قسمت یک ماتریس دودویی از آنتی‌بادی‌ها با سایز $M\times N$ تولید می‌شود. اندازه $M$ برابر با اندازه جمعیت اولیه است که در این تحقیق به دلخواه ۱۰۰ در نظر گرفته شده و اندازه $N = N_C + N_\sigma + N_f$ است یعنی جمعیت اولیه آنتی‌بادی‌ها یک ماتریس دودویی $\left(100 \times \left(N_C + N_\sigma + N_f\right)\right)$ است.

**گام۳؛ تبدیل دودویی به دهدهی:** در این مرحله مقادیر دودویی پارامترهای $C$ و $\sigma$ موجود در بردار آنتی‌بادی را با کمک رابطه (۵-۲) به مقادیر دهدهی تبدیل می‌شوند.

**گام۴؛ تشکیل بردار آنتی‌بادی:** پس از محاسبه مقادیر دهدهی برای پارامترهای $C$ و $\sigma$ بردار سه قسمتی نهایی آنتی‌بادی شامل قسمت دودویی ماسک ویژگی‌ها و دو قسمت دهدهی پارامترهای $C$ و $\sigma$ تشکیل می‌گردد.

**گام۵؛ محاسبه میل ترکیبی:** در این قسمت به کمک رابطه (۵-۳) میل ترکیبی میان آنتی-بادی با هر آنتی‌ژن محاسبه می‌شود.

**گام۶؛ کلونی سازی:** در این مرحله به تعداد $n_1$ تا از بهترین آنتی‌بادی‌ها که میل ترکیبی بالاتری با آنتی‌ژن‌ها دارند انتخاب شده و از آن‌ها برای کلونی‌سازی استفاده می‌گردد.

### ۵-۲-۴-۱ انتخاب کلونال

در بسیاری از کارهای انجام شده از جمله [۹]، که مرتبط به مسأله کاهش ویژگی است، هنگام استفاده از الگوریتم انتخاب کلونال، در هر مرحله تعداد کلون‌های تولید شده تعداد ثابتی است اما در این تحقیق تعداد کلون‌ها با کمک رابطه (۵-۵) تعیین می‌شوند. طبق رابطه (۵-۵) تعداد کلون‌ها کاملاً وابسته به میل ترکیبی میان آنتی‌بادی و آنتی‌ژن است. بنابراین آنتی‌بادی با میل ترکیبی بالاتر تعداد کلونی بیشتری را تولید می‌کند.

$$Clone\_Number = round\left(c\_Rate \times \left(Affinity(Ag, CurrentAB)\right)\right) \quad (۵-۵)$$

هدف اولیه در مسأله انتخاب ویژگی، انتخاب ویژگی‌های کارای یک نمونه است، به‌طوری‌که

کارایی دسته‌بندکننده در حد قابل قبولی باقی مانده و در عین حال بر سرعت دسته‌بندی آن افزوده شود. بنابراین در این تحقیق قصد داریم با استفاده از الگوریتم انتخاب کلونال تطبیقی زمان اجرا را در کنار حفظ دقت و کاهش ویژگی مناسب، کاهش دهیم.

### ۵-۲-۴-۲ انتخاب کلونال تطبیقی

تمامی مراحل الگوریتم انتخاب کلونال تطبیقی مانند الگوریتم انتخاب کلونال است با این تفاوت که تعداد کلون‌ها با استفاده از رابطه (۵-۶) تعیین می‌شوند.

$$Clone\_Number(i) = round\left(\frac{\beta \times N_{Sel} \times cc}{i}\right) \quad (۵-۶)$$

در رابطه (۵-۶)، پارامتر $\beta$ عامل ضرب، $cc$ عامل شتاب با مقدار اولیه کمتر از یک ($cc<1$) و مقدار $N_{Sel}$ برابر با $n_1$ می‌باشند.

مقدار فاکتور شتاب $cc$ در هر مرحله از اجرای الگوریتم با کمک رابطه (۵-۷)، به روزرسانی می‌گردد.

$$cc = cc \times \gamma \quad (۵-۷)$$

در رابطه (۵-۷) مقدار $\gamma$ در بازه $[0.5,1.1]$ قرار دارد. با افزایش تعداد تکرارهای الگوریتم اندازه کلون‌ها به دلیل استفاده از این رابطه کاهش یافته، لذا همگرایی الگوریتم سریع‌تر خواهد بود.

**گام۷؛ فراجهش:** پس از کلونی‌سازی آنتی‌بادی‌ها عمل فراجهش روی آن‌ها انجام خواهد شد. به این صورت که هرچه میل ترکیبی بین آنتی‌بادی و آنتی‌ژن کمتر باشد جهش بیشتری در بیت‌های آنتی‌بادی اتفاق می‌افتد و بر عکس. بنابراین عمل اکتشاف[1] به خوبی اتفاق می‌افتد و از واگرایی الگوریتم جلوگیری می‌کند. این عملگر مانند جهش در الگوریتم ژنتیک در صورت برقرار بودن شرطی به ازای یک بیت، مقدار آن بیت را از یک به صفر یا به عکس تغییر خواهد داد. در [۹] فراجهش با نرخ ثابتی اتفاق می‌افتد. از آنجا که فراجهش یک گام بسیار مهم در الگوریتم AIS است و تأثیر به سزایی در همگرایی آن دارد، در این تحقیق به جای انجام فراجهش با نرخ ثابت، از روشی همانند الگوریتم

[1] Exploraition

شکل ۵-۱۸، استفاده می‌کنیم. در این حالت عمل جهشی که روی بیت‌های آنتی‌بادی اتفاق می‌افتد کاملاً متناظر با میل ترکیبی آن با آنتی‌ژن است.

```
for   i=1 : Clone_Number
      MutateRate = 1-Affinity (Ag,Ab)
      for  j=1 : Feature_Number
           d = (-MutateRate- MutateRate) × rand + MutateRate
           MutateAb ( i,j ) = Clone ( i,j ) + d × MutateRate
           if  MutateAb ( i,j ) ≤ 0
                 MutateAb ( i,j ) = 0
           elseif  MutateAb ( i,j ) ≥ 1
                  MutateAb ( i,j ) = 1
          end
      end
end
```

شکل ۵-۱۸: الگوریتم فراجهش در روش کاهش ویژگی با الگوریتم AIS

**گام ۸؛ جایگزینی:** پس از انجام عمل فراجهش برای ایجاد تنوع در آنتی‌بادی‌هایی که به مرحله بعد وارد می‌شوند و افزایش پراکندگی[1] الگوریتم، به تعداد $n_2$ عدد $(n_2 < n_1)$ از بدترین آنتی‌بادی‌ها (آنتی‌بادی‌هایی که میل ترکیبی کمی دارند) را انتخاب کرده و آن‌ها را با آنتی‌بادی‌های تصادفی جایگزین می‌کنیم. این کار موجب می‌شود تنوع در جمعیت اولیه ایجاد شده و تمامی فضای جستجو پوشش داده شود لذا باعث افزایش اکتشاف نیز خواهد شد.

**گام ۹؛ معیار توقف:** الگوریتم پس از رسیدن به تعداد تکرار معین متوقف خواهد شد.

[1] Diversity

# فصل۶ نتایج شبیه‌سازی

در فصل پنجم الگوریتم پیشنهادی جهت استخراج و انتخاب ویژگی معرفی و تشریح شد. در این فصل نتایج حاصل از اعمال این الگوریتم بر روی مجموعه داده‌های مختلف UCI و مجموعه ویژگی-های استخراج شده از تصویر بررسی و مقایسه‌های لازم در قالب جدول‌ها ارائه خواهد شد.

## ۶-۱ مجموعه داده‌های UCI استفاده شده در ارزیابی

برای ارزیابی دقت روش پیشنهادی از مجموعه داده‌های UCI که در مقالات مختلف استفاده شده‌اند و دارای ویژگی‌های عددی هستند استفاده می‌گردد. در جدول ۶-۱، مجموعه داده‌های استفاده شده با مشخصات مربوط به تعداد ویژگی‌ها، تعداد نمونه‌ها و تعداد کلاس‌های آن‌ها معرفی شده‌اند. پیاده‌سازی کلیه الگوریتم‌ها در این تحقیق در محیط نرم‌افزار Matlab 2013 و پیاده‌سازی کلاسه‌بند ماشین بردار پشتیبان با استفاده از کتابخانه LibSVM تحت نرم‌افزار متلب انجام شده‌اند. تمامی الگوریتم‌ها و روش‌ها روی سیستمی با مشخصات Intel Core i7 CPU 2.67 GHz و RAM 4.00 GB اجرا شده‌اند.

