بهبود کارایی معیار طبقه‌بند چیرنف به‌کمک الگوریتم تابع پرکننده

جواد حمیدزاده* **و منا مرادی**

دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه سجاد، مشهد، ایران

چکیده

تحلیل تفکیک‌کننده خطی یکی از روش‌های پرکاربرد در حوزه کاهش ابعاد فضای ویژگی و طبقه‌بندی داده‌ها به‌وسیله بیشینه‌سازی نسبت پراکندگی بین طبقه‌ها به پراکندگی درون طبقه‌ها است. این روش مبتنی بر معیار فیشر بوده و از تحلیل واریانس برای بیان تفکیک‌پذیری طبقه‌ها استفاده می‌کند. مهم‌ترین محدودیت این معیار در مواجهه با داده‌های ناهمگن است. برای رفع این محدودیت، استفاده از فواصل توزیعی نظیر معیار چیرنف پیشنهاد شده است. معیار چیرنف با درنظرگرفتن فاصله چیرنف میان دو توزیع داده، قادر به اندازه‌گیری فواصل میان توابع چگالی احتمال و استخراج ویژگی‌هایی با بیش‌ترین قابلیت تفکیک‌کنندگی است؛ اما ایراد این روش آن است که چنانچه دو توزیع طبقه داده‌های ناهمگن از یکدیگر فاصله کمی داشته باشند، موجب هم‌پوشانی طبقه‌ها در فضای نگاشت شده و باعث افزایش خطای طبقه‌بندی می‌شود. این مقاله، با معرفی روش انتخاب نمونه با نام حاشیه بیشینه‌ای به شناسایی نمونه‌های مرزی و غیرمرزی پرداخته و با بهره‌گیری از نمونه‌های مرزی، ماتریس پراکندگی مطلوبی برای افزایش کارایی تحلیل تفکیک‌کننده خطی ایجاد می‌کند. در روش پیشنهادی، فرایند انتخاب نمونه همانند یک مسأله بهینه‌سازی مقید دودویی در نظر گرفته شده و جواب‌های مسأله با استفاده از تابع پرکننده به‌دست می‌آیند. عملکرد روش پیشنهادی بر روی داده‌های برگرفته‌شده از پایگاه داده UCI به‌وسیلهٔ روش اعتبارسنجی ضرب‌دری ده‌تایی ارزیابی و با طبقه بندهای سنتی و مرز دانش مقایسه شده است. آزمایش‌ها نشان‌دهنده برتری روش پیشنهادی از نظر صحت طبقه‌بندی و زمان محاسبه است.

واژگان کلیدی: طبقه‌بندی داده‌ها، معیار چیرنف، حاشیه حداکثری، تابع پرکننده.

Improving Chernoff criterion for classification by using the filled function

Javad Hamidzadeh* & Mona Moradi

Faculty of computer engineering and information technology,
Sajjad University, Mashhad, Iran

Abstract

Linear discriminant analysis is a well-known matrix-based dimensionality reduction method. It is a supervised feature extraction method used in two-class classification problems. However, it is incapable of dealing with data in which classes have unequal covariance matrices. Taking this issue, the Chernoff distance is an appropriate criterion to measure distances between distributions. In the proposed method, for data classification, LDA is used to extract most discriminative features but instead of its Fisher criterion, the Chernoff distance is employed to preserve the discriminatory information for the several classes with heteroscedastic data. However, the Chernoff distance cannot handle the situations where the component means of distributions are close and leads to the component distribution overlap and underperforming classification. To overcome this issue, the proposed method designs an instance selection method that provides the appropriate covariance matrices. Aiming to improve LDA-based feature selection, the proposed method includes two phases: (1) it removes non-border instances and keeps border ones by introducing a maximum margin sampling method. The basic idea of this phase is

* نویسندهٔ عهده‌دار مکاتبات

* Corresponding author

• تاریخ ارسال مقاله: ۱۳۹۹/۳/۱۰ • تاریخ پذیرش: ۱۴۰۰/۹/۲۰ • تاریخ انتشار: ۱۴۰۱/۱۰/۳ • نوع مطالعه: پژوهشی

[DOI: 10.52547/jsdp.19.3.105]

based on keeping the hyperplane that separates a two-class data and provides large margin separation. In this way, the most representative instances are selected. (2) It extracts features on selected instances by the proposed extension of LDA which generates a desirable scatter matrix to increase the efficiency of LDA. In the proposed method, the instance selection process is considered a constrained binary optimization problem with two contradicting objects, and the problem solutions are obtained by using a heuristic method named filled function. This optimization method does not easily get stuck in local minima; meanwhile, it is not affected by improper initial points. The performance of the proposed method on data collected from the UCI database is evaluated by 10-fold validation. The results of experiments are compared to several competing methods, which show the superiority of the proposed method in terms of classification accuracy percentage and computational time.

Keywords: Chernoff criterion; Data classification; Instance selection; Filled function, Maximum margin.

۱- مقدمه

طبقه‌بندی یکی از مباحث مهم داده‌کاوی است. الگوریتم‌های طبقه‌بندی با تحلیل مجموعه‌داده و شناسایی الگوی پنهان در آن‌ها، سعی در تعیین برچسب نمونه‌های دیده‌نشده دارند. در کاربردهای واقعی، مجموعه‌داده‌ها اغلب دارای تعداد زیادی نمونه با ابعاد (ویژگی‌های) بسیار بالا بوده که علاوه‌بر کاهش کیفیت توصیف داده‌ها منجر به افزایش بار محاسباتی، زمان محاسبه و فضای ذخیره‌سازی می‌شوند [1,2]. در این شرایط اعمال روش‌های کاهش تعداد نمونه و تعداد ابعاد در مرحله پیش‌پردازش، می‌توانند منجر به افزایش کارایی طبقه‌بندها شوند [3].

روش‌های کاهش نمونه برای انتخاب زیرمجموعه مناسب از مجموعه‌داده اصلی به سه شیوه مختلف عمل می‌کنند: (۱) روش‌های مبتنی بر مرز تصمیم؛ این روش‌ها تنها نمونه‌های نزدیک به مرز طبقه‌بندی را حفظ می‌کند. (۲) روش‌های مبتنی بر همسایگی؛ این روش‌ها، نمونه‌های ناهمخوان با همسایه‌ها را به‌عنوان نوفه در نظر گرفته و از مجموعه‌داده حذف می‌کنند. (۳) روش‌های ترکیبی؛ این روش‌ها به‌گونه‌ای کوچک‌ترین زیرمجموعه داده‌ها را انتخاب می‌کنند که دقت طبقه‌بند برای داده‌های آزمایشی حفظ و یا بیشتر شود. روش‌های ترکیبی اصولاً سعی می‌کنند نقاط حاشیه‌ای را با استفاده از معیار ترکیبی از دو روش یادشده حفظ کنند.

در یک دسته‌بندی کلی، روش‌های کاهش ابعاد به دو دسته خطی و غیرخطی تقسیم می‌شوند. در سال‌های اخیر، روش‌های خطی نقش مهمی در حلی مسائلی که با تعداد زیادی داده مواجه‌اند ایفا کرده است. به‌عنوان‌مثال، در مسئله شناسایی مؤثرترین ویژگی‌های یک طبقه و یا در مسئله طراحی طبقه‌بندهایی که قادر به برچسب‌گذاری بردارهای ویژگی جدید باشند [4-7]. سه دلیل عمده برای محبوبیت روش‌های خطی وجود دارد: (۱) چنانچه توزیع طبقه‌ها به‌صورت خطی تفکیک‌پذیر باشند، به‌طورمعمول تعداد اندکی از نمونه‌ها برای انجام وظیفه[1] موردنظر کفایت می‌کنند. (۲) برای بیشتر مسائل، می‌توان رابطه تجزیه مقادیر ویژه تعریف کرد. در این حوزه، الگوریتم‌های کارآمدی ارائه شده‌اند [8,9]. (۳) با بهره‌گیری از کرنل‌های غیرخطی، می‌توان مسائل مرتبط با داده‌های پیچیده و غیرخطی را مدیریت کرد. البته روش‌های خطی با مشکلاتی روبه‌رو هستند؛ به‌عنوان مثال (۱) در روش‌های کاهش بُعد خطی، با تغییر ساختار داده طی برداری‌کردن، موقعیت محلی داده‌ها نسبت به یکدیگر تغییر کرده و پیوستگی و روابط میان آن‌ها از بین می‌رود. (۲) این روش‌ها دارای پیچیدگی محاسباتی زیادی هستند. برای مثال در روش‌های خطی نظیر تحلیل تفکیک‌کننده خطی[2] (LDA) [10] و تحلیل مؤلفه‌های اصلی[3] (PCA)، نیاز به حل مسأله ماتریس پراکندگی مربعی به ابعادی برابر با تعداد ویژگی‌ها داریم. (۳) در این روش‌ها با مشکل نحوست ابعاد مواجهیم، بدین معنی که با افزایش تعداد ویژگی‌های داده، تعداد نمونه‌های لازم برای داشتن یک روش یادگیری مطمئن افزایش یافته و در صورت عدم دسترسی به تعداد داده کافی، خطای طبقه‌بندی رخ می‌دهد. برای مثال در LDA، ناکافی‌بودن داده‌ها منجر به بدحالتی ماتریس پراکندگی و بروز خطا در طبقه‌بندی می‌شود.

