

تحلیل ممیز غیر پارامتریک بهبودیافته برای دسته‌بندی

تصاویر ابرطیفی با نمونه آموزشی محدود

آزاده کیانی سرکله^۱ و محمد حسن قاسمیان^{۲*}^۱ دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، گروه مهندسی برق، تهران، ایران^۲ دانشگاه تربیت مدرس، دانشکده مهندسی برق و کامپیوتر، تهران، ایران

چکیده

استخراج ویژگی نقش مهمی در بهبود دسته‌بندی تصاویر ابرطیفی دارد. روش‌های استخراج ویژگی غیرپارامتریک، نسبت به روش‌های پارامتریک، برای داده‌های با توزیع غیر نرمال کارایی بهتری دارند و می‌توانند ویژگی‌های بیشتری را استخراج کنند. روش‌های استخراج ویژگی غیرپارامتریک از ماتریس‌های پراکندگی غیرپارامتریک برای محاسبه ماتریس انتقال استفاده می‌کنند. تحلیل ممیز غیرپارامتریک^۱ یکی از روش‌های غیرپارامتریک در استخراج ویژگی است که در آن برای تشکیل ماتریس‌های پراکندگی غیرپارامتریک، از میانگین‌های محلی هر نمونه و تابع وزن استفاده می‌شود. میانگین محلی با استفاده از k نمونه همسایه به دست می‌آید و تابع وزن، بر روی نمونه‌های مرزی در تشکیل ماتریس پراکندگی بین‌دسته‌ای تأکید می‌کند. در این مقاله، NDA بهبودیافته^۲ به منظور اصلاح NDA معرفی شده است. در MNDA، تعداد نمونه‌های همسایه در محاسبه میانگین محلی با توجه به موقعیت نمونه در فضای ویژگی به دست می‌آید. روش پیشنهادی از توابع وزن جدید در تشکیل ماتریس‌های پراکندگی استفاده می‌کند. توابع وزن پیشنهادی تأکید روی نمونه‌های مرزی در تشکیل ماتریس پراکندگی بین‌دسته‌ای و تأکید روی نمونه‌های نزدیک به میانگین دسته، در تشکیل ماتریس پراکندگی درون دسته‌ای دارند. علاوه بر این، به منظور اجتناب از تکین شدن ماتریس پراکندگی درون دسته‌ای، از تنظیم آن استفاده شده است. نتایج آزمایش‌ها روی تصاویر ایندیانا و سالیاناس نشان می‌دهد که MNDA کارایی بهتری نسبت به روش‌های استخراج ویژگی پارامتریک و غیرپارامتریک مورد مقایسه داشته است. بیشترین مقدار صحت متوسط دسته‌بندی برای داده ایندیانا ۸۰/۳۴٪ است که با ۱۸ نمونه آموزشی، دسته‌بند ماشین بردار پشتیبان و ۱۰ ویژگی استخراج شده از MNDA به دست آمده است. برای داده سالیاناس، بیشترین مقدار صحت متوسط دسته‌بندی، ۹۴/۳۱٪ است که با ۱۸ نمونه آموزشی، دسته‌بند ماشین بردار پشتیبان و ۹ ویژگی استخراج شده از MNDA به دست آمده است. آزمایش‌ها نشان می‌دهند که با استفاده از توابع وزن پیشنهادی و ماتریس پراکندگی درون دسته‌ای تنظیم شده، روش پیشنهادی نتایج بهتری را در دسته‌بندی تصاویر ابرطیفی با نمونه‌های آموزشی محدود به دست آورده است.

واژگان کلیدی: تصاویر ابرطیفی، استخراج ویژگی، دسته‌بندی نظارت‌شده، پدیده هیوز، نمونه‌های آموزشی محدود.