جدول ۶-۱: مشخصات مجموعه داده های UCI

| شماره | نام مجموعه داده | تعداد ویژگی‌ها | تعداد نمونه‌ها | تعداد کلاس‌ها |
|---|---|---|---|---|
| ۱ | Australian | ۱۴ | ۶۹۰ | ۲ |
| ۲ | Breast Cancer | ۱۰ | ۶۸۳ | ۲ |
| ۳ | Diabetes | ۸ | ۷۶۸ | ۲ |
| ۴ | Glass | ۱۰ | ۲۱۴ | ۷ |
| ۵ | German | ۲۴ | ۱۰۰۰ | ۲ |
| ۶ | Ionosphere | ۳۲ | ۳۵۱ | ۲ |
| ۷ | Iris | ۴ | ۱۵۰ | ۳ |
| ۸ | Sonar | ۶۰ | ۲۰۸ | ۲ |
| ۹ | Vowel | ۱۳ | ۹۹۰ | ۱۱ |
| ۱۰ | WDBC | ۳۰ | ۵۶۹ | ۲ |
| ۱۱ | Wine | ۱۳ | ۱۷۸ | ۳ |

## ۶-۲ روش k-Fold Cross Validation

برای تضمین معتبر بودن نتایج به دست آمده از مجموعه داده‌ها، در ابتدا داده‌های ورودی به صورت تصادفی به دو مجموعه مستقل آموزش و تست با کمک روش k-Fold Cross Validation تقسیم می‌گردند. در این روش داده‌های ورودی به $k$ دسته تقسیم شده که هر یک از این $k$ قسمت یک بار به صورت مستقل به عنوان مجموعه تست و سایر $k-1$ قسمت باقی مانده به عنوان زیرمجموعه آموزش انتخاب خواهند شد. در این روش بهتر است برای مجموعه داده‌های کوچکتر، جهت افزایش تعداد نمونه‌ها در مرحله آموزش، اندازه $k$ بزرگتر در نظر گرفته شود [۷۵]. اما در این تحقیق به ازای همه مجموعه داده‌های استفاده شده مقدار $k=10$ در نظر گرفته شده است.

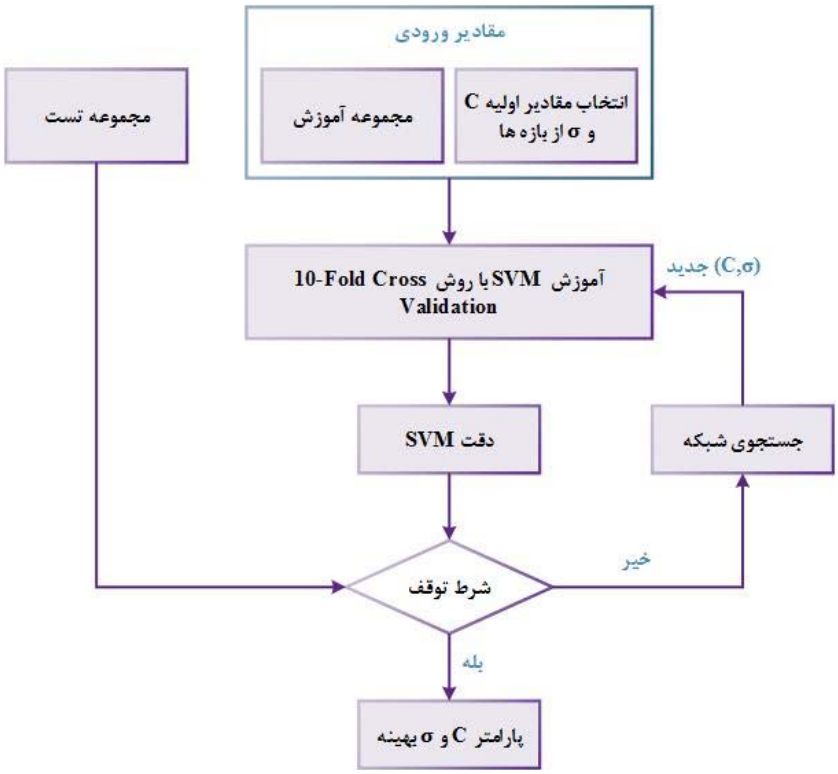
## ۶-۳ الگوریتم جستجوی شبکه

الگوریتم جستجوی شبکه یک روش معمول برای یافتن بهترین نتیجه برای پارامترهای $C$ و $\sigma$ در زمان استفاده از کرنل RBF می‌باشد. شکل ۶-۱، نشان دهنده فرآیند انجام شده در الگوریتم جستجوی شبکه هنگام استفاده از SVM است. در این الگوریتم هر بار یک جفت $(C,\sigma)$ از بازه‌های معرفی شده توسط کاربر، انتخاب می‌شوند و به عنوان پارامترهای کرنل RBF در ماشین بردار پشتیبان قرار می‌گیرند سپس SVM را با داده‌های آموزش به روش 10-Fold Cross Validation آموزش داده و با داده‌های تست، دقت نهایی حاصل از این جفت محاسبه می‌شود. در نهایت دقت‌های به دست آمده از SVM مقایسه شده و جفتی که دقت بالاتری را ارائه دهند (بهترین ناحیه در شبکه)، به عنوان پارامترهای ورودی در مرحله مقداردهی اولیه به بردار آنتی‌بادی، استفاده خواهند شد. اشکال این روش وقت‌گیر بودن آن و نداشتن عملکرد مناسب است [۴, ۷]. همچنین این روش قادر نیست عمل کاهش ویژگی را انجام دهد [۷].

در این تحقیق در دو قسمت، از الگوریتم جستجوی شبکه برای محاسبه پارامترهای $C$ و $\sigma$ در کرنل RBF، استفاده شده است:

۱) در روش پیشنهادی پارامترهای $C$ و $\sigma$ به صورت مقادیر دودویی در نظر گرفته می‌شوند و در هر تکرار الگوریتم AIS آموزش دیده و اصلاح می‌شوند، سپس نتایج روش پیشنهادی با نتایج حاصل از استفاده از الگوریتم جستجوی شبکه برای انتخاب پارامترهای $C$ و $\sigma$، مقایسه خواهد شد.

۲) در مرحله استخراج ویژگی از الگوریتم جستجوی شبکه برای انتخاب پارامترها و آموزش بهتر SVM استفاده شده است.




شکل ۶-۱: فرآیند انتخاب پارامترهای $C$ و $\sigma$ با استفاده از الگوریتم جستجوی شبکه و SVM

در جدول ۶-۲، بازه‌های انتخاب شده برای پارامترهای $C$ و $\sigma$ در الگوریتم جستجوی شبکه نشان داده شده‌اند.

جدول ۶-۲: بازه انتخاب شده برای پارامترهای $C$ و $\sigma$ در الگوریتم جستجوی شبکه

| الگوریتم | نام پارامتر | مقدار |
|---|---|---|
| جستجوی شبکه | $C$ | بازه $\{2^{-2},2^{-1},\ldots,2^{11},2^{12}\}$ |
| | $\sigma$ | بازه $\{2^{-10},2^{-9},\ldots,2^{3},2^{4}\}$ |

## ۶-۳-۱ دقت حاصل از روش جستجوی شبکه و الگوریتم ایمنی مصنوعی بر پایه انتخاب کلونال در تعیین پارامترهای $C$ و $\sigma$

در جدول ۶-۳ دقت، تعداد ویژگی‌های انتخاب شده، پارامترهای $C$ و $\sigma$ مناسب در دو روش الگوریتم ایمنی مصنوعی بر پایه انتخاب کلونال[1] و الگوریتم جستجوی شبکه برای پایگاه داده German به ازای هر Fold نشان داده شده‌اند.