در دسته‌بندی دیگر، روش‌های کاهش ابعاد به دو گروه کلی انتخاب ویژگی و استخراج ویژگی دسته‌بندی می‌شوند [11]. روش‌های انتخاب ویژگی با انتخاب زیرمجموعه‌ای از ویژگی‌ها ابعاد داده را کاهش می‌دهند؛ اما روش‌های استخراج ویژگی با هدف تولید ویژگی‌های بامعناتر، چند ویژگی را به‌گونه‌ای با یکدیگر ترکیب می‌کند که دارای تمام یا بخش زیادی از اطلاعات موجود در ویژگی‌های اولیه باشند. روش LDA از پرکاربردترین روش‌های کاهش ابعاد خطی و استخراج ویژگی‌های

[1] Task
[2] Linear Discriminant Analysis (LDA)
[3] Principal Component Analysis (PCA)

[DOI: 10.52547/jsdp.19.3.105]

تفکیک‌کننده برای داده‌های همگن است که ضمن کاهش ابعاد به روش استخراج ویژگی، سعی در تفکیک بهتر طبقه‌ها از یکدیگر دارد. این روش با فرض برقراری شرایطی نظیر

- وجود ماتریس‌های پراکندگی برابر
- وجود توزیع نرمال طبقه‌های داده و
- همگن‌بودن داده‌ها

به‌صورت کارا عمل کرده اما داده‌های دنیای واقعی به‌طورمعمول از چنین شرایطی پیروی نمی‌کنند. استفاده از فواصل توزیعی همچون معیار چیرنف از روش‌های غلبه بر الزام شرط همگن‌بودن داده‌ها است. هدف معیار چیرنف، یافتن ترکیب خطی است که بتواند فاصله چیرنف میان دو توزیع را بیشینه کند. فاصله چیرنف از ماتریس کوواریانس طبقه‌ها برای کسب اطلاعات تفکیک‌کنندگی استفاده می‌کند. عیب معیار چیرنف آن است که احتمال وجود همه طبقه‌ها را یکسان در نظر می‌گیرد. در این صورت، اگر توزیع دو طبقه از یکدیگر فاصله کمی داشته باشند، موجب هم‌پوشانی طبقه‌ها در فضای نگاشت شده و نرخ طبقه‌بندی کاهش می‌یابد.

ایده اصلی این مقاله، بهبود شاخص تفکیک‌پذیری روش LDA است. یعنی به‌دنبال تولید ویژگی‌هایی هستیم که با ایجاد بیشترین تفکیک میان طبقه‌ها، نقشی تعیین‌کننده در نتایج طبقه‌بندی داشته باشند. بدین منظور، ماتریس پراکندگی LDA با استفاده معیار چیرنف پیشنهادی، بهبود می‌یابد. داده‌های ورودی به این معیار، نمونه‌های مرزی شناسایی‌شده به‌وسیلهٔ روش پیشنهادی بیشینه حاشیه است. ازآن‌جاکه در پیاده‌سازی روش پیشنهادی تنها از نمونه‌های مرزی استفاده می‌شود، بنابراین کاهش نمونه، کاهش فضای جستجو و منابع مورد نیاز از مزایای آن است. نتایج حاصل از آزمایش‌ها نشان می‌دهند که روش پیشنهادی در مقایسه با روش‌های رقیب از نرخ صحت طبقه‌بندی و زمان اجرای مناسبی برخوردار است.

نوآوری‌های مقاله در دو حوزه انتخاب نمونه و استخراج ویژگی شامل موارد زیر است:

- ارائه مدلی برای شناسایی نمونه‌های مرزی.
- بهره‌گیری از نمونه‌های مرزی شناسایی‌شده به‌منظور بهبود کارایی طبقه‌بند چیرنف

در این مقاله از روش انتخاب نمونه جدیدی به‌عنوان حاشیه بیشینه‌ای و متقارن مبتنی بر قواعد نزدیک‌ترین همسایه استفاده شده است. در این روش، نمونه‌هایی که بر روی مرز تصمیم‌گیری تأثیری ندارند، نادیده گرفته شده و نمونه‌های مرزی با هدف ایجاد ماتریس پراکندگی تحلیل تفکیک‌کننده خطی حفظ می‌شوند. در روش پیشنهادی، فرایند انتخاب نمونه همانند یک مسئله بهینه‌سازی مقید دودویی در نظر گرفته شده و جواب‌های آن با استفاده از تابع پرکننده به‌نحوی به‌دست می‌آید که کاهش هم‌پوشانی بین طبقه‌ها و افزایش نرخ طبقه‌بندی را به دنبال دارد.

ساختار مقاله به شرح زیر است: بخش دوم به بررسی مطالعات مرتبط می‌پردازد. روش پیشنهادی در بخش سوم معرفی شده و نتایج و تفسیر آن‌ها در بخش چهارم و نتیجه‌گیری و کارهای آینده در بخش پنجم بیان شده است.

۲- مطالعات مرتبط

روش‌های کاهش بُعد خطی به‌دلیل پیچیدگی زمانی خطی در کاربردهای تشخیص الگو از اهمیت بالایی برخوردارند. در [12] آنالیز تفکیک‌کننده چیرنف به‌عنوان روش کاهش ابعاد خطی در فضای نگاشت ارائه شده است. هدف این معیار، بیشینه‌سازی جداپذیری توزیع‌های پراکندگی طبقه‌ها در فضای نگاشت داده‌شده است که با اندازه‌گیری فاصله بین دو طبقه در فضای نگاشت داده‌شده با ابعاد کم تحقّق می‌یابد. در [13] روش کاهش بعد خطی مبتنی بر گرادیان کاهشی ارائه شد که هدف آن بیشینه‌کردن فاصله چیرنف در فضای انتقال یافته است. این روش منجر به افزایش تفکیک‌پذیری طبقه‌ها در این فضا می‌شود. در [14] از معیار چیرنف برای اندازه‌گیری فاصله میان توزیع‌های اصلی و وزن‌دار استفاده شد. در [15] یک روش کاهش بعد خطی ناهمگن مبتنی بر مقدار ویژه و معیار چیرنف ارائه شد. در [16] یک روش کاهش بعد کرنلی مبتنی بر فضای گرادیان توابع کانونیک ارائه شد. در [17] روش کاهش بُعد تفکیک‌پذیر بدون ناظر با استفاده از کرنل ارائه شد. در این روش، با بهره‌گیری از یادگیری گراف تطبیقی و یادگیری ویژگی، به‌طور هم‌زمان ساخت گراف بهینه و کاهش بعد انجام شد. در [18] روش کاهش بعد چند کرنلی مبتنی بر رگرسیون خطی برای طبقه‌بندی تصاویر ارائه شد. هدف روش یادشده عبارت است از یادگیری خودکار کرنل بهینه از چند کرنل پایه و تبدیل نگاشت به‌گونه‌ای که در زیرفضای با ابعاد کم نگاشت‌یافته، خطای بازسازی بین طبقه‌ای افزایش و خطای بازسازی

[DOI: 10.52547/jsdp.19.3.105]

درون طبقه‌ای کاهش یابد. در این روش، فشردگی درون طبقه‌ای افزایش می‌یابد.

انتخاب ویژگی مبتنی بر یادگیری معیار فاصله رویکردی دیگر است که تاکنون مطالعات زیادی را به خود اختصاص داده است [22-19]. هدف این رویکرد، یافتن معیار فاصله، برای کاربردهایی نظیر خوشه‌بندی و یا طبقه‌بندهای مبتنی بر فاصله نظیر kNN، به‌گونه‌ای است که بتوان به‌نحو مطلوبی ارتباط میان ویژگی‌ها را توصیف و محدودیت‌های میان نمونه‌ها (شباهت و تفاوت) را اقناع کرد. بر خلاف روش‌های انتخاب ویژگی که برای هر ویژگی وزن خاصی در نظر می‌گیرند، بیشتر روش‌های یادگیری معیار فاصله، اهمیت یک‌سانی برای تمام ویژگی‌ها قائل هستند. جهت حل این مسأله، در [23] یک معیار فاصله ترکیبی که دربرگیرنده مشخصه‌هایی نظیر وزن‌دهی به ویژگی‌های بااهمیت و ارتباطاتشان بود ارائه شد. [24] روش‌هایی را برای استخراج ویژگی‌های عمیق به‌وسیلهٔ شبکه‌های عصبی مبتنی بر soft-max ارائه کرد. در [25] یک روش طبقه‌بندی جدید با استفاده از سطح تصمیم مبتنی بر فاصله با رویکرد تصویرسازی نزدیک‌ترین همسایگی به نام DDC ارائه شد. این روش دارای ویژگی‌های نظیر عدم نیاز به روند آموزش مرسوم (مانند kNN)، عدم نیاز به زمان جستجوی بالا به‌منظور مکان‌یابی kNN و عدم نیاز به فرایند بهینه‌سازی (برخلاف روش‌های طبقه‌بندی مانند ماشین بردار پشتیبان) است. در [26] یک روش تفکیک‌کننده غیرخطی با به‌کارگیری کرنل ترکیبی مدل‌های محلی با درون‌یابی ارائه شده است.

LDA [27]، یک روش کاهش ابعاد با نظارت است که با فرض نرمال بودن طبقه‌ها، نگاشت خطی که معیار فیشر[1] را بیشینه کند به‌دست می‌آورد (طبق معادله (۱)). ازآنجاکه این روش در پی جداسازی میانگین طبقه‌ها، با استفاده از فاصله ماهالانوبیس بوده و فرض می‌کند که هر طبقه می‌تواند توسط توزیع گاوسی مدل شده و همه طبقه‌ها ماتریس‌های کوواریانس مشابهی دارند، در برخورد با داده‌های ناهمگن ناتوان است [28].

$$J_{FDA}(A) = tr\left\{\left(AS_W A^T\right)^{-1}\left(AS_B A^T\right)\right\} \qquad (1)$$

$$S_W = \sum_{i=1}^{k} p_i S_i$$

$$S_B = \sum_{i=1}^{k} p_i (m_i - \bar{m})(m_i - \bar{m})^T$$

$$\bar{m} = \sum_{i=1}^{k} p_i m_i$$

که k تعداد طبقه‌ها، S_W ماتریس پراکندگی درون‌طبقه‌ای، S_B ماتریس پراکندگی بین طبقه‌ای، m_i بردار میانگین طبقه i، $\bar{m}$ میانگین تخمینی، p_i احتمال پیشین طبقه i و S_i ماتریس کوواریانس درون‌طبقه‌ای است. با استفاده از روش LDA و با کمک ماتریس انتقال A ($d \times n$ بعدی)، بردارهای ویژگی از فضای $\boldsymbol{n}$ بُعدی به فضای d بعدی به‌گونه‌ای منتقل می‌شوند که $J_{FDA}(A)$ بیشینه شود.