Modified Nonparametric Discriminant Analysis for Classification of Hyperspectral Images with Limited Training Samples

Azadeh Kianisarkaleh¹ & Mohammad Hassan Ghassemian^{2*}¹Department of Electrical Engineering, Science and Research Branch, Azad Islamic University, Tehran, Iran.²Faculty of Electrical and Computer Engineering, University of Tarbiat Modarres, Tehran, Iran¹ Nonparametric Discriminant Analysis (NDA)² Modified NDA (MNDA)

* Corresponding author

* نویسنده عهده‌دار مکاتبات

Abstract

Feature extraction performs an important role in improving hyperspectral image classification. Compared with parametric methods, nonparametric feature extraction methods have better performance when classes have no normal distribution. Besides, these methods can extract more features than what parametric feature extraction methods do. Nonparametric feature extraction methods use nonparametric scatter matrices to compute transformation matrix. Nonparametric Discriminant Analysis (NDA) is one of the nonparametric feature extraction methods in which, to form nonparametric scatter matrices, local means of samples and weight function are used. Local mean is calculated by k nearest neighbors of each sample and weight function emphasizes on boundary samples in between class scatter matrix formation. In this paper, modified NDA (MNDA) is proposed to improve NDA. In MNDA, the number of neighboring samples, when measuring local mean, are determined considering position of each sample in feature space. MNDA uses new weight functions in scatter matrix formation. Suggested weight functions emphasizes on boundary samples in between class scatter matrix formation and focus on samples close to class mean in within class scatter matrix formation. Moreover, within class scatter matrix is regularized to avoid singularity. Experimental results on Indian Pines and Salinas images show that MNDA has better performance compared to other parametric, nonparametric feature extraction methods. For Indian Pines data set, the maximum average classification accuracy is 80.34%, which is obtained by 18 training samples, support vector machine (SVM) classifier and 10 extracted features achieved by MNDA method. For Salinas data set, the maximum average classification accuracy is 94.31%, which is obtained by 18 training samples, SVM classifier and 9 extracted features achieved by MNDA method. Experiments show that using suggested weight functions and regularized within class scatter matrix, the proposed method obtained better results in hyperspectral image classification with limited training samples.

Keywords: Hyperspectral images, Feature extraction, Supervised classification, Hughes Phenomenon, Limited training samples.

روش‌های نیمه‌نظارتی، ترکیبی از روش‌های نظارت‌شده و نظارت‌نشده هستند و از خاصیت‌های آنها برخوردارند. در این روش‌ها، از نمونه‌های برچسب‌دار و بدون برچسب استفاده می‌شود [6]–[7]. در روش‌های نظارت‌نشده مانند تحلیل مؤلفه‌های اساسی^۵ [8] و افکندش با نگهداری موقعیت محلی [9]، از تمامی نمونه‌ها استفاده می‌شود. این روش‌ها، ساختار کلی داده‌ها را حفظ کرده و روی اطلاعات تمایزی نمونه‌ها تمرکز ندارند؛ بنابراین ممکن است در کاربردهای دسته‌بندی کارایی مناسبی نداشته باشند.

در روش‌های نظارت‌شده از نمونه‌های برچسب‌دار استفاده می‌شود. در این روش‌ها، تمایز دسته‌ها با استفاده از نمونه‌های برچسب‌دار در نظر گرفته شده و می‌توانند برای کاربردهای دسته‌بندی مناسب باشند [10]. در این مقاله، روش‌های استخراج ویژگی نظارت‌شده مورد توجه قرار گرفته‌اند. روش‌های استخراج ویژگی نظارت‌شده را می‌توان به دو دسته روش‌های پارامتریک و غیرپارامتریک تقسیم‌بندی کرد. در روش‌های پارامتریک، از اطلاعات کلی نمونه‌های آموزشی مانند پارامترهای میانگین و کواریانس داده‌ها در فرآیند استخراج ویژگی استفاده می‌شود. در روش‌های غیرپارامتریک، اطلاعات محلی هر نمونه آموزشی

۱- مقدمه

یکی از کاربردهای تصاویر سنجش از دور، شناسایی و تشخیص اجسام یا پدیده‌ها بدون تماس مستقیم با آنها، است. با پیشرفت سنجش‌گرها، دقت طیفی تصاویر سنجش از دور به چندصد باند در تصاویر ابرطیفی افزایش یافته است. بنابراین، طیف اندازه‌گیری‌شده به طیف واقعی پیکسل نزدیکتر بوده و می‌توان عمل شناسایی را با دقت بالاتری انجام داد؛ ولی از آنجایی که تهیه نمونه‌های آموزشی قابل اطمینان بسیار سخت و هزینه‌بر و گاهی اوقات غیرعملی است، در دسته‌بندی بانظارت تصاویر ابرطیفی با پدیده هیوز^۱ مواجه خواهیم شد [1]. برای غلبه بر پدیده هیوز در دسته‌بندی داده‌های با ابعاد بالا، روش‌هایی مورد استفاده قرار می‌گیرد. یکی از این روش‌ها، استفاده از الگوریتم‌های استخراج ویژگی است [2]–[5]. در استخراج ویژگی، از یک تبدیل یا نگاشت برای انتقال تمامی پیکسل‌ها در تصویر ابرطیفی از فضای با بعد d به زیرفضایی با بعد r ($r \leq d$) استفاده می‌شود.