جدول ۶-۳: بررسی مجموعه داده German با دو روش جستجوی شبکه و AISCS+SVM در هر Fold

| AISCS+SVM | | | | Grid Search | | | |
|---|---|---|---|---|---|---|---|
| تعداد ویژگی | $\sigma$ | $C$ | دقت کلی | $\sigma$ | $C$ | دقت کلی | # Fold |
| 7 | 0.4474 | 277.93 | **79** | 0.25 | 2 | **81** | 1 |
| 4 | 0.5625 | 1.19 | **84** | 0.5 | 1 | **79** | 2 |
| 5 | 0.2504 | 50.87 | **83** | 0.25 | 4 | **84** | 3 |
| 24 | 0.072 | 47.3 | **85** | 0.5 | 1 | **75** | 4 |
| 7 | 0.0547 | 108.5 | **82** | 0.5 | 0.5 | **75** | 5 |
| 5 | 0.5 | 100.75 | **83** | 1 | 4094 | **81** | 6 |
| 3 | 0.5 | 1 | **81** | 0.0125 | 16 | **84** | 7 |
| 7 | 0.0073 | 340.29 | **76** | 9.76e-4 | 2048 | **80** | 8 |
| 12 | 0.5237 | 229.07 | **86** | 0.0625 | 4 | **83** | 9 |
| 19 | 0.1987 | 48.61 | **79** | 0.0156 | 1024 | **75** | 10 |

نتایج به دست آمده از دو روش جستجوی شبکه و الگوریتم ایمنی مصنوعی به ازای همه پایگاه داده‌های معرفی شده UCI، در جدول ۶-۴ مشاهده می‌گردد. دقت حاصل از روش AISCS+SVM همراه با کاهش ویژگی نسبت به دقت به دست آمده از طریق الگوریتم جستجوی شبکه بهتر است که نشان دهنده انتخاب مقادیر مناسب برای پارامترهای $C$ و $\sigma$ توسط این الگوریتم می‌باشد. روش AISCS+SVM یک روش یادگیرنده است که در طی تکرارهای متوالی و انجام عمل فراجهش متناسب با دقت به دست آمده، پارامترها را جهش می‌دهد تا به نتیجه مطلوب‌تری برای پارامترهای $C$

[1] Artificial Immune System based on Clonal Selection (AISCS)

و $\sigma$ دست یابد.

جدول ۶-۴: نتایج حاصل از دو روش AISCS+SVM و جستجوی شبکه

| Grid Search | AISCS+SVM | |
|---|---|---|
| میانگین دقت کلی | میانگین دقت کلی | مجموعه داده |
| 89.40 ± 3.8 | **91.59 ± 3.91** | **Australian** |
| 97.51 ± 1.82 | **99.42 ± 1.02** | **Breast Cancer** |
| 79.95 ± 3.91 | **82.95 ± 2.67** | **Diabetes** |
| 81.36 ± 6.09 | **83.14 ± 6.85** | **Glass** |
| 79.7 ± 3.62 | **81.8 ± 3.08** | **German** |
| 97.69 ± 3.69 | **99.15 ± 1.91** | **Ionosphere** |
| 98.33 ± 3.44 | **100** | **Iris** |
| 92.26 ± 5.21 | **99.21 ± 1.37** | **Sonar** |
| 96.79 ± 0.64 | **100** | **Vowel** |
| 98.59 ± 1.10 | **99.65 ± 0.73** | **WDBC** |
| 99.47 ± 1.6 | **100** | **Wine** |

## ۶-۴ انتخاب کلونال با نرخ جهش ثابت

همان‌طور که در فصل قبل گفته شد، برای سادگی استفاده از الگوریتم انتخاب کلونال در برخی از تحقیقات انجام شده مانند [۹] در سال ۲۰۰۹ که در زمینه کاهش ویژگی می‌باشد، نرخ فراجهش و تعداد کلون‌های تولید شده در هر تکرار یک مقدار ثابت در نظر گرفته شده است. در این تحقیق از الگوریتم شکل ۵-۱۸ برای انجام فراجهش و رابطه (۵-۵) برای کلونی سازی براساس میزان میل ترکیبی بین آنتی‌بادی آنتی‌ژن استفاده می‌گردد. در جدول ۶-۵ پارامترهای استفاده شده در این قسمت معرفی شده‌اند.

نتایج حاصل از روش الگوریتم سیستم ایمنی مصنوعی بر پایه انتخاب کلونال (AISCS) با روش الگوریتم سیستم ایمنی مصنوعی بر پایه انتخاب کلونال ثابت[1] (AISCCS+SVM) که در آن فراجهش

[1] Artificial Immune System based on Constant Clonal Selection (AISCCS)

با نرخ ثابت ۰/۰۱ انجام گردد در جدول ۶-۶ نشان داده شد.

جدول ۶-۵: مقادیر پارامترهای دو روش AISCS+SVM و AISCCS+SVM

| الگوریتم | نام پارامتر | مقدار |
|---|---|---|
| انتخاب کلونال | $c\_Rate$ | 10 |
| | نرخ ثابت فراجهش | 0.01 |
| AISCS | اندازه جمعیت | 100 |
| | تعداد تکرار | 20 |
| | $C$ | [1,400] |
| | $\sigma$ | [0.00001,1] |
| | $N_C$ | 22 |
| | $N_\sigma$ | 22 |

با توجه به نتایج جدول ۶-۶، میانگین زمان اجرا در حالتی که از روش AISCCS+SVM استفاده کرده‌ایم به علت انجام فراجهش با نرخ ثابت و عدم نیاز به محاسبه نرخ جهش با استفاده از الگوریتم شکل ۵-۱۸، کمتر است. در تمامی مجموعه داده‌ها تعداد ویژگی‌های کاهش یافته و دقت محاسبه شده در حالتی که از روش AISCS+SVM استفاده شده است نسبت به حالتی که از روش AISCCS+SVM استفاده شده است، مطلوب‌تر می‌باشد به طوری که میانگین درصد کاهش ویژگی در الگوریتم AISCS+SVM، ۶۸/۶۲ درصد با دقت میانگین ۹۴/۲۶ است در حالی‌که در الگوریتم AISCCS، میانگین درصد کاهش ویژگی ۶۵/۴۲ با دقت ۹۴/۰۹ می‌باشد. این نتایج نشان‌دهنده تأثیر فراجهش متناسب با میل ترکیبی در افزایش دقت روش‌های مبتنی بر الگوریتم سیستم ایمنی مصنوعی می‌باشد.

جدول ۶-۶: نتایج حاصل از دو روش AISCS+SVM و AISCCS+SVM

| | AISCS+SVM | | | AISCCS+SVM | | |
|---|---|---|---|---|---|---|
| مجموعه داده | زمان اجرا | میانگین تعداد ویژگی | میانگین دقت کلی | زمان اجرا | میانگین تعداد ویژگی | میانگین دقت کلی |
| **Australian** | 7897.22 | **5.4 ± 2.55** | **91.59 ± 3.91** | **1366.09** | 7.1 ± 2.96 | 91.28 ± 3.62 |
| **Breast Cancer** | 929.64 | **2.2 ± 1.25** | **99.42 ± 1.02** | **1131.75** | 2.3 ± 1.41 | 98.98 ± 1.81 |
| **Diabetes** | 2886.78 | **3.5 ± 1.33** | **82.95 ± 2.67** | **1552.96** | 3.6 ± 1.25 | 82.64 ± 2.12 |
| **Glass** | 656.58 | **4.6 ± 1.35** | **83.14 ± 6.85** | **463.16** | 5.2 ± 2.78 | 83.19 ± 6.82 |
| **German** | 8951.21 | **9.3 ± 6.98** | **81.8 ± 3.08** | **3988.97** | 10.5 ± 2.06 | 81.6 ± 3.13 |
| **Ionosphere** | 915.04 | **5.8 ± 1.4** | **99.15 ± 1.91** | **816.54** | 6.5 ± 4.74 | 99.14 ± 1.38 |
| **Iris** | 193.41 | **1.1 ± 0.32** | **100** | **186.02** | 1.1 ± 0.85 | 100 |
| **Sonar** | 841.58 | **12.8 ± 2.54** | **99.21 ± 1.37** | **544.47** | 13.9 ± 1.56 | 98.61 ± 1.28 |
| **Vowel** | 4042.68 | **7.6 ± 1.4** | **100** | **3961.79** | 7.8 ± 0.79 | 100 |
| **WDBC** | 895.64 | **4.1 ± 2.07** | **99.65 ± 0.73** | **864.60** | 5.4 ± 2.12 | 99.65 ± 0.76 |
| **Wine** | 267.77 | **2.2 ± 0.42** | **100** | **237.34** | 2.2 ± 0.41 | 100 |
| **میانگین** | **2507.45** | | **94.26** | **1376.7** | | **94.09** |

## ۶-۵ روش‌های انتخاب کلونال و انتخاب کلونال تطبیقی

طبق نتایج جدول ۶-۶، میانگین زمان اجرای الگوریتم AISCS+SVM زیاد است جهت کاهش مقدار زمان اجرا می‌توان از الگوریتم سیستم ایمنی مصنوعی بر پایه انتخاب کلونال تطبیقی[1] استفاده کرد. با توجه به فصل قبل، در روش انتخاب کلونال برای کلونی‌سازی در هر مرحله از رابطه (۵-۵) استفاده می‌شود اما در روش انتخاب کلونال تطبیقی معرفی شده از رابطه (۵-۶) استفاده می‌گردد. در این الگوریتم رابطه (۵-۷) تعداد کلون‌های تولید شده در هر تکرار را کاهش داده و موجب افزایش سرعت همگرایی می‌شود بنابراین پاسخ سریع‌تری خواهیم داشت. در جدول ۶-۷ پارامترهای استفاده شده در این قسمت معرفی شده‌اند.