مشکل تکینی[2] یک مسأله مهم در LDA آن است و در شرایطی به‌وجود می‌آید که تعداد ویژگی‌ها بسیار بیشتر از تعداد داده‌ها باشد. برای حل این مشکل، روش‌های یادگیری زیرفضای چندخطی ارائه شدند [29-31]. توسعه چندخطی LDA نیز موردمطالعه قرارگرفته است [32-34]. در [35] روش LDA دو بُعدی Lp-norm تعمیم‌یافته ($p > 0$) ارائه شده است. روش یادشده جهت محاسبه پراکندگی درون‌طبقه‌ای و بین طبقه‌ای از Lp-norm استفاده کرده و با انتخاب p مناسب در مسائل کاربردی مختلف، می‌تواند به مقاومت به نوفه دست یابد.

ویژگی‌های ناهمبسته به‌طورمعمول ویژگی‌های مطلوبی در تشخیص الگو به‌شمار می‌روند؛ زیرا این نوع ویژگی‌ها شامل اطلاعات تفکیک‌کنندگی بیشتری نسبت به ویژگی‌های همبسته در همان بُعد هستند. روش ناهمبستگی تحلیل تفکیک‌کننده خطی [36]، بردارهای تفکیک‌کننده‌ای به‌دست می‌آورد که معیار فیشر را بیشینه می‌کنند؛ هرچند ویژگی‌های استخراج‌شده با محدودیت‌هایی مواجه هستند که ازلحاظ آماری ناهمبسته گفته می‌شوند. در این روش، بردارهای تفکیک‌کننده شامل ویژگی‌های تفکیک‌کننده‌ای هستند که علاوه‌بر این بیشینه‌کردن معیار فیشر، ارتباط بین ویژگی‌های استخراج‌شده را کمینه می‌کنند. در [37] مدل بهبودیافته روش ناهمبستگی ناهمگنی LDA با یک‌پارچه‌سازی معیار وزنی چیرنف ارائه شده است. با استفاده از این روش، اطلاعات تفکیک‌کننده موجود در میانگین و ماتریس کوواریانس هر طبقه استخراج می‌شود. LDA چند منظر[3]ی [38,39] ترکیبی از ترکیب دو روش تحلیل همبستگی متعارف و LDA است که در آن با درنظرگرفتن تابع هدف مناسب، تفکیک‌پذیری هر منظر و همبستگی میان دو منظر به‌طور هم‌زمان بیشینه می‌شود. LDA ناهمبسته [40,41] نسخه‌ای از LDA است که با افزودن تعدادی قید به تابع هدف آن، ویژگی‌های استخراج‌شده آن

[1] Fisher Separability Criterion

[2] Singularity Problem

[3] Multi-view

[DOI: 10.52547/jsdp.19.3.105]

دارای کمترین افزونگی است. در [42] با ترکیب دو روش تحلیل همبستگی متعارف و LDA ناهمبسته روشی برای ارائه استخراج ویژگی‌ها در مسائل چندمنظری ارائه شد. همچنین جهت غلبه بر مسائل تفکیک‌ناپذیر خطی، نسخه کرنلی آن معرفی شد.

استفاده از روش‌های آماری، رویکردی دیگر از روش‌های کاهش بُعد است. در [43] روش آماری چند متغیره برای طبقه‌بندی نمونه‌ها ارائه شد. در [44] نسخه تعمیم‌یافته این روش، با ترکیب دو روش LDA و کوادراتیک برای طبقه‌بندی نمونه‌های جمعیت آماری چندمتغیره مختلف ارائه شد. ایراد این روش حساسیت به نویز است. روش [45] مشکل حساسیت به نوفه روش [46] را برطرف کرد. روش [46] یک مجموعه داده را به‌صورت منیفلد در نظر گرفته و عدم شباهت را با جمع فاصله مؤلفه‌های خطی محلی محاسبه می‌کند. [47] مجموعه تصاویر را به‌صورت یک نقطه در روی منیفلد گراسمان در نظر می‌گیرد، آنگاه با هدف کاهش بُعد و با شناسایی ویژگی‌های متمایزکننده، آن را به فضای منیفلدی با ابعاد کم نگاشت می‌دهد.

در دهه اخیر، شبکه‌های عصبی کانولوشنی موفقیت زیادی در حوزه استخراج ویژگی از داده‌های با ابعاد بالا داشته‌اند و بدین منظور، کتابخانه‌های متعددی از یادگیری عمیق ارائه شده است [50-48] اما مهم‌ترین چالش کار با شبکه‌های عصبی عمیق، ظرفیت محدود حافظه و تعداد داده‌های آموزشی ناکافی به‌خصوص برای داده‌های با ابعاد بالا است. از این‌رو، برای بهبود عملکرد طبقه‌بند در حضور تعداد محدود داده آموزشی، میتوان از دو رویکرد اساسی استفاده کرد: (۱) استخراج ویژگی به شیوه سنتی به‌طور موازی همراه با شبکه عصبی عمیق اقدام به شناسایی ویژگی‌های مؤثر در طبقه‌بندی کنند (شکل ۱- (الف)). (۲) ابتدا استخراج ویژگی به شیوه سنتی و سپس با کمک شبکه عصبی عمیق انجام پذیرد (شکل ۱- (ب)).

۳- روش پیشنهادی

روش پیشنهادی، با هدف انتخاب داده‌های مرزی و نزدیک به مرز به‌عنوان داده‌های مؤثر در تفکیک طبقه‌ها از یکدیگر، سعی در ایجاد ماتریس پراکندگی مطلوب به‌منظور افزایش کارآیی LDA دارد. روش پیشنهادی در دو مرحله اجرا می‌شود:

۱- شناسایی نمونه‌های مرزی بر اساس بیشینه حاشیه

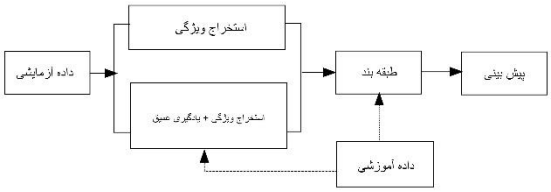
۲- طبقه‌بندی داده‌ها با معیار چیرنف

حال، به بیان جزئیات روش پیشنهادی پرداخته می‌شود.

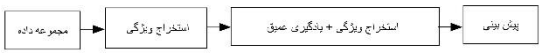
۳-۱- شناسایی نمونه‌های مرزی بر اساس بیشینه حاشیه

کاهش نمونه یک مسئله بهینه‌سازی چندهدفه است که تلاش می‌کند کیفیت طبقه‌بندی را با حفظ نمونه‌های باکیفیت افزایش و هم‌زمان اندازه مجموعه را با حذف نمونه‌های بی‌تأثیر درنتیجه طبقه‌بندی کاهش دهد. ازاین‌رو در کاهش نمونه‌ها، با مسأله برقراری مصالحه بین اندازه مجموعه آموزشی و کیفیت طبقه‌بندی مواجه هستیم. روش پیشنهادی با هدف افزایش صحت طبقه‌بندی داده‌ها، نمونه‌های مرزی و نزدیک مرزها را حفظ و نمونه‌های داخلی و یا دور از مرزها را حذف می‌کند. ذکر این نکته ضروری است که تعیین مرز طبقه‌ها و انتخاب نمونه به‌صورت توأم انجام می‌شود.

(الف)



(ب)

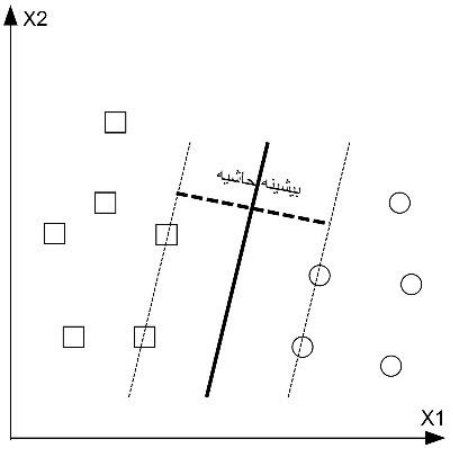


(شکل-۱): روندنمای استخراج ویژگی (الف) یادگیری سنتی و یادگیری عمیق به طور هم‌زمان ویژگی‌های متمایزکننده را شناسایی می‌کنند. (ب) ویژگی‌های استخراج‌شده توسط یادگیری سنتی به‌عنوان ورودی در یادگیری عمیق استفاده می‌شوند.

(Figure-1): The feature extraction process. (a) Both traditional and deep learning algorithms detect discriminative features simultaneously. (b) The features extracted from a traditional algorithm are considered as the input of a deep learning algorithm.

به‌منظور شناسایی نمونه‌های مؤثر در طبقه‌بندی، ابتدا لازم است حاشیه‌ها شناسایی و سپس میزان تأثیر (وزن) هر نمونه در ایجاد مرزها محاسبه شود. با این رویکرد، تأثیر نمونه‌های مؤثر در ایجاد مرزهای جداکننده تشدید شده و از تأثیر داده‌های غیرمرزی چشم‌پوشی می‌شود. شکل (۲) نمایی از بیشینه حاشیه متقارن در فضای دوبُعدی و نمونه‌های قرارگرفته بر روی حاشیه را نشان می‌دهد.