الگوریتم‌های استخراج ویژگی را می‌توان به سه دسته نظارت‌شده^۲، نظارت‌نشده^۳ و نیمه‌نظارتی^۴ تقسیم کرد.

¹ Hughes Phenomenon

² Supervised

³ Unsupervised

⁴ Semi supervised

⁵ Principal Component Analysis (PCA)

ماتریس‌های پراکندگی غیرپارامتریک، تعدادی از معایب روش LDA را پوشش داده است؛ ولی همچنان از مشکلاتی رنج می‌برد. این مشکلات عبارتند از: ۱. بهترین مقدار برای تعداد نمونه‌های همسایه در محاسبه میانگین محلی، با انجام آزمایش به دست می‌آید. ۲. با استفاده از تابع وزن در NDA، وزن نمونه‌های مرزی و تأثیر آنها در تشکیل S_b کاهش پیدا می‌کند. ۳. ماتریس S_w در روش NDA با روش LDA در مواردی که دسته‌ها درهم‌فرورفتگی داشته باشند، مشابه است. بنابراین هنوز مشکل تکین شدن آن وجود خواهد داشت [12].

استخراج ویژگی غیرپارامتریک وزن‌دار^۷ برای غلبه بر مشکلات LDA و NDA پیشنهاد شده است [11]. ایده اصلی این روش، تشکیل میانگین محلی وزن‌دار با استفاده از تمامی نمونه‌های برچسب‌دار است. بنابراین تعداد نمونه‌ها در محاسبه میانگین محلی، مشخص است؛ همچنین تابع وزن پیشنهادی در محاسبه ماتریس‌های پراکندگی، پارامتر آزاد ندارد و با استفاده از آن، نمونه‌های مرزی وزن بیشتری در تشکیل S_b دارند. در این روش، از فن تنظیم برای رفع مشکل تکین شدن S_w استفاده شده است. این روش با وجود مزیت‌های ذکر شده، زمان زیادی را نسبت به LDA و NDA برای انجام محاسبات صرف می‌کند. این مسأله می‌تواند در مواردی که نمونه‌های برچسب‌دار افزایش می‌یابند، قابل توجه باشد.

در استخراج ویژگی غیرپارامتریک بر اساس فاصله کسینوسی^۸، از فاصله کسینوسی برای توسعه تابع وزن استفاده شده است [12]. با استفاده از این فاصله، CNFE از اندازه اطلاعات محلی در اطلاعات کلی، برای محاسبه وزن نمونه استفاده می‌کند. در این روش، بهترین مقدار برای تعداد نمونه‌های همسایه در محاسبه میانگین محلی، با انجام آزمایش به دست می‌آید. تحلیل ممیز و نخستین مؤلفه اساسی^۹، روشی غیرپارامتریک بر مبنای تحلیل ممیز است [13]. در این روش، از نخستین مؤلفه اساسی برای محاسبه وزن نمونه‌ها استفاده می‌شود. بنابراین، علاوه بر استفاده از اطلاعات تمایزی دسته‌ها، از اطلاعات بازنمایی داده‌ها برای حفظ ساختار غالب آنها، استفاده می‌شود. در هر دو روش CNFE و DA-PC1، از فن تنظیم برای رفع مشکل تکین شدن S_w استفاده شده است.