[1] Artificial Immune System based on Adaptive Clonal Selection (AISACS)

جدول ۶-۷: جدول مقادیر اولیه پارامترهای استفاده شده در الگوریتم‌های AISCS، AISACS و انتخاب کلونال تطبیقی

| الگوریتم | نام پارامتر | مقدار |
|---|---|---|
| انتخاب کلونال تطبیقی | $cc$ | 0.8971 |
| | $\beta$ | 1 |
| | $\gamma$ | 0.9 |
| AISCS و AISACS | اندازه جمعیت | 100 |
| | تعداد تکرار | 20 |
| | $C$ | [1,400] |
| | $\sigma$ | [0.00001,1] |
| | $N_C$ | 22 |
| | $N_\sigma$ | 22 |

## ۶-۵-۱ نتایج به دست آمده از روش الگوریتم سیستم ایمنی مصنوعی بر پایه انتخاب کلونال تطبیقی

در جدول ۶-۸ نتایج حاصل از روش AISACS+SVM نمایش داده شده است. در زمان استفاده از الگوریتم ایمنی مصنوعی بر پایه الگوریتم انتخاب کلونال تطبیقی میانگین زمان اجرا نسبت به الگوریتم ایمنی مصنوعی بر پایه الگوریتم انتخاب کلونال تقریباً $\frac{1}{4}$ شده است. میزان میانگین درصد کاهش ویژگی روش AISACS+SVM به ازای تمامی پایگاه داده‌ها ۶۴/۴۶ درصد است که نسبت به روش AISCS+SVM به میزان۴/۱۶ درصد ضعیف‌تر عمل کرده است و میانگین دقت به ۹۴/۱۱ رسیده است که به میزان ۰/۱۵ کاهش داشته است. در مجموع با توجه به موفقیت این روش در کاهش زمان اجرا استفاده از آن مناسب‌تر خواهد بود.

جدول ۶-۸: نتایج حاصل از روش AISACS+SVM

| مجموعه داده | زمان اجرا | میانگین تعداد ویژگی | میانگین دقت کلی |
|---|---|---|---|
| **Australian** | 789.34 | 6.3 ± 3.43 | 90.89 ± 2.69 |
| **Breast Cancer** | 328.01 | 2.2 ± 0.42 | 98.99 ± 0.99 |
| **Diabetes** | 804.81 | 3.6 ± 1.14 | 82.18 ± 3.76 |
| **Glass** | 255.29 | 5.7 ± 2.11 | 83.97 ± 4.94 |
| **German** | 2091.93 | 9.9 ± 4.31 | 81.8 ± 3.60 |
| **Ionosphere** | 286.22 | 7.1 ± 3.9 | 99.18 ± 1.36 |
| **Iris** | 68.65 | 1.2 ± 0.67 | 100 |
| **Sonar** | 231.10 | 14.2 ± 2.82 | 98.61 ± 2.34 |
| **Vowel** | 1572.10 | 7.8 ± 1.22 | 100 |
| **WDBC** | 293.19 | 6 ± 2.87 | 99.65 ± 1.09 |
| **Wine** | 88.16 | 2.8 ± 0.79 | 100 |
| میانگین | **618.98** | | **94.11** |

## ۶-۶ الگوریتم ژنتیک

با توجه به اینکه روش پیشنهادی یک روش الهام گرفته شده از بیولوژی طبیعی، تکاملی و در عین حال یادگیرنده بدن است، نتایج حاصل از این تحقیق (کاهش ویژگی با استفاده از الگوریتم سیستم ایمنی مصنوعی) با الگوریتم ژنتیک که یک الگوریتم تکاملی و الهام گرفته شده از بیولوژی طبیعی است، مقایسه خواهد شد. در جدول ۶-۹ اصطلاحات دو روش AIS و GA در مسأله کاهش ویژگی با یکدیگر مقایسه شده است.

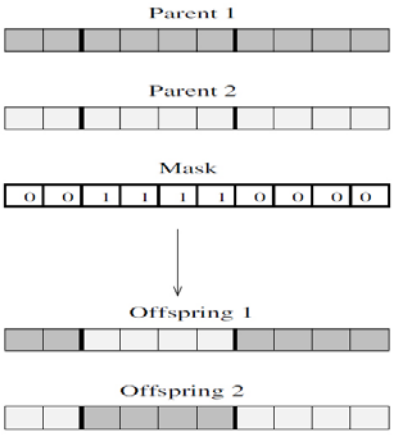
در الگوریتم ژنتیک نحوه تولید جمعیت اولیه کروموزوم‌ها[1] دقیقاً مانند تولید جمعیت اولیه آنتی-بادی‌ها است. یعنی جمعیت اولیه کروموزوم‌ها یک ماتریس دودویی $(100\times(N_C+N_\sigma+N_f))$ است، که دو قسمت مربوط به مقادیر دودویی پارامترهای C و $\sigma$ از فضای ژنوتایپ[2] با کمک رابطه (۵-۲) به

[1] Chromosome
[2] Genotype

حالت فنوتایپ[1] تبدیل شده‌اند. مجموعه داده‌های ورودی نیز به کمک رابطه (۵-۴) نرمال‌سازی می‌شوند سپس از رابطه (۵-۳) به عنوان تابع ارزیابی در الگوریتم ژنتیک استفاده کرده و کیفیت کروموزوم‌ها را ارزیابی می‌کنیم. پس از انتخاب دو بهترین کروموزوم به عنوان والدین[2] با کمک روش چرخ رولت[3]، با استفاده از Two-Point Crossover همانند شکل ۶-۲، فرزندان[4] را تولید کرده سپس از عملگر جهش به صورت تصادفی برای انجام جهش در فرزندان استفاده شده است. در این الگوریتم نخبه‌گرایی[5] نیز در هر تکرار اعمال شده است. در جدول ۶-۱۰ مقادیر پارامترهای استفاده شده در روش GA معرفی شده‌اند.

جدول ۶-۹: مقایسه اصطلاحات دو الگوریتم AIS و GA در مسأله کاهش ویژگی

| AIS | GA | کاربرد در مسأله کاهش ویژگی |
|---|---|---|
| آنتی‌بادی | کروموزوم | بردار شامل ماسک ویژگی‌ها، پارامترهای $C$ و $\sigma$ در کرنل $RBF$ |
| تابع محاسبه میل ترکیبی | تابع ارزیابی | محاسبه کیفیت بردار ویژگی انتخاب شده |
| کلونی‌سازی و فراجهش | Crossover و جهش | افزایش دقت حاصل از کاهش ویژگی و محاسبه پارامترهای $C$ و $\sigma$ جهت افزایش دقت $SVM$ |




شکل ۶-۲: Two-Point Crossover [۴۰]

[1] Phenotype
[2] Parent
[3] Roulette Wheel
[4] Offspring
[5] Elitism

جدول ۶-۱۰: جدول مقادیر اولیه پارامترهای استفاده شده در الگوریتم ژنتیک

| نام پارامتر | مقدار |
|---|---|
| نرخ Crossover (pc) | 0.75 |
| نرخ جهش (pm) | 0.02 |
| اندازه جمعیت | 100 |
| تعداد تکرار | 20 |
| $C$ | [1,400] |
| $\sigma$ | [0.00001,1] |
| $N_C$ | 22 |
| $N_\sigma$ | 22 |

### ۶-۶-۱ نتایج حاصل از الگوریتم ژنتیک

در جدول ۶-۱۱ نتایج به دست آمده با الگوریتم ژنتیک (GA+SVM) و پارامترهای تعیین شده نمایش داده شده است. نتیجه حاصل از این روش در پایگاه داده‌های Glass ، Diabetes و Vowel نسبت به تمامی روش‌های الگوریتم سیستم ایمنی مصنوعی مطلوب‌تر است، اما به صورت میانگین میزان کاهش ویژگی و میانگین دقت به دست آمده در این روش از دو روش AISCS+SVM و AISACS+SVM ضعیف‌تر می‌باشد. دلایل این ضعف در الگوریتم ژنتیک را می‌توان در همگرایی کندتر و احتمال گیر افتادن در کمینه محلی نسبت به روش‌های مبتنی بر الگوریتم سیستم ایمنی مصنوعی دانست. الگوریتم GA به دلیل انجام جهش به صورت تصادفی زمان اجرای سریع‌تری نسبت به روش AISCS+SVM دارد.