[DOI: 10.52547/jsdp.19.3.105]



(شکل-۲): بیشینه حاشیه متقارن در فضای دوبُعدی

(Figure-2): A symmetric maximum margin model in a 2-dimensional space

شکل (۳) مدل پیشنهادی را نشان می‌دهد. این مدل، بیشینه حاشیه با نمونه‌های کاهش داده‌شده را به‌عنوان مسئله برنامه‌ریزی غیرخطی مقید در نظر می‌گیرد. قید مهم آن، عدم تغییر بیشینه حاشیه بین دو طبقه، پس از اعمال کاهش نمونه است.

$$\min \sum_{i=1}^{n} a_i$$
$$s.t.\ \forall x_i \in \text{Training set}$$
constraint (I):
if $x \in class+1$
$$\min\{a_i \rho(x,x_i) \mid x_i \in \text{class}+1\} - \min\{a_i \rho(x,x_i) \mid x_i \in \text{class}-1\} \le \min\{\rho(x,x_i) \mid x_i \in \text{class}+1\} - \min\{\rho(x,x_i) \mid x_i \in \text{class}-1\} + q,$$
otherwise
$$\min\{a_i \rho(x,x_i) \mid x_i \in \text{class}-1\} - \min\{a_i \rho(x,x_i) \mid x_i \in \text{class}+1\} \le \min\{\rho(x,x_i) \mid x_i \in \text{class}-1\} - \min\{\rho(x,x_i) \mid x_i \in \text{class}+1\} + q,$$
Constraint (II):
$$a_i \in \{0,1\} \text{ for } i = 1,\ldots,n$$

(شکل-۳): مدل حداکثر حاشیه پیشنهادی به‌عنوان مسأله برنامه‌ریزی دودویی ترکیبی

(Figure-3): The proposed maximum margin model as a binary compound programing problem

در مدل ارائه‌شده در شکل (۳)، ضرایب α_i برای تعیین تأثیر هر نمونه در ساخت مرزهای جداکننده به‌کار می‌روند. چنانچه α_i برابر صفر باشد، آن‌گاه نمونه i در ایجاد مرز بین دو طبقه بی‌تأثیر بوده و می‌توان آن را از مجموعه آموزشی حذف کرد. هدف از تابع $\min \sum_{i=1}^{n} \alpha_i$ یافتن بیشترین تعداد α_i ای است که برابر با صفر شوند، است. در این صورت، نمونه‌های بیشتری حذف می‌شوند. قید (I) به تأثیر نمونه‌ها بر روی قواعد نزدیک‌ترین همسایه در طبقه‌بندی داده آزمایشی x توجه دارد. در قید (I) مدل حداکثر حاشیه $\min\{\partial(x,x_i) \mid x_i \in \text{class}+1\}$ برای نشان‌دادن فاصله بین داده آزمایشی x و نزدیک‌ترین همسایه‌های آن در طبقه $+1$ به کار می‌رود. در این قید، $\min\{\partial(x,x_i) \mid x_i \in \text{class}-1\} - \min\{\partial(x,x_i) \mid x_i \in \text{class}+1\}$ برای اندازه‌گیری اختلاف فاصله داده x از نزدیک‌ترین همسایه‌هایش در دو طبقه استفاده شده است. علامت این عبارت، طبقه x را مشخص می‌کند. $\partial(x,x_i)$ فاصله اقلیدسی میان داده آزمایشی x و داده آموزشی x_i را محاسبه می‌کند. ذکر این نکته ضروری است که در زمان آموزش، داده آموزشی به‌عنوان داده آزمایشی مورد استفاده قرار می‌گیرد. اگر تعیین طبقه x به‌وسیلهٔ x_i تحت تأثیر قرار نگیرد آن‌گاه α_i به صفر تغییر کرده و x_i را از مجموعه آموزشی حذف می‌کنیم. پارامتر q یک ورودی قابل تنظیم توسط کاربر است و برای مقابله با نوفه استفاده می‌شود.

مدل بیشینه حاشیه معرفی‌شده در شکل (۳) به شکل دیگری قابل بیان است:

$$\text{find } f(g) \qquad (۲)$$
$$s.t.\ f(x) = \begin{cases} -1\min\{\alpha_i \partial(x,x_i) \mid x_i \in \text{class}-1\} < \min\{\alpha_i \partial(x,x_i) \mid x_i \in \text{class}+1\} \\ , \\ +1\min\{\alpha_i \partial(x,x_i) \mid x_i \in \text{class}+1\} < \min\{\alpha_i \partial(x,x_i) \mid x_i \in \text{class}-1\} \end{cases}$$

به‌عبارتی‌دیگر، داریم:

$$f(x) = sign\begin{pmatrix} \min\{\alpha_i \partial(x,x_i) \mid x_i \in \text{class}-1\} \\ -\min\{\alpha_i \partial(x,x_i) \mid x_i \in \text{class}+1\} \end{pmatrix} \qquad (۳)$$

بنابراین، فاصله از هر نقطه روی سطح تصمیم‌گیری تا نزدیک‌ترین نمونهٔ طبقه -1 برابر با فاصله از آن نقطه از سطح تصمیم‌گیری تا نزدیک‌ترین نمونه طبقه $+1$ است. ازاین‌رو $f(x)$ نشان‌دهنده بیشینه حاشیه متقارن و $f(x)=0$ نشان‌دهنده صفحه جداکننده در میانه مرزی بین دو طبقه است. ازآنجاکه مدل بیشینه حاشیه در شکل (۳) دربرگیرنده متغیرهای گسسته و پیوسته است، می‌توان آن را یک مسأله برنامه‌ریزی دودویی ترکیبی در نظر گرفت؛ بنابراین، قیدهای گسسته $\alpha_i \in \{0,1\} \text{ for } i = 1,\ldots,n$ به‌صورت پیوسته آن، یعنی $\alpha_i(\alpha_i - 1) = 0, \text{ for } i = 1,\ldots,n$ نمایش داده می‌شوند. شکل (۴) مسأله برنامه‌ریزی غیرخطی صفر-یک با قید پیوسته را نشان می‌دهد. در روش پیشنهادی، از روش تابع

[DOI: 10.52547/jsdp.19.3.105]

پرکننده [17] برای حل برنامه‌ریزی غیرخطی، استفاده شده است.

$\min \sum_{i=1}^{n} a_i$
$s.t.\ \forall x_i \in \text{Training set}$
constraint (I):
if $x \in \text{class}+1$
$\min\{a_i \rho(x,x_i) \mid x_i \in \text{class}+1\} - \min\{a_i \rho(x,x_i) \mid x_i \in \text{class}-1\} \le$
$\min\{\rho(x,x_i) \mid x_i \in \text{class}+1\} - \min\{\rho(x,x_i) \mid x_i \in \text{class}-1\} + q,$
otherwise
$\min\{a_i \rho(x,x_i) \mid x_i \in \text{class}-1\} - \min\{a_i \rho(x,x_i) \mid x_i \in \text{class}+1\} \le$
$\min\{\rho(x,x_i) \mid x_i \in \text{class}-1\} - \min\{\rho(x,x_i) \mid x_i \in \text{class}+1\} + q,$
Constraint (P):
$a_i(a_i - 1) = 0 \quad \text{for } i = 1,\ldots,n$

(شکل-۴): مسأله برنامه‌ریزی غیرخطی پیشنهادی
(Figure-4): The proposed nonlinear programing problem

تابع پرکننده به‌وسیلهٔ رابطه (۴) بیان می‌شود.

$$P_{r,c,q,x^*}(x) = \frac{1}{\|x - x^*\|^2 + 1} f_{r,c}\left(g_r(f(x) - f(x^*)) + \sum_{i=1}^{m} g_{\frac{r}{q}}(g_i(x) - 2r)\right) \quad (۴)$$

که در آن x^* کمینه‌ساز تابع $f(x)$ در مسأله P و q,c,r پارامترهای قابل تنظیم‌اند. توابع پیوسته f و g نیز به‌صورت زیر تعریف می‌شوند:

$$f_{r,c}(t) = \begin{cases} c & \text{if } t \ge 0 \\ -\frac{2c}{r^3}t^3 - \frac{3c}{r^2} + c & \text{if } -r < t \le 0 \\ 0 & \text{if } t \le -r \end{cases} \quad (۵)$$

$$g_r(t) = \begin{cases} t+2 & \text{if } t \ge 0 \\ \frac{r-4}{t^3}t^3 + \frac{2r-6}{t^2}t^2 + t + 2 & \text{if } -r < t \le 0 \\ 0 & \text{if } t \le -r \end{cases} \quad (۶)$$

تابع f دارای نقش پرکننده در $P_{r,c,q,x^*}(x)$ است و تابع g به‌عنوان تابع آزاد برای کار با قیدها استفاده می‌شود. اگر نقطه کمینه محلی روش نخست از مسأله اصلی یک نقطه کمینه کلی نباشد، آن‌گاه می‌توان نقطه کمینه دیگری برای مسئلهٔ اصلی با اعمال جستجوی محلی و تکرار این عمل تا زمانی که یک نقطه کمینه‌ی تقریبی از آن حاصل شود، به‌دست آورد؛ درنهایت، مسأله بهینه‌سازی مقید بسته‌ای به نام مسأله تابع پرکننده به‌صورت زیر معرفی می‌شود:

$$(\bar{P}) \min_{x \in X} P_{r,c,q,x^*}(x) \quad (۷)$$

که تابع پرکننده $P_{r,c,q,x^*}(x)$ در رابطه (۴) تعریف شده است. چنانچه کمینه‌کننده محلی فعلی x^* یک کمینه‌کننده برای (P) نباشد، نقطه $\bar{x} \in S$ با شرط $f(\bar{x}) < f(x^*)$ با استفاده از برخی طرح‌های جستجوی محلی برای مسئله $\bar{P}$ قابل محاسبه است؛ بنابراین، می‌توان کمینه‌کننده محلی بهتری برای (P) با اعمال جستجوی محلی به مسألهٔ (P) با شروع از $\bar{x}$ به‌دست آورد. در الگوریتم انتخاب نمونه بیشینه حاشیه باید توجه کرد که $f(x)$ همانند تابع هدف در مسأله برنامه‌ریزی غیرخطی در شکل (۴) است و $g_i(x)$ برای $i = 1,\ldots,m$ به‌صورت زیر تعریف می‌شود:

قید (I) به‌وسیله رابطه (۸) تعریف شده است.

$$\begin{aligned} &\text{if } x \in \text{class}+1 \\ &\min\{\alpha_i \partial(x,x_i) \mid x_i \in \text{class}+1\} - \\ &\min\{\alpha_i \partial(x,x_i) \mid x_i \in \text{class}-1\} - \\ &\min\{\partial(x,x_i) \mid x_i \in \text{class}+1\} + \\ &\min\{\partial(x,x_i) \mid x_i \in \text{class}-1\} + \theta, \\ &\text{otherwise} \\ &\min\{\alpha_i \partial(x,x_i) \mid x_i \in \text{class}-1\} - \\ &\min\{\alpha_i \partial(x,x_i) \mid x_i \in \text{class}+1\} - \\ &\min\{\partial(x,x_i) \mid x_i \in \text{class}-1\} + \\ &\min\{\partial(x,x_i) \mid x_i \in \text{class}+1\} - \theta. \end{aligned} \quad (۸)$$

قید (P) به‌صورت رابطه (۹) تعریف می‌شود.

$$\begin{aligned} &\alpha_i(\alpha_i - 1) \le 0, \ \text{for } i = 1,\ldots,n \\ &\alpha_i(1 - \alpha_i) \le 0, \ \text{for } i = 1,\ldots,n \end{aligned} \quad (۹)$$

در الگوریتم انتخاب نمونه بیشینه حاشیه، m یک عدد مثبت کوچک و M یک عدد بزرگ مثبت است و n تعداد ویژگی‌های مجموعه داده را مشخص می‌کند. الگوریتم تابع پرکننده پیشنهادی جهت انتخاب نمونه بیشینه حاشیه در شکل (۵) نشان داده شده است.

گام ۱: $k = 2n$، $r = 1$، $c = 1$، $l = 1$، $j = 1$، $q = 1000$، $M = 10^8$، $m = 10^{-8}$ و $e_1, e_2, \ldots, e_k$ بردارهای واحد با ابعاد n.
گام ۲: محاسبه تابع پرکننده طبق رابطه (۴).
گام ۳: اگر $j \le k$ باشد، آنگاه یک l $(l \le 1)$ نامنفی

[DOI: 10.52547/jsdp.19.3.105]

$y^j = x^* + l\, e_j$ انتخاب نموده، در غیر این صورت به گام ۶ برو.

گام ۴: جستجو برای یافتن y^* نقطه کمینه‌کننده محلی تابع پرکننده، با استفاده از روش شیب توأم. همچنین $x^0 = y^*$ را تنظیم کن.

گام ۵: یک نقطه کمینه‌کننده محلی مسئله اصلی با استفاده از روش SQP[1] با شروع از x^0 پیدا کن و به گام ۲ برو.

گام ۶: اگر $q \leq M$، آنگاه q را در ۱۰ ضرب کن و $j = 1$ را قرار بده، در غیر این صورت به گام ۷ برو.

گام ۷: اگر $c \leq M$، آنگاه c را در ۱۰ ضرب کن، و $j = 1$ و $q = 1000$ را قرار بده و به گام ۲ برو، در غیر این صورت به گام ۸ برو.

گام ۸: اگر m r^3، آنگاه r را بر ۱۰ تقسیم کن و $c = 1$، $q = 1000$ قرار بده و به گام ۲ برو، در غیر این صورت الگوریتم را متوقف کن و x^* یک کمینه‌کننده کلی و یا تقریبی از آن برای مسئله اصلی است.

(شکل-۵): الگوریتم تابع پُرکننده پیشنهادی برای انتخاب نمونه بیشینه حاشیه

(Figure-5): The proposed filled function for selecting instances on the maximum margin

۳-۲- ماتریس پراکندگی چیرنف با داده‌های کاهش‌یافته

در این مرحله، با هدف شناسایی ویژگی‌های مؤثر در طبقه‌بندی و کسب اطلاعات تفکیک‌کننده موجود در طبقه‌ها، از فاصله توزیع معیار چیرنف استفاده می‌شود. مزیت استفاده از این معیار آن است که برخلاف روش LDA که تفکیک‌کننده بهینه را برای دو طبقه $l = 2$ پیدا می‌کند، معیار چیرنف می‌تواند در چند طبقه را از یکدیگر تفکیک کند. با درنظرگرفتن این مسأله، ماتریس‌های پراکندگی میان‌طبقه‌ای و درون‌طبقه‌ای با استفاده از روابط (۱۰) و (۱۱) محاسبه و در معیار چیرنف (رابطه (۱۲)) جایگذاری می‌شوند. نمونه‌های مرزی و غیرمرزی به‌دست‌آمده از مرحله قبل، به‌عنوان ورودی این روابط در نظر گرفته می‌شوند. روند این مرحله از روش پیشنهادی در شکل (۶) ارائه شده است:

$$S^{(B)} = \sum_{i=1}^{l} \sum_{j=1}^{n^{(B)}} \left(x_j^{(B)} - m_i\right)\left(x_j^{(B)} - m_i\right)^T \qquad (۱۰)$$

$$S^{(W)} = \sum_{i=1}^{l} \sum_{j=i} \left(x_j^{(NB)} - m_i\right)\left(x_j^{(NB)} - m_i\right)^T \qquad (۱۱)$$

که $x^{(B)}$ نمونه‌های مرزی و $x^{(NB)}$ نمونه‌های غیرمرزی است.

[1] Sequential Quadratic Programming

(۱۲)

$$J_C(A) = \sum_{i=1}^{l-1} \sum_{j=i+1}^{l} P_i P_j tr\left\{\left(AS_W A^T\right)^{-1} \times AS_W^{\frac{1}{2}} \left\{ \begin{array}{l} \left(S_W^{-\frac{1}{2}} S_{ij}^{(W)} S_W^{-\frac{1}{2}}\right)^{-\frac{1}{2}} \times \left(S_W^{-\frac{1}{2}} S_{ij}^{(B)} S_W^{-\frac{1}{2}}\right) \\ \left(S_W^{-\frac{1}{2}} S_{ij}^{(W)} S_W^{-\frac{1}{2}}\right)^{-\frac{1}{2}} \\ + \frac{1}{\pi_i \pi_j} \left(\begin{array}{l} \log\left(S_W^{-\frac{1}{2}} S_{ij}^{(W)} S_W^{-\frac{1}{2}}\right) \\ -\pi_i \log\left(S_W^{-\frac{1}{2}} S_i^{(W)} S_W^{-\frac{1}{2}}\right) \\ -\pi_j \log\left(S_W^{-\frac{1}{2}} S_j^{(W)} S_W^{-\frac{1}{2}}\right) \end{array} \right) \end{array} \right\} AS_W^{\frac{1}{2}} \right\}$$

برای تعیین بردار A، ماتریس تجزیه مقدار ویژه رابطه (۱۳) تشکیل شده $(S_C W = l\, S_W W)$؛ و $(A = [W_1, \ldots, W_d])$ معادل با بردار ویژه بیشترین مقدار ویژه انتخاب می‌شود.

line

$$S_C = \sum_{i=1}^{l-1} \sum_{j=i+1}^{l} P_i P_j tr\left\{\left(S_W\right)^{-1} \times S_W^{\frac{1}{2}} \left\{ \begin{array}{l} \left(S_W^{-\frac{1}{2}} S_{ij}^{(W)} S_W^{-\frac{1}{2}}\right)^{-\frac{1}{2}} \times \left(S_W^{-\frac{1}{2}} S_{ij}^{(B)} S_W^{-\frac{1}{2}}\right) \\ \left(S_W^{-\frac{1}{2}} S_{ij}^{(W)} S_W^{-\frac{1}{2}}\right)^{-\frac{1}{2}} \\ + \frac{1}{\pi_i \pi_j} \left(\begin{array}{l} \log\left(S_W^{-\frac{1}{2}} S_{ij}^{(W)} S_W^{-\frac{1}{2}}\right) \\ -\pi_i \log\left(S_W^{-\frac{1}{2}} S_i^{(W)} S_W^{-\frac{1}{2}}\right) \\ -\pi_j \log\left(S_W^{-\frac{1}{2}} S_j^{(W)} S_W^{-\frac{1}{2}}\right) \end{array} \right) \end{array} \right\} S_W^{\frac{1}{2}} \right\}$$

ورودی: ماتریس داده‌های آموزشی $A = \{a_i \mid a_i \in \mathbb{R}^d\}$ و برچسب $B_i \in \{1, 2, \ldots, l\}$.

خروجی: ماتریس انتقال W_C.

۱. به‌دست‌آوردن ماتریس پراکندگی از نمونه‌های مرزی و غیرمرزی

۲. برای هر طبقه $i = 1, \ldots, l$ موارد زیر را محاسبه کن:

الف) $X_{(i)} \leftarrow \{X_j\}_{y_j = i}$

ب)

$S^{(B)} \leftarrow S^{(B)} + (X^{(B)} - m_i (l_{n^{(B)}})^T)(X^{(B)} - m_i (l_{n^{(B)}})^T)^T$

که در آن $l_{n^{(B)}}$ یک بردار با مقدار یک است.

ج)

[DOI: 10.52547/jsdp.19.3.105]

$$S^{(W)} \leftarrow S^{(W)} + (X^{(NB)} - m_i (l_{n(NB)})^T)(X^{(NB)} - m_i (l_{n(NB)})^T)^T$$

که در آن $l_{n(NB)}$ یک بردار با مقدار یک است.