## ۶-۷ بررسی نتایج حاصل از مجموعه داده‌های UCI در روش‌های معرفی شده

در این قسمت نتایج حاصل از الگوریتم سیستم ایمنی مصنوعی بر مبنای انتخاب کلونال (AISCS+SVM)، الگوریتم سیستم ایمنی مصنوعی بر مبنای انتخاب کلونال تطبیقی (AISACS+SVM) و الگوریتم ژنتیک (GA+SVM) در مجموعه‌داده‌های انتخاب شده از پایگاه‌داده UCI بررسی شده است.

جدول ۶-۱۱: نتایج حاصل از الگوریتم GA+SVM

| مجموعه داده | زمان اجرا | میانگین تعداد ویژگی | میانگین دقت کلی |
|---|---|---|---|
| **Australian** | 2417.11 | 6.4 ± 2.01 | 90.88 ± 2.77 |
| **Breast Cancer** | 460.24 | 2.4 ± 1.35 | 99.27 ± 0.77 |
| **Diabetes** | 3143.8 | 3.7 ± 1.16 | 83.89 ± 4.61 |
| **Glass** | 288.33 | 3.4 ± 1.17 | 84.08 ± 5.16 |
| **German** | 6286.9 | 10.6 ± 2.12 | 80.1 ± 1.8 |
| **Ionosphere** | 352.06 | 11.3 ± 2.83 | 99.14 ± 1.93 |
| **Iris** | 82.51 | 1.2 ± 0.63 | 100 |
| **Sonar** | 322.57 | 15 ± 2.94 | 98.05 ± 2.00 |
| **Vowel** | 2261.45 | 7.5 ± 2.33 | 99.9 ± 0.35 |
| **WDBC** | 418.10 | 9.9 ± 4.8 | 99.82 ± 0.55 |
| **Wine** | 105.68 | 3.3 ± 1.42 | 100 |
| **میانگین** | **952.8** | | **94.10** |

در جدول ۶-۱۲ نتایج نهایی حاصل از سه روش AISCS+SVM، AISACS+SVM و GA+SVM به صورت خلاصه مشاهده می‌گردد. در این جدول میانگین دقت ویژگی‌های انتخاب شده در طول ۲۰ تکرار با اعمال روش 10-Fold Cross Validation و میانگین ویژگی‌های کاهش یافته به ازای همه مجموعه داده‌ها در این ۲۰ تکرار ارائه شده است.

جدول ۶-۱۲: نتایج نهایی حاصل از سه روش AISCS+SVM، AISACS+SVM و GA+SVM در مجموعه داده-های UCI

| روش | میانگین زمان اجرا | میانگین دقت | میانگین درصد کاهش ویژگی |
|---|---|---|---|
| **AISCS+SVM** | 2507.45 | **94.26** | **68.62** |
| **AISACS+SVM** | **618.98** | 94.11 | 64.76 |
| **GA+SVM** | 952.8 | 94.10 | 63.77 |

نتایج در جدول ۶-۱۲ نشان می‌دهند، روش AISCS+SVM در زمینه کاهش ویژگی در کنار

دستیابی به بالاترین دقت، موفقیت بیشتری کسب کرده است اما زمان اجرای این روش بسیار طـولانی است بنابراین استفاده از روش AISACS+SVM از لحاظ زمان اجرا بهینه‌تر است.

## ۶-۸ مقایسه نتایج حاصل از مجموعه داده‌های UCI در روش‌های معرفـی شـده در این تحقیق و کارهای پیشین

در این قسمت نتایج حاصل از دو روش AISCS+SVM و AISACS+SVM بـا کارهـای مشـابه انجام شده در زمینه کاهش ویژگی که در قسمت مقدمه به اختصار معرفی شده‌اند، مقایسه مـی‌گـردد. روش‌های بررسی شده در مراجـع [۴] (رویکـرد مبتنـی بـر الگـوریتم GA)، [۵] (رویکـرد مبتنـی بـر الگوریتم اجتماع ذرات) و [۹] (رویکرد مبتنی بر الگوریتم انتخاب کلونال) به ترتیـب بـا نـام‌هـای GA Based+SVM، PSO+SVM و CSA+SVM معرفی شده‌اند که در جدول ۶-۱۳ نیز با همین اسـامی مورد مقایسه قرار گرفته‌اند. نتایج جدول ۶-۱۳ برتری روش AISCS+SVM نسبت به سایر روش‌هـای استفاده شده در مراجع مختلف را نشان می‌دهند.

## ۶-۹ نتیجه نهایی حاصل از مجموعه داده‌های UCI

الگوریتم های ایمنی نظیر الگوریتم سیستم ایمنی مصنوعی دقت بالاتری نسبت به الگـوریتم‌هـای تکاملی نظیر الگوریتم ژنتیک دارند. دلایل این برتری را می‌توان در موارد زیر دانست:

- سرعت بالای همگرایی الگوریتم های ایمنی
- خصوصیت مهم بهینه‌سازی و تکامل‌گرا بودن الگـوریتم انتخـاب کلونـال و الگـوریتم انتخـاب کلونال تطبیقی
- وجود کلونی‌سازی متناسب با میزان میل ترکیبی و عملگر فراجهش متناسب بـا عکـس میـل ترکیبی با رویکرد بهبود روند حل مسأله در روش‌های مبتنی بر الگوریتم انتخاب کلونال
- ناگزیر بودن الگوریتم های تکاملی به اجرا در تعداد نسل های بـالا بـرای دسـتیابی بـه پاسـخ مطلوب

جدول ۶-۱۳: نتایج حاصل از روش‌های AISCS+SVM، AISACS+SVM، GA Based+SVM، PSO+SVM و CSA+SVM در مجموعه داده‌های UCI

| مجموعه داده | روش | میانگین تعداد ویژگی | میانگین دقت کلی |
|---|---|---|---|
| **Australian** | **AISCS+SVM** | **5.4** | **91.59** |
| | AISACS+SVM | 6.3 | 90.89 |
| | GA Based+SVM | 3 | 88.1 |
| | PSO+SVM | 9 | 91.03 |
| | CSA+SVM | 6.7 | 90.82 |
| **Breast cancer** | **AISCS+SVM** | **2.2** | **99.42** |
| | AISACS+SVM | 2.2 | 98.99 |
| | GA Based+SVM | 1 | 96.19 |
| | PSO+SVM | 6 | 99.18 |
| | CSA+SVM | 1 | 97.2 |
| **German** | **AISCS+SVM** | **9.3** | **81.8** |
| | AISACS+SVM | 9.9 | 81.8 |
| | GA Based+SVM | 13 | 85.6 |
| | PSO+SVM | 18 | 81.62 |
| | CSA+SVM | 12 | 86.4 |
| **Ionosphere** | **AISCS+SVM** | **5.8** | **99.15** |
| | AISACS+SVM | 7.1 | 99.18 |
| | GA Based+SVM | 6 | 98.56 |
| | PSO+SVM | 21 | 99.01 |
| | CSA+SVM | 6 | 98.56 |
| **Sonar** | **AISCS+SVM** | **12.8** | **99.21** |
| | AISACS+SVM | 14.2 | 98.61 |
| | GA Based+SVM | 15 | 98 |
| | PSO+SVM | 37 | 96.26 |
| | CSA+SVM | 13.7 | 98.8 |

## ۶-۱۰کاهش ویژگی در تصاویر دریافتی از موبایل ربات

زمانی که بردار ویژگی‌ها با ازای تک تک پیکسل‌های تصاویر دریافتی از موبایل ربات استخراج شد در نهایت یک مجموعه داده $(M \times N) \times (17)$ به ازای هر تصویر داریم. در هر مجموعه داده تعداد سطرها (نمونه‌ها) برابر با تعداد پیکسل‌های تصویر$(M \times N)$، تعداد ستون‌های آن شامل سیزده ویژگی استخراج شده بافت به روش هارالیک مبتنی بر GLCM، سه ویژگی رنگ در حوزه HSV و ستون آخر برچسب پیکسل‌های متمایز شده جاده و غیرجاده با کمک روش SVM است. برای تست هر مجموعه داده، $\frac{2}{3}$ داده‌ها به عنوان داده‌های آموزش و $\frac{1}{3}$ به عنوان داده تست در نظر گرفته شده‌اند. تصاویر مورد استفاده در این مرحله، تصاویر دریافتی از موبایل ربات نمایش داده شده در شکل ۵-۸) تا شکل ۵-۱۵) می‌باشند که در جدول ۶-۱۴ با عنوان تصویر (۱) تا تصویر (۸) معرفی شده‌اند.