ه)

$$S_C = \sum_{i=1}^{l-1}\sum_{j=i+1}^{l} P_i P_j tr\left\{(S_W)^{-1}\times S_W^{1/2}\begin{pmatrix}\left(S_W^{-1/2}S_{ij}^{(W)}S_W^{-1/2}\right)^{-1/2}\\ \times\left(S_W^{-1/2}S_{ij}^{(B)}S_W^{-1/2}\right)\\ \left(S_W^{-1/2}S_{ij}^{(W)}S_W^{-1/2}\right)^{-1/2}\\ +\frac{1}{\pi_i\pi_j}\begin{pmatrix}\log\left(S_W^{-1/2}S_{ij}^{(W)}S_W^{-1/2}\right)\\ -\pi_i\log\left(S_W^{-1/2}S_i^{(W)}S_W^{-1/2}\right)\\ -\pi_j\log\left(S_W^{-1/2}S_j^{(W)}S_W^{-1/2}\right)\end{pmatrix}\end{pmatrix}S_W^{1/2}\right\}$$

۳. محاسبه ماتریس پراکندگی

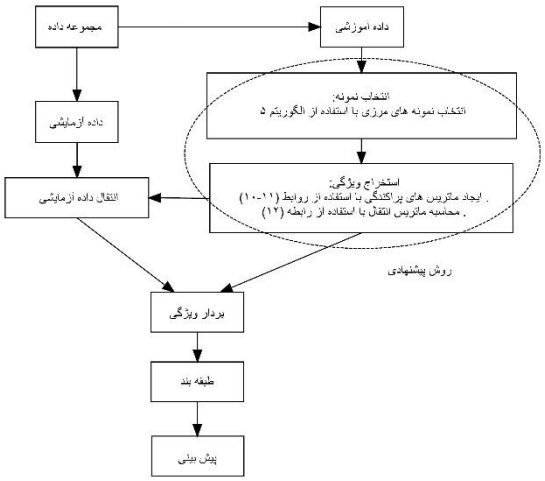
الف) $S_C W = l S_W W,\ l_1 \geq l_2 \geq \ldots \geq l_d$

ب) $A = [W_1,\ldots,W_d]$

(شکل-۶): روش پیشنهادی

(Figure-6): The proposed method

جهت درک بهتر روش پیشنهادی، روندنمای آن در شکل (۷) نشان داده شده است.



(شکل-۷): روندنمای روش پیشنهادی

(Figure-7): The flowchart of the proposed method

۴- ارزیابی و مقایسه روش پیشنهادی

در این بخش، نتایج حاصل از پیاده‌سازی روش پیشنهادی بر روی سیستم رایانه‌ای با پردازنده اینتل ۲.۷۰ گیگا هرتز و ۸ گیگا بایت رم و با نرم‌افزار MATLAB R2016 بر روی هفده مجموعه‌داده مستخرج از پایگاه داده UCI بیان شده است. خصوصیات مجموعه‌داده‌ها در جدول (۱) نمایش داده شده است. در ارزیابی عملکرد، علاوه‌بر روش مرز دانش [35][1]، از طبقه‌بندهای سنتی درخت تصمیم، شبکه عصبی و kNN با اعمال دو روش استخراج ویژگی LDA و روش پیشنهادی استفاده شده است. روش‌های یادشده از نظر معیارهای صحت و زمان اجرا مقایسه شده‌اند. اعتبارسنجی به شیوه ضربدری ده‌تایی[2] انجام شده است.

(جدول-۱): مجموعه‌داده‌ها

(Table-1): Datasets

مجموعه داده	تعداد داده	تعداد طبقه	تعداد ویژگی
Iris	150	3	5
Liver	345	2	6
Hepatitis	137	2	34
Balance-scale	150	3	5
Hayes roth	132	3	5
German credit	1000	2	38
New-thyroid	215	3	6
Banknote authentication	1370	2	3
Wine	178	3	13
Breast cancer Winconsin	699	2	11
Zoo	101	2	16
Vote	435	2	16
Australian credit	653	2	51
Primary tumor	336	2	15
Haberman	31	2	3
Ads	3279	2	1558
Census	29926	2	409

۵- نتایج

به‌منظور درک بهتر عملکرد هر روش، نتایج حاصل از اجرای طبقه‌بندها بر روی هر مجموعه‌داده رتبه‌بندی و سپس میانگین رتبه هر روش در آخرین سطر جداول نمایش داده شده است. جهت نمایش رتبه، از علامت پرانتز استفاده شده است.

برای بررسی تأثیر استخراج ویژگی بر صحت و زمان اجرای طبقه‌بندی، آزمایش‌ها در دو حالت قبل و بعد از اجرای پیش‌پردازش ویژگی انجام شد. طبق نتایج مندرج در جدول (۲)، شبکه عصبی بیشترین و طبقه‌بند kNN کمترین میزان صحت را بر روی مجموعه‌داده‌های استاندارد داشته‌اند. همچنین طبق نتایج مندرج در جدول (۳)، طبقه‌بند شبکه عصبی با اجرای استخراج ویژگی به روش پیشنهادی، همچنان صحت بالای طبقه‌بندی را

[1] http://www.optimal-group.org/Resources/Code/G2DLDA.html

[2] 10-fold cross validation

[DOI: 10.52547/jsdp.19.3.105]

حفظ کرده است. طبقه‌بند درخت تصمیم با اجرای استخراج ویژگی به روش پیشنهادی نیز رتبه دوم را دارد.

(جدول-۲): نرخ صحت

(Table-2): Comparison of Accuracy (%)

مجموعه داده	*k*NN	درخت تصمیم	شبکه عصبی
Iris	81.79(3)	82.74(2)	90.62(1)
Liver	74.26(2)	75.53(1)	70.02(3)
Hepatitis	59.88(2)	59.86(3)	75.99(1)
Balance-scale	50.37(3)	51.85(2)	65.64(1)
Hayes roth	78.25(3)	78.86(2)	86.24(1)
German credit	56.25(3)	67.66(2)	68.98(1)
New-thyroid	36.77(3)	37.54(2)	89.17(1)
Banknote authentication	67.89(2)	67.00(3)	97.59(1)
Wine	79.71(3)	80.63(2)	90.53(1)
Breast cancer Winconsin	84.41(3)	85.35(2)	95.75(1)
Zoo	74.54(3)	75.32(2)	86.13(1)
Vote	74.76(2)	74.71(3)	94.69(1)
Australian credit	64.26(3)	65.17(2)	82.43(1)
Primary tumor	59.18(3)	60.12(2)	70.34(1)
Haberman	65.46(3)	76.27(1)	74.15(2)
Ads	75.21(3)	79.17(2)	84.05(1)
Census	78.16(2)	73.19(3)	81.36(1)
Average rank	2.69(3)	2.13(2)	1.19(1)

(جدول-۳): نرخ صحت ٪ (با اعمال پیش‌پردازش ویژگی)

(Table-3): Comparison of Accuracy (%) (Feature Preprocessing)

مجموعه داده	[35]	*k*NN (LDA)	درخت تصمیم (LDA)	شبکه عصبی (LDA)	*k*NN (روش پیشنهادی)	درخت تصمیم (روش پیشنهادی)	شبکه عصبی (روش پیشنهادی)
Iris	91.49(3)	85.84(6)	86.79(4)	71.62(7)	86.35(5)	91.72(2)	94.67(1)
Liver	66.57(3)	55.13(7)	56.46(6)	60.14(4)	56.53(5)	66.72(2)	70.89(1)
Hepatitis	63.49(5)	63.93(3)	63.91(4)	63.15(7)	63.46(6)	64.37(2)	80.04(1)
Balance-scale	54.84(6)	54.87(5)	56.35(3)	54.64(7)	56.25(4)	67.23(2)	70.14(1)
Hayes roth	82.33(3)	80.77(6)	81.38(5)	78.44(7)	82.15(4)	85.75(2)	88.76(1)
German credit	62.61(3)	60.77(6)	62.18(5)	57.39(7)	62.42(4)	64.28(2)	73.5(1)
New-thyroid	95.21(1)	40.62(7)	41.39(6)	44.89(4)	42.04(5)	75.21(3)	93.02(2)
Banknote authentication	86.99(4)	59.13(5)	58.24(7)	94.24(3)	58.87(6)	98.31(2)	98.83(1)
Wine	86.81(4)	83.89(7)	84.81(5)	86.85(3)	84.25(6)	91.49(2)	94.71(1)
Breast cancer Winconsin	78.82(7)	86.57(5)	87.51(4)	86.21(6)	87.81(3)	95.55(2)	97.91(1)
Zoo	93.56(1)	79.41(6)	80.19(5)	78.42(7)	80.29(4)	85.19(3)	91(2)
Vote	75.06(4)	74.78(5)	74.73(6)	93.23(3)	74.32(7)	94.47(2)	94.71(1)
Australian	66.9	65.7	66.6	65.9	66.8	73.4	83.9

طبق نتایج مندرج در جدول (۴)، شبکه عصبی کمترین و طبقه‌بند kNN بیشترین زمان محاسبه (بر حسب ثانیه) بر روی مجموعه‌داده‌های استاندارد داشته‌اند. همچنین طبق نتایج مندرج در جدول (۵)، طبقه‌بند kNN با اجرای استخراج ویژگی به روش پیشنهادی کمترین زمان محاسبه را داشته است. گفتنی است جهت زمان محاسبه، میانگین زمان اجرا در اعتبارسنجی ضربدری ده‌تایی در نظر گرفته شده است.

(جدول-۴): زمان محاسبه (s)

(Table-4): Comparison of Computational Time (s)

مجموعه داده	*k*NN	درخت تصمیم	شبکه عصبی
Iris	6.53(2)	7.52(3)	5.24(1)
Liver	7.61(2)	8.35(3)	6.84(1)
Hepatitis	8.52(3)	6.84(1)	7.36(2)
Balance-scale	49.47(3)	47.99(1)	48.66(2)
Hayes roth	50.14(3)	48.35(1)	48.95(2)
German credit	41.39(1)	42.78(2)	43.96(3)
New-thyroid	35.93(3)	35.62(2)	33.03(1)
Banknote authentication	71.16(2)	71.78(3)	71.01(1)
Wine	68.20(3)	66.34(2)	65.45(1)
Breast cancer Winconsin	105.73(1)	106.58(3)	106.01(2)
Zoo	54.87(2)	55.22(3)	53.52(1)
Vote	24.63(3)	23.63(1)	24.44(2)
Australian credit	24.48(1)	26.92(3)	25.64(2)
Primary tumor	31.10(1)	33.05(2)	33.93(3)
Haberman	20.74(2)	21.15(3)	20.14(1)
Ads	179.68(3)	178.47(1)	178.78(2)
Census	168.12(3)	166.67(1)	166.78(2)
Average rank	2.25(3)	2.00(2)	1.75(1)

(جدول-۵): زمان محاسبه (s) (با اعمال پیش‌پردازش ویژگی)

(Table-5): Comparison of Time (s) (Feature Preprocessing)

مجموعه داده	[35]	*k*NN (LDA)	درخت تصمیم (LDA)	شبکه عصبی (LDA)	*k*NN (روش پیشنهادی)	درخت تصمیم (روش پیشنهادی)	شبکه عصبی (روش پیشنهادی)
Iris	9.75(6)	6.96(4)	12.69(7)	8.54(5)	5.63(1)	6.83(3)	5.8(2)
credit	(3)	4(7)	5(5)	(6)	2(4)	6(2)	1(1)
Primary tumor	60.96(4)	59.28(7)	60.22(6)	62.95(3)	60.58(5)	64.89(2)	70.44(1)
Haberman	65.38(3)	59.97(6)	60.78(5)	54.65(7)	61.44(4)	73.85(2)	78.66(1)
Ads	71.38(3)	57.58(7)	62.23(5)	73.12(2)	58.18(6)	64.09(4)	77.93(1)
Census	72.83(2)	61.04(7)	65.22(5)	72.48(3)	62.92(6)	67.19(4)	75.38(1)
Average rank	3.5(3)	6.00(6)	5.13(5)	4.94(4)	4.94(4)	2.38(2)	1.13(1)

[DOI: 10.52547/jsdp.19.3.105]

6.17 (3)	7.18 (6)	4.55 (1)	6.95 (5)	10.71(7)	6.32 (4)	5.94 (2)	Liver
10.71(5)	6.47 (1)	7.67 (2)	9.76 (4)	12.13(6)	12.45(7)	8.73 (3)	Hepatitis
47.7 (3)	46.66(1)	47.22(2)	50.3 (6)	50.11(5)	49.42(4)	56.08(7)	Balance-scale
52.62(5)	50.01(2)	48.77(1)	51.61(4)	50.83(3)	53.41(6)	54.07(7)	Hayes roth
47.08(5)	42.88(2)	41.66(1)	43.61(3)	44.88(4)	47.24(6)	50.53(7)	German credit
34.26(3)	32.69(1)	34.14(2)	36.35(5)	39.07(6)	36.06(4)	43.15(7)	New-thyroid
69.45(2)	69.38(1)	69.72(3)	74.73(7)	72.96(6)	70.64(4)	70.66(5)	Banknote authentication
67.06(3)	66.66(2)	65.27(1)	67.17(4)	67.35(5)	67.4 (7)	67.38(6)	Wine
106.13(3)	104.48(1)	105.18(2)	106.85(5)	107.1(6)	107.78(7)	106.67(4)	Breast cancer Wisconsin
57.2 (4)	54.31(1)	54.75(2)	58.45(6)	55.23(3)	57.58(5)	60.96(7)	Zoo
27.87(6)	24.28(2)	23.79(1)	24.72(3)	27.76(5)	28.53(7)	26.9 (4)	Vote
26.56(3)	24.77(2)	21.39(1)	27.72(5)	30.13(7)	27.85(6)	26.59(4)	Australian credit
31.24(3)	29.91(1)	30.44(2)	37.95(7)	32.81(5)	32.93(6)	31.79(4)	Primary tumor
19.53(2)	19.69(3)	17.83(1)	23.02(6)	22.8 (5)	20.12(4)	26.29(7)	Haberman
183.26(4)	180.21(2)	178.95(1)	182.88(3)	183.94(7)	183.46(6)	183.32(5)	Ads
168.04(4)	164.92(1)	165.22(2)	173.72(7)	169.03(5)	169.39(6)	166.84(3)	Census
3.63 (3)	1.81 (2)	1.56 (1)	5.00 (4)	5.31 (6)	5.56 (7)	5.13 (5)	**Average rank**

۶- نتیجه‌گیری

در این مقاله به‌منظور استخراج ویژگی و با هدف بهبود عملکرد LDA در حل مسائل چند طبقه‌ای و همچنین مواجهه با داده‌های ناهمگن، روشی مبتنی بر فاصله توزیعی چیرنف معرفی شده است. نخست، روش پیشنهادی با ارائه مسأله بهینه‌سازی دودویی مقید و حل آن به‌وسیله الگوریتم‌های پرکننده به کاهش داده‌ها می‌پردازد. در روش پیشنهادی، نمونه‌های غیرمرزی حذف و تنها نمونه‌های مرزی حفظ می‌شوند. فرآیند انتخاب نمونه پیشنهادی مبتنی بر ابرصفحه جداکننده برای ایجاد یک حاشیه بیشینه‌ای جداکننده است؛ سپس، ماتریس‌های پراکندگی معیار چیرنف تنها با استفاده از داده‌های مرزی ساخته می‌شود. روش پیشنهادی علاوه‌بر روش مرز دانش، با طبقه‌بندهای شبکه عصبی، درخت تصمیم و kNN از نظر صحت طبقه‌بندی و زمان محاسبه مقایسه شد. نتایج نشان می‌دهند در صورت استفاده از روش پیشنهادی، علاوه‌بر حفظ عملکرد طبقه‌بندها، زمان محاسبه نیز کاهش می‌یابد. از آنجاکه در روش پیشنهادی از تابع پرکننده استفاده می‌شود، مقداردهی دقیق پارامترهای آن چالش روش پیشنهادی است.

جهت کارهای آینده پیشنهاد می‌شود از روش انتخاب ویژگی پیشنهادی، جهت بهبودی کارایی سایر طبقه‌بندها استفاده شود. همچنین، می‌توان از روش پیشنهادی به‌عنوان رویکرد تکمیلی روش‌های یادگیری عمیق جهت غلبه بر مشکل کمبود داده‌های برچسب‌دار استفاده کرد.

۷- مراجع 7- References

[1] S. Hakak, M. Alazab, S. Khan, T. R. Gadekallu, P. K. R. Maddikunta, and W. Z. Khan, "An ensemble machine learning approach through effective feature extraction to classify fake news," *Future Generation Computer Systems*, vol. 117, pp. 47-58, 2021, doi: https://doi.org/10.1016/j.future.2020.11.022.

[2] C. H. Shen, "Feature Extraction and Classification of Heart Murmurs Based on Acoustic Qualities," *IRBM*, 2021, doi: https://doi.org/10.1016/j.irbm.2021.02.002.

[3] T. Luo, C. Hou, F. Nie, and D. Yi, "Dimension Reduction for Non-Gaussian Data by Adaptive Discriminative Analysis," *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 933-946, 2019, doi: 10.1109/TCYB.2018.2789524.

[4] F. Wang, L. Zhu, L. Xie, Z. Zhang, and M. Zhong, "Label propagation with structured graph learning for semi-supervised dimension reduction," *Knowledge-Based Systems*, pp. 107130, 2021.

[5] A. Deshpande and R. Pratap, "Sampling-based dimension reduction for subspace approximation with outliers," *Theoretical Computer Science*, vol. 858, pp. 100-113, 2021, doi: https://doi.org/10.1016/j.tcs.2021.01.021.

[6] S. Fazili, J. Grover, S. Wazir, and I. Mehta, "Recent Trends in Dimension Reduction Methods," *ICIDSSD*, pp. 68, 2021.

[7] O. Mokhlessi, S. Seyed Mahdavi Chabok, and A. Alirezaee, "Selecting effective features from Phonocardiography by Genetic Algorithm based on Pearson`s Coefficients Correlation," *jsdp*, vol. 17, no. 3, pp. 157-176, 2020, doi: 10.29252/jsdp.17.3.157.

[۱۱] ا. مخلصی ، ج. سید مهدوی چابک ، آ. علیرضائی. انتخاب ویژگی‌های مؤثر در ناهنجاری‌های دریچه‌ای

[DOI: 10.52547/jsdp.19.3.105]

قلب با استفاده از الگوریتم ژنتیک بر اساس ارزیابی همبستگی پیرسون. پردازش علائم و داده‌ها ۱۳۹۹؛ ۱۷ (۳): ۱۵۷-۱۷۶

[8] S. Weiss, I. K. Proudler, and F. K. Coutts, "Eigenvalue decomposition of a parahermitian matrix: extraction of analytic eigenvalues," *IEEE Transactions on Signal Processing*, vol. 69, pp. 722-737, 2021.

[9] B. Feng and G. Wu, "Revisiting the low-rank eigenvalue problem," *Applied Mathematics Letters*, vol. 112, p. 106706, 2021.

[10] Z. Fan, Y. Xu, and D. Zhang, "Local linear discriminant analysis framework using sample neighbors," *IEEE Transactions on Neural Networks*, vol. 22, no. 7, pp. 1119-1132, 2011.