در جدول ۶-۱۴، تعداد ویژگی‌ها به ازای بیشترین دقت، کمترین تعداد ویژگی‌ها و میانگین تعداد ویژگی‌های انتخاب شده در طی ۲۰ تکرار قابل مقایسه‌اند. از آنجا که همواره میانگین تعداد ویژگی‌ها در تکرارهای یک روش قابل استنادتر است در جدول ۶-۱۵ میانگین کاهش ویژگی نهایی، میانگین دقت به دست آمده و میانگین زمان اجرا به ازای ۸ تصویر ورودی در صورتی که میانگین تعداد ویژگی‌ها در ۲۰ تکرار انتخاب شوند ارائه شده است.

طبق نتایج در جدول ۶-۱۴، روش GA+SVM در زمینه کاهش ویژگی در تصاویر موبایل ربات در تصاویر (۲) و (۸) نتایج مطلوب‌تری نسبت به دو روش AISCS+SVM و AISACS+SVM به دست آورده است اما در دو حالت نتایج ضعیف‌تری ارائه کرده است:

جدول ۶-۱۴: نتایج حاصل از سه روش AISCS+SVM، AISACS+SVM و GA+SVM در تصاویر دریافتی از موبایل ربات

| تصویر | روش | زمان | بیشترین دقت در ۲۰ تکرار | | کمترین تعداد ویژگی در ۲۰ تکرار | | میانگین در ۲۰ تکرار | |
|---|---|---|---|---|---|---|---|---|
| | | | # ویژگی | دقت | # ویژگی | دقت | # ویژگی | دقت |
| ۱ | AISCS | 8835.82 | 10 | 98.62 | 4 | 98.39 | 7.7±2.54 | 98.34±0.20 |
| | AISACS | 3106.10 | 8 | 98.74 | 7 | 98.38 | 8.8±1.74 | 98.34±0.21 |
| | GA | 4299.01 | 13 | 99.07 | 2 | 96.82 | 8.5±4.03 | 98.78±0.12 |
| ۲ | AISCS | 13187.25 | 10 | 95.75 | 5 | 95.52 | 8.3±0.78 | 94.92±0.69 |
| | AISACS | 5239.59 | 6 | 95.67 | 6 | 95.67 | 8.5±0.39 | 94.12±0.36 |
| | GA | 9419.84 | 4 | 96.48 | 2 | 96.25 | 7.7±1.64 | 96.10±0.41 |
| ۳ | AISCS | 11206.59 | 6 | 85.36 | 2 | 85.16 | 3.8±0.54 | 85.24±0.26 |
| | AISACS | 5840.59 | 5 | 85.55 | 2 | 85.02 | 3.9±2.15 | 85.13±0.16 |
| | GA | 17660.56 | 7 | 86.64 | 3 | 86.30 | 6.8±3.12 | 85.77±0.39 |
| ۴ | AISCS | 14844.37 | 2 | 79.18 | 1 | 72.37 | 1.05±0.22 | 72.33±1.62 |
| | AISACS | 5541.46 | 1 | 72.06 | 1 | 72.06 | 1±0 | 71.51±0.42 |
| | GA | 276.91 | 15 | 72.28 | 4 | 36.96 | 7.15±2.83 | 49.85±16.23 |
| ۵ | AISCS | 26900.61 | 7 | 82.96 | 3 | 81.79 | 7.33±2.56 | 82.24±0.39 |
| | AISACS | 10325.03 | 8 | 82.71 | 4 | 82.11 | 8.15±1.98 | 81.90±0.43 |
| | GA | 19832.95 | 12 | 83.68 | 3 | 82.77 | 7.9±3.81 | 83.09±0.52 |
| ۶ | AISCS | 10471.40 | 3 | 74.06 | 1 | 70.99 | 3.35±0.74 | 69.14±3.60 |
| | AISACS | 4941.61 | 4 | 73.19 | 1 | 71.09 | 4.1±0.30 | 70.45±1.39 |
| | GA | 422.92 | 6 | 72.63 | 3 | 58.98 | 8±3.56 | 61.27±4.89 |
| ۷ | AISCS | 35603.05 | 10 | 82.75 | 1 | 82.44 | 5.05±3.54 | 82.28±0.19 |
| | AISACS | 23027.03 | 9 | 82.83 | 2 | 81.96 | 6.25±2.65 | 82.11±0.40 |
| | GA | 61527.07 | 6 | 83.26 | 3 | 82.95 | 7.6±2.79 | 82.83±0.21 |
| ۸ | AISCS | 23996.42 | 10 | 79.92 | 4 | 74.27 | 8.35±1.81 | 77.27±1.22 |
| | AISACS | 15129.62 | 13 | 80.75 | 4 | 75.85 | 8.55±2.52 | 77.19±1.85 |
| | GA | 26250.01 | 5 | 81.27 | 2 | 80.47 | 6.4±2.7 | 80.38±0.52 |

۱) برای تصاویر (۴) و (۶) زمان اجرای روش GA+SVM نسبت به دو روش دیگر به صورت چشم‌گیری کمتر است اما میانگین تعداد ویژگی‌های انتخاب شده زیاد و دقت به دست آمده از آن بسیار کم است، این نتیجه به دلیل افتادن در کمینه محلی به دست آمده است.

۲) در تصویر (۷) زمان اجرای GA+SVM تقریباً دو برابر زمان اجرای AISCS+SVM و سه برابر زمان اجرای AISACS+SVM است در حالی‌که نتایج خروجی آن نسبتاً با دو روش دیگر یکسان است، که نشان دهنده همگرایی کند این روش می‌باشد.

با توجه به نتایج جدول ۶-۱۵ زمان اجرای روش‌های AISCS+SVM و GA+SVM به ترتیب ۱/۹۸ و ۱/۹۱ برابر زمان اجرای روش AISACS+SVM می‌باشند. بنابراین از لحاظ زمان اجرا الگوریتم سیستم ایمنی مصنوعی مبتنی بر انتخاب کلونال تطبیقی بهینه‌تر است.

جدول ۶-۱۵: میانگین نتایج سه روش AISCS+SVM، AISACS+SVM و GA+SVM در تصاویر دریافتی از موبایل ربات

| روش | میانگین زمان اجرا | میانگین دقت | میانگین درصد کاهش ویژگی |
|---|---|---|---|
| **AISCS+SVM** | 18130.69 | **83.97** | **64.90** |
| **AISACS+SVM** | **9143.77** | 82.69 | 61.68 |
| **GA+SVM** | 17470.16 | 79.75 | 53.08 |

در نهایت در مسأله کاهش ویژگی تصاویر موبایل ربات نیز روش AISCS+SVM در زمینه کاهش ویژگی در کنار دستیابی به بالاترین دقت، موفقیت بیشتری کسب کرده است اما با درنظر گرفتن زمان اجرا و دقت به دست آمده از روش AISACS+SVM، این روش نتایج مناسب‌تری نسبت به دو روش دیگر دارد.

# فصل ۷ نتیجه‌گیری و پیشنهادات

# فصل ۸

در این تحقیق انتخاب ویژگی با کمک الگوریتم سیستم ایمنی مصنوعی و ماشین بردار پشتیبان جهت محاسبه میل ترکیبی میان آنتی‌بادی و آنتی‌ژن و بهینه‌سازی همزمان پارامترهای SVM و کرنل RBF بررسی شده است. همچنین برای استخراج ویژگی از تصاویر موبایل ربات، از روش استخراج ویژگی بافت هارالیک مبتنی بر ماتریس هم‌رخداد سطح خاکستری که یک روش آماری مرتبه دو است و ویژگی‌های رنگ تصویر در کانال HSV به کمک روش SVM استفاده شده است. در این روش طول بردار ویژگی به ازای هر پیکسل تصویر ۱۶ است.