[11] H. R. Ghaffari and A. Jalali Mojahed, "Feature extraction based on the more resolution of the classes using auxiliary classifiers," (in eng), *Signal and Data Processing*, Research vol. 18, no. 2, pp. 29-44, 2021

[۱۱] ح. غفاری ، آ. جلالی مجاهد. استخراج ویژگی مبتنی بر تفکیک‌پذیری بیشتر رده‌ها با استفاده از طبقه‌بندهای کمکی. پردازش علائم و داده‌ها ۱۴۰۰؛ ۱۸ (۲): ۲۹-۴۴

[12] L. Rueda and M. J. P. R. Herrera, "Linear dimensionality reduction by maximizing the Chernoff distance in the transformed space," *Pattern Recognition*, vol. 41, no. 10, pp. 3138-3152, 2008.

[13] L. Rueda and M. Herrera, "Linear dimensionality reduction by maximizing the Chernoff distance in the transformed space," *Pattern Recognition*, vol. 41, no. 10, pp. 3138-3152, 2008, doi: https://doi.org/10.1016/j.patcog.2008.01.016.

[14] K. M. Nair, P. Sankaran, and S. Smitha, "Chernoff distance for truncated distributions," *Statistical Papers*, vol. 52, no. 4, pp. 893-909, 2011.

[15] R. P. W. Duin and M. Loog, "Linear dimensionality reduction via a heteroscedastic extension of LDA: the Chernoff criterion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 732-739, 2004, doi: 10.1109/TPAMI.2004.13.

[16] C. Tao and J. Feng, "Canonical kernel dimension reduction," *Computational Statistics Data Analysis*, vol. 107, pp. 131-148, 2017.

[17] Y. Gao, S. Luo, J. Pan, Z. Wang, and P. Gao, "Kernel Alignment Unsupervised Discriminative Dimensionality Reduction," *Neurocomputing*, 2021.

[18] W. Yan, Q. Sun, H. Sun, Y. Li, and Z. Ren, "Multiple kernel dimensionality reduction based on linear regression virtual reconstruction for image set classification," *Neurocomputing*, vol. 361, pp. 256-269, 2019.

[19] D. Zabihzadeh, R. Monsefi, and H. S. Yazdi, "Sparse Bayesian approach for metric learning in latent space," *Knowledge-Based Systems*, vol. 178, pp. 11-24, 2019.

[20] S. A. R. Al-Obaidi, D. Zabihzadeh, H. J. I. J. o. M. L. Hajiabadi, and Cybernetics, "Robust metric learning based on the rescaled hinge loss," vol. 11, pp. 2515-2528, 2020.

[21] S. Bell and K. J. A. t. o. g. Bala, "Learning visual similarity for product design with convolutional neural networks," vol. 34, no. 4, pp. 1-10, 2015.

[22] H. Oh Song, Y. Xiang, S. Jegelka, and S. Savarese, "Deep metric learning via lifted structured feature embedding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4004-4012.

[23] H.-J. Ye, D.-C. Zhan, X.-M. Si, and Y. Jiang, "Learning feature aware metric," in *Asian Conference on Machine Learning*, 2016: PMLR, pp. 286-301.

[24] S. Horiguchi, D. Ikami, and K. Aizawa, "Significance of Softmax-Based Features in Comparison to Distance Metric Learning-Based Features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 5, pp. 1279-1285, 2020, doi: 10.1109/TPAMI.2019.2911075.

[25] J. Hamidzadeh, R. Monsefi, and H. S. Yazdi, "DDC: distance-based decision classifier," *Neural Computing and Applications*, vol. 21, no. 7, pp. 1697-1707, 2012.

[26] Z. Shi and J. Hu, "Local linear discriminant analysis with composite kernel for face recognition," in *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 2012: IEEE, pp. 1-5.

[27] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Mullers, "Fisher discriminant analysis with kernels," in *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)*, 1999: Ieee, pp. 41-48.

[28] A. Zollanvari, J. Hua, and E. R. Dougherty, "Analytical study of performance of linear discriminant analysis in stochastic settings," *Pattern Recognition*, vol. 46, no. 11, pp. 3017-3029, 2013, doi: https://doi.org/10.1016/j.patcog.2013.04.002.

[29] Y. Guo, Y. Zhou, and Z. Zhang, "Fault diagnosis of multi-channel data by the CNN with the multilinear principal component analysis," *Measurement*, vol. 171, pp. 108513, 2021.

[DOI: 10.52547/jsdp.19.3.105]

Engineering, vol. 18, no. 10, pp. 1312-1322, 2006.

[42] S. Sun, X. Xie, and M. Yang, "Multiview Uncorrelated Discriminant Analysis," *IEEE Transactions on Cybernetics*, vol. 46, no. 12, pp. 3272-3284, 2016, doi: 10.1109/TCYB.2015.2502248.

[43] H. Nakanishi and Y. Sato, "The performance of the linear and quadratic discriminant functions for three types of non-normal distribution," *Communications in Statistics - Theory and Methods*, vol. 14, no. 5, pp. 1181-1200, 1985, doi: 10.1080/03610928508828970.

[44] S. Bose, A. Pal, R. SahaRay, and J. Nayak, "Generalized quadratic discriminant analysis," *Pattern Recognition*, vol. 48, no. 8, pp. 2676-2684, 2015, doi: https://doi.org/10.1016/j.patcog.2015.02.016.

[45] A. Ghosh, R. SahaRay, S. Chakrabarty, and S. Bhadra, "Robust Generalised Quadratic Discriminant Analysis," *Pattern Recognition*, p. 107981, 2021.

[46] R. Wang, S. Shan, X. Chen, and W. Gao, "Manifold-manifold distance with application to face recognition based on image set," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008: IEEE, pp. 1-8.

[47] Z. Huang, R. Wang, S. Shan, and X. Chen, "Projection metric learning on Grassmann manifold with application to video based face recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 140-149.

[48] J. S. Abbasi, F. Bashir, K. N. Qureshi, M. Najam ul Islam, and G. Jeon, "Deep learning-based feature extraction and optimizing pattern matching for intrusion detection using finite state machine," *Computers & Electrical Engineering*, vol. 92, p. 107094, 2021, doi: https://doi.org/10.1016/j.compeleceng.2021.107094.

[49] G. Li, L. Yu, and S. Fei, "A deep-learning real-time visual SLAM system based on multi-task feature extraction network and self-supervised feature points," *Measurement*, vol. 168, p. 108403, 2021, doi: https://doi.org/10.1016/j.measurement.2020.108403.

[50] C. Huang, Y. Zong, Y. Ding, X. Luo, K. Clawson, and Y. Peng, "A new deep learning approach for the retinal hard exudates detection based on superpixel multi-feature extraction and patch-based CNN," Neurocomputing, 2020, doi: https://doi.org/10.1016/j.neucom.2020.07.145.

[30] S. Du, Y. Shi, G. Shan, W. Wang, and Y. Ma, "Tensor low-rank sparse representation for tensor subspace learning," *Neurocomputing*, vol. 440, pp. 351-364, 2021.

[31] J. Liu, C. Zhu, Z. Long, H. Huang, and Y. Liu, "Low-rank tensor ring learning for multi-linear regression," *Pattern Recognition*, vol. 113, p. 107753, 2021.

[32] W. Yin, Z. Ma, and Q. Liu, "Riemannian-based Discriminant Analysis for Feature Extraction and Classification," *arXiv preprint arXiv:.08032*, 2021.

[33] Q. Wang, Z. Qin, F. Nie, and Y. Yuan, "Convolutional 2D LDA for Nonlinear Dimensionality Reduction," in *IJCAI*, 2017, pp. 2929-2935.

[34] X. Xiao, Y. Chen, Y.-J. Gong, and Y. Zhou, "2D quaternion sparse discriminant analysis," *IEEE Transactions on Image Processing*, vol. 29, pp. 2271-2286, 2019.

[35] C.-N. Li, Y.-H. Shao, W.-J. Chen, Z. Wang, and N.-Y. Deng, "Generalized two-dimensional linear discriminant analysis with regularization," *Neural Networks*, vol. 142, pp. 73-91, 2021, doi: https://doi.org/10.1016/j.neunet.2021.04.030.

[36] I. Czarnowski, "Cluster-based instance selection for machine classification," *Knowledge Information Systems*, vol. 30, no. 1, pp. 113-133, 2012.

[37] H.-D. Cheng, J. Shan, W. Ju, Y. Guo, and L. Zhang, "Automated breast cancer detection and classification using ultrasound images: A survey," *Pattern recognition*, vol. 43, no. 1, pp. 299-317, 2010.

[38] M. Yang and S. Sun, "Multi-view uncorrelated linear discriminant analysis with applications to handwritten digit recognition," in *2014 International Joint Conference on Neural Networks (IJCNN)*, 2014: IEEE, pp. 4175-4181.

[39] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *2012 IEEE conference on computer vision and pattern recognition*, 2012: IEEE, pp. 2160-2167.

[40] Z. Jin, J.-Y. Yang, Z.-S. Hu, and Z. Lou, "Face recognition based on the uncorrelated discriminant transformation," *Pattern recognition*, vol. 34, no. 7, pp. 1405-1416, 2001.

[41] J. Ye, R. Janardan, Q. Li, and H. Park, "Feature reduction via generalized uncorrelated linear discriminant analysis," *IEEE Transactions on Knowledge Data*

[DOI: 10.52547/jsdp.19.3.105]

جواد حمید زاده در حال حاضر دانشیار دانشکده مهندسی فن‌آوری اطلاعات و ارتباطات دانشگاه سجاد است. از علاقه‌مندی‌های ایشان می‌توان به یادگیری ماشین، محاسبات نرم و بازشناسی الگو اشاره کرد.
نشانی رایانامه ایشان عبارت است از:
J_hamidzadeh@sadjad.ac.ir

منا مرادی دارای مدرک کارشناسی ارشد مهندسی کامپیوتر در گرایش نرم‌افزار است. از علاقه‌مندی‌های ایشان می‌توان به یادگیری ماشین، محاسبات نرم و بازشناسی الگو اشاره کرد.
نشانی رایانامه ایشان عبارت است از:
MonaMoradi0@gmail.com

[DOI: 10.52547/jsdp.19.3.105]