نتایج حاصل از الگوریتم سیستم ایمنی مصنوعی مبتنی بر انتخاب کلونال با نرخ جهش ثابت 0.01 (AISCCS)، الگوریتم سیستم ایمنی مصنوعی مبتنی بر انتخاب کلونال با نرخ جهش متناسب با میل ترکیبی آنتی‌بادی و آنتی‌ژن (AISCS) و الگوریتم سیستم ایمنی مصنوعی مبتنی بر انتخاب کلونال تطبیقی (AISACS) بررسی شده است و نتایج حاصل از این سه روش با الگوریتم ژنتیک (GA) مقایسه گردید. طبق نتایج به دست آمده در فصل ۶ الگوریتم‌های پیشنهاد داده شده بر اساس الگوریتم سیستم ایمنی مصنوعی به واسطه خصوصیت مهم بهینه‌سازی و تکامل‌گرا بودن الگوریتم انتخاب کلونال، کارایی بهتر و دقت قابل قبولی را نسبت به الگوریتم ژنتیک در مسأله کاهش ویژگی ارائه داده‌اند. الگوریتم AISCS دقت قابل قبول‌تر و الگوریتم AISACS زمان اجرا کوتاه‌تری را نسبت به سایر روش‌های معرفی شده دارند.

در این تحقیق می‌توان با اعمال تغییراتی در مراحل الگوریتم انتخاب کلونال کلاسیک، سرعت اجرا و کارایی را افزایش بیشتری داد و یا به جای الگوریتم سیستم ایمنی مصنوعی از سایر الگوریتم-های بهینه‌سازی و تکاملی بهره برد. نکته حائز اهمیت انتخاب تابع ارزیابی است، چراکه امر مهم در رسیدن الگوریتم‌های بهینه‌سازی به بهترین جواب، انتخاب تابع ارزیابی مناسب است. برای تابع ارزیابی ماشین بردار پشتیبان با کرنل RBF می توان از انواع دیگر کرنل‌ها بهره برد و پارامترهای آن‌ها را نیز بهینه کرد چون تأثیر پارامترهای کرنل‌های مختلف در دقت ماشین بردار پشتیبان بسیار با اهمیت است. همچنین به جای SVM می‌توان از رگرسیون ماشین بردار استفاده کرد.

در مبحث استخراج ویژگی هم می‌توان روش‌هایی نظیر تبدیل موجک و یا روش الگـوی دودویـی محلی[1] برای استخراج بردار ویژگی از تصویر بهره برد.

[1] Local Binary Pattern (LBP)

# فصل۹ فهرست مراجع

[۱] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE,* vol. 67, pp. 786-804, 1979.

[۲] L. Jack and A. Nandi, "Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms," *Mechanical systems and signal processing,* vol. 16, pp. 373-390, 2002.

[۳] H. Fröhlich, O. Chapelle, and B. Schölkopf, "Feature Selection for Support Vector Machines by Means of Genetic Algorithms," Master's thesis, University of Marburg, 2002. http://www-ra/informatik. unituebingen. de/mitarb/froehlich, 2002.

[۴] C.-L. Huang and C.-J. Wang, "A GA-based feature selection and parameters optimizationfor support vector machines," *Expert Systems with applications,* vol. 31, pp. 231-240, 2006.

[۵] S.-W. Lin, K.-C. Ying, S.-C. Chen, and Z.-J. Lee, "Particle swarm optimization for parameter determination and feature selection of support vector machines," *Expert Systems with Applications,* vol. 35, pp. 1817-1824, 2008.

[۶] C.-L. Huang and J.-F. Dun, "A distributed PSO–SVM hybrid system with feature selection and parameter optimization," *Applied Soft Computing,* vol. 8, pp. 1381-1391, 2008.

[۷] E. Avci, "Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm–support vector machines: HGASVM," *Expert Systems with Applications,* vol. 36, pp. 1391-1402, 2009.

[۸] عادلی مسبب، احسان و فتحی محمود ، ۱۳۸۶، مدلی مبتنی بر الگوریتم ژنتیک برای انتخاب/استخراج ویژگی و بهینه‌سازی پارامترهای SVM، *اولین کنگره مشترک سیستم های فازی و سیستم های هوشمند*، مشهد، دانشگاه فردوسی مشهد

[۹] S. Ding and S. Li, "Clonal selection algorithm for feature selection and parameters optimization of support vector machines," in *Knowledge Acquisition and Modeling, 2009. KAM'09. Second International Symposium on*, 2009, pp. 17-20.

[۱۰] S. Shojaie and M. Moradi, "An evolutionary artificial immune system for feature selection and parameters optimization of support vector machines for ERP assessment in a P300-based GKT," in *Biomedical Engineering Conference, 2008. CIBEC 2008. Cairo International*, 2008, pp. 1-5.

[۱۱] تقی زاده، گلاره و افتخاری مقدم امیر مسعود، ۱۳۸۸، کاهش ابعاد داده با رویکرد الگوریتم CLONALG مبتنی بر سیستم های ایمنی مصنوعی، *پانزدهمین کنفرانس بین المللی سالانه انجمن کامپیوتر ایران.*

[۱۲] I. K. Fodor, "A survey of dimension reduction techniques," ed: Technical Report UCRL-ID-148494, Lawrence Livermore National Laboratory, 2002.

[۱۳] G. Dudek, "Tournament searching method to feature selection problem," in *Artifical Intelligence and Soft Computing*, ed: Springer, 2010, pp. 437-444.

[۱۴] A. Meyer-Baese and V. J. Schmid, *Pattern Recognition and Signal Analysis in Medical Imaging*: Elsevier, 2014.

[۱۵] S. E. Umbaugh, *Computer imaging: digital image analysis and processing*: CRC

Press, 2005.

[۱۶] J. R. Smith and S.-F. Chang, "Automated binary texture feature sets for image retrieval," in *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, 1996, pp. 2239-2242.

[۱۷] J. Sklansky, "Image segmentation and feature extraction," *Systems, Man and Cybernetics, IEEE Transactions on,* vol. 8, pp. 237-247, 1978.

[۱۸] J. K. Hawkins, "Textural properties for pattern recognition," *Picture processing and psychopictorics,* pp. 347-370, 1970.

[۱۹] M. Tuceryan and A. K. Jain, "Texture analysis," *The handbook of pattern recognition and computer vision,* vol. 2, pp. 207-248, 1998.

[۲۰] T. A. Pham, "Optimization of Texture Feature Extraction Algorithm," MSc Thesis, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2010.

[۲۱] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on,* pp. 610-621, 1973.

[۲۲] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 19, pp. 153-158, 1997.

[۲۳] W. Siedlecki and J. Sklansky, "On automatic feature selection," *International Journal of Pattern Recognition and Artificial Intelligence,* vol. 2, pp. 197-220, 1988.

[۲۴] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial intelligence,* vol. 97, pp. 245-271, 1997.

[۲۵] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial intelligence,* vol. 97, pp. 273-324, 1997.

[۲۶] M. Dash and H. Liu, "Feature selection for classification," *Intelligent data analysis,* vol. 1, pp. 131-156, 1997.

[۲۷] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *ICML*, 1997, pp. 412-420.

[۲۸] D. L. Swets and J. J. Weng, "Efficient content-based image retrieval using automatic feature selection," in *Computer Vision, 1995. Proceedings., International Symposium on*, 1995, pp. 85-90.

[۲۹] L. N. De Castro and F. J. Von Zuben, "The clonal selection algorithm with engineering applications," in *Proceedings of GECCO*, 2000, pp. 36-39.

[۳۰] W. Lee, S. J. Stolfo, and K. W. Mok, "Adaptive intrusion detection: A data mining approach," *Artificial Intelligence Review,* vol. 14, pp. 533-567, 2000.

[۳۱] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 17, pp. 491-502, 2005.

[۳۲] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artificial intelligence,* vol. 151, pp. 155-176, 2003.
[۳۳] M. Dash and P. W. Koot, "Feature selection for clustering," in *Encyclopedia of database systems*, ed: Springer, 2009, pp. 1119-1125.
[۳۴] S. Das, "Filters, wrappers and a boosting-based hybrid for feature selection," in *ICML*, 2001, pp. 74-81.
[۳۵] K.-J. Wang, K.-H. Chen, and A. M. Adrian, "An improved artificial immune recognition system with the opposite sign test for feature selection," *Knowledge-Based Systems,* 2014.
[۳۶] E. Leopold and J. Kindermann, "Text categorization with support vector machines. How to represent texts in input space?," *Machine Learning,* vol. 46, pp. 423-444, 2002.
[۳۷] K. Ng and H. Liu, "Customer retention via data mining," *Artificial Intelligence Review,* vol. 14, pp. 569-590, 2000.
[۳۸] E. P. Xing, M. I. Jordan, and R. M. Karp, "Feature selection for high-dimensional genomic microarray data," in *ICML*, 2001, pp. 601-608.
[۳۹] D. Dasgupta and F. Nino, *Immunological computation: theory and applications*: CRC Press, 2008.
[۴۰] A. P. Engelbrecht, *Computational intelligence: an introduction*: John Wiley & Sons, 2007.
[۴۱] L. N. De Castro and F. J. Von Zuben, "Artificial immune systems: Part I–basic theory and applications," *Universidade Estadual de Campinas, Dezembro de, Tech. Rep,* vol. 210, 1999.
[۴۲] A. K. Abbas, A. H. Lichtman, and S. Pillai, *Cellular and molecular immunology*: Elsevier Health Sciences, 1994.
[۴۳] P. J. Delves, S. J. Martin, D. R. Burton, and I. M. Roitt, *Roitt's essential immunology* vol. 20: John Wiley & Sons, 2011.
[۴۴] L. Martin, "What is the function of the human appendix? Did it once have a purpose that has since been lost," *Sci Am,* 1999.
[۴۵] L. N. De Castro and J. Timmis, *Artificial immune systems: a new computational intelligence approach*: Springer, 2002.
[۴۶] D. Dasgupta, *Artificial immune systems and their applications* vol. 1: Springer, 1999.
[۴۷] J. Timmis, M. Amos, W. Banzhaf, and A. Tyrrell, "" Going back to our roots": second generation biocomputing," *arXiv preprint cs/0512071,* 2005.
[۴۸] S. Forrest, A. S. Perelson, L. Allen, and R. Cherukuri, "Self-nonself discrimination in a computer," in *2012 IEEE Symposium on Security and Privacy*, 1994, pp. 202-202.
[۴۹] J. E. Hunt and D. E. Cooke, "An adaptive, distributed learning system based on the immune system," in *Systems, Man and Cybernetics, 1995. Intelligent Systems for*

*the 21st Century., IEEE International Conference on*, 1995, pp. 2494-2499.

[۵۰] U. Aickelin and S. Cayzer, "The danger theory and its application to artificial immune systems," *arXiv preprint arXiv:0801.3549,* 2008.

[۵۱] A. S. Perelson and G. F. Oster, "Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination," *Journal of theoretical biology,* vol. 81, pp. 645-670, 1979.

[۵۲] F. González, D. Dasgupta, and J. Gómez, "The effect of binary matching rules in negative selection," in *Genetic and Evolutionary Computation—GECCO 2003*, 2003, pp. 195-206.

[۵۳] J. K. Percus, O. E. Percus, and A. S. Perelson, "Predicting the size of the T-cell receptor and antibody combining region from consideration of efficient self-nonself discrimination," *Proceedings of the National Academy of Sciences,* vol. 90, pp. 1691-1695, 1993.

[۵۴] F. A. González and D. Dasgupta, "Anomaly detection using real-valued negative selection," *Genetic Programming and Evolvable Machines,* vol. 4, pp. 383-403, 2003.

[۵۵] T. Stibor, J. Timmis, and C. Eckert, "A comparative study of real-valued negative selection to statistical anomaly detection techniques," in *Artificial Immune Systems*, ed: Springer, 2005, pp. 262-275.

[۵۶] T. Stibor, J. Timmis, and C. Eckert, "On the use of hyperspheres in artificial immune systems as antibody recognition regions," in *Artificial Immune Systems*, ed: Springer, 2006, pp. 215-228.

[۵۷] L. N. de Castro and J. Timmis, "Artificial immune systems: a novel approach to pattern recognition," 2002.

[۵۸] G. Dudek, "An artificial immune system for classification with local feature selection," *Evolutionary Computation, IEEE Transactions on,* vol. 16, pp. 847-860, 2012.

[۵۹] L. N. De Castro and F. J. Von Zuben, "Learning and optimization using the clonal selection principle," *Evolutionary Computation, IEEE Transactions on,* vol. 6, pp. 239-251, 2002.

[۶۰] E. Hart and J. Timmis, "Application areas of AIS: The past, the present and the future," *Applied soft computing,* vol. 8, pp. 191-201, 2008.

[۶۱] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, pp. 273-297, 1995.

[۶۲] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery,* vol. 2, pp. 121-167, 1998.

[۶۳] Y. Bazi and F. Melgani, "Toward an optimal SVM classification system for hyperspectral remote sensing images," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 44, pp. 3374-3385, 2006.

[۶۴] D. Meyer and F. T. Wien, "Support vector machines," *The Interface to libsvm in package e1071, January,* vol. 10, 2014.

[۶۵] R. Fletcher, *Practical methods of optimization*: John Wiley & Sons, 2013.

[۶۶] K. Hornik, D. Meyer, and A. Karatzoglou, "Support vector machines in R," *Journal of statistical software,* vol. 15, pp. 1-28, 2006.

[۶۷] I. Aydin, M. Karakose, and E. Akin, "A multi-objective artificial immune algorithm for parameter optimization in support vector machine," *Applied Soft Computing,* vol. 11, pp. 120-129, 2011.

[۶۸] M. Pal, "Multiclass approaches for support vector machine based land cover classification," *arXiv preprint arXiv:0802.2411,* 2008.

[۶۹] S. Knerr, L. Personnaz, and G. Dreyfus, "Single-layer learning revisited: a stepwise procedure for building and training a neural network," in *Neurocomputing*, ed: Springer, 1990, pp. 41-50.

[۷۰] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *Neural Networks, IEEE Transactions on,* vol. 13, pp. 415-425, 2002.

[۷۱] J. C. Platt, N. Cristianini, and J. Shawe-Taylor, "Large Margin DAGs for Multiclass Classification," in *nips*, 1999, pp. 547-553.

[۷۲] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST),* vol. 2, p. 27, 2011.

[۷۳] S. Zhou, J. Gong, G. Xiong, H. Chen, and K. Iagnemma, "Road detection using support vector machine based on online learning and evaluation," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*, 2010, pp. 256-261.

[۷۴] R. C. Gonzalez and R. E. Woods, "Digital image processing," ed: Prentice hall Upper Saddle River, NJ:, 2002.

[۷۵] S. L. Salzberg, "On comparing classifiers: Pitfalls to avoid and a recommended approach," *Data mining and knowledge discovery,* vol. 1, pp. 317-328, 1997.

## Abstract

The feature reduction increases the speed of processing and reduces the storage capabilities. Therefore, there has been a lot of attention in feature selection and feature extraction methods. In recent years Artificial Immune System algorithm (AIS) based on clonal selection with its optimization and evolutionary properties highly regarded. By choosing appropriate method for calculating affinity and perform hypermutation and generating clones proportional to result of antibody and antigen affinity, the AIS algorithm has been achieved to acceptable results efficiently.

In this thesis a feature vector consist of color properties in HSV channel and texture features calculated by Haralick method, extraced from each pixel of a picture that taken by mobile robot. For determining the road and non-road pixel the Support Vector Machine (SVM) method has been applied. AIS algorithm has the potential to generate both the optimal feature subset and SVM parameters at the same time. Our research objective is to optimize the parameters and feature subset simultaneously, without degrading the SVM classification accuracy. Clonal selection in AIS algorithm have high time complexity thus to reduce it, the adaptive clonal selection algorithm is used. Several public datasets of UCI repository are employed to calculate the classification accuracy rate. The experimental results conducted in this study with AIS algorithm based on clonal selection algorithm and SVM as a function for affinity calculation, show that the average feature reduction rate is 68.62% with 94.26 average accuracy rate and for dataset extracted from mobile robot image, average feature reduction rate is 64.90% with 83.97 average accuracy rate.

**Keywords:** Adaptive Clonal Selection Algorithm, Artificial Immune System Algorithm, Clonal Selection Algorithm, Feature Reduction, Gray Level Co-occurrence Matrix, Support Vector Machine, Texture Extraction.

**Shahrood University of Technology**

**Department of Electrical and Robotic Engineering**

# Feature Reduction of Image by Mobile Robot Using Artificial Immune System Algorithm

**Maryam Sadat Hashemipour**

**Supervisor:**

**Dr. Seyyed Ali Soleimani**

**January 2014**