در فرآیند استخراج ویژگی مورد استفاده قرار می‌گیرد. تحلیل ممیز خطی^۱، یکی از روش‌های نظارت‌شده پارامتریک در استخراج ویژگی است. در LDA، از بردار میانگین و ماتریس کواریانس هر دسته استفاده می‌شود. در این روش، ویژگی‌های استخراج‌شده با در نظر گرفتن بیشینه تمایز بین دسته‌ها به دست می‌آیند و می‌توانند صحت دسته‌بندی را افزایش دهند [8]. البته استفاده از این روش، محدودیت‌های زیر را به همراه دارد که روی کارایی آن تأثیرگذارند. ۱. در مواردی که توزیع داده‌ها غیر نرمال یا ترکیبی از چند توزیع باشد، ویژگی‌های به دست آمده از LDA رضایت‌بخش نیست. ۲. رتبه ماتریس پراکندگی بین دسته‌ای^۲ برابر با تعداد دسته‌ها منهای یک است؛ بنابراین ویژگی‌های استخراج‌شده، ممکن است برای بازنمایی داده‌های واقعی کافی نباشند. ۳. مشکل تکین شدن ماتریس پراکندگی درون دسته‌ای^۳ در مسائل با ابعاد بالا وجود دارد. ماتریس S_w از مجموع ماتریس‌های کواریانس داده‌های آموزشی در هر دسته محاسبه می‌شود. اگر N_i تعداد نمونه‌های آموزشی در هر دسته و d تعداد ویژگی‌ها باشد، ماتریس کواریانس در هر دسته از جمع N_i ماتریس $d \times d$ با رتبه یک، به دست می‌آید. بنابراین در صورتی ماتریس کواریانس تکین نخواهد بود که $N_i > d$ باشد. محدودیت نمونه‌های آموزشی، باعث تکین شدن ماتریس کواریانس و S_w در مسائل با ابعاد می‌شود؛ بنابراین ویژگی‌های به دست آمده از LDA در شرایط نمونه آموزشی محدود، کارایی قابل قبولی در دسته‌بندی داده‌ها نخواهند داشت [11].

برای غلبه بر این محدودیت‌ها، روش‌های پارامتریک غیرپارامتریک فراوانی بر مبنای LDA توسعه یافته‌اند. یکی از این روش‌ها، تحلیل ممیز غیرپارامتریک^۴ است [8]. در این روش، ماتریس‌های پراکندگی به صورت غیرپارامتریک و با استفاده از میانگین‌های محلی^۵ هر نمونه محاسبه می‌شوند. ماتریس‌های پراکندگی غیرپارامتریک مزیت‌هایی شامل استفاده از اطلاعات نمونه‌های همسایه، به دست آمدن رتبه بزرگتر از تعداد دسته‌ها منهای یک برای S_b و کاهش تأثیر نمونه‌های پرت^۶ را دارند و برای داده‌هایی که توزیع غیرنرمال دارند، مناسب می‌باشند. روش NDA با استفاده از

¹ Linear Discriminant Analysis (LDA)

² Between classes Scatter Matrix (S_b)

³ Within classes Scatter Matrix (S_w)

⁴ Nonparametric Discriminant Analysis (NDA)

⁵ Local mean

⁶ Outlier

⁷ Nonparametric Weighted Feature Extraction (NWFE)

⁸ Cosine-based Nonparametric Feature Extraction (CNFE)

⁹ Discriminant Analysis and the First Principal Component (DA-PC1)

روش پیشنهادی با افزایش تراکم درون‌دسته‌ای و فاصله بین‌دسته‌ای، تمایز دسته‌ها را افزایش می‌دهد. به‌منظور ارزیابی کارایی روش پیشنهادی، آزمایش‌ها روی دادگان واقعی ابرطیفی ایندیانا^۴ و سالیناس^۵ انجام شده است. نتایج نشان می‌دهند که روش پیشنهادی با استفاده از توابع وزن پیشنهادی و تعیین تعداد همسایگی‌ها به کمک الگوریتم ارائه‌شده، کارایی قابل توجهی در استخراج ویژگی‌های تمایزی داشته و باعث افزایش صحت دسته‌بندی در شرایط نمونه آموزشی محدود، شده است.

ادامه مقاله به شکل زیر سازمان یافته است: در بخش بعدی، روش استخراج ویژگی NDA معرفی می‌شود؛ سپس روش پیشنهادی معرفی خواهد شد. نتایج به دست آمده، در بخش چهارم و نتیجه‌گیری در بخش آخر ارائه خواهد شد.

۲- تحلیل ممیز غیرپارامتریک

روش استخراج ویژگی NDA، یک روش نظارت‌شده خطی است و از میانگین محلی نمونه‌ها در تشکیل ماتریس‌های پراکندگی غیرپارامتریک استفاده می‌کند [8]. با انتخاب تعداد نمونه‌های همسایه، میانگین محلی هر نمونه در دسته z ام، طبق رابطه (۱) محاسبه می‌شود [8].

$$M_j(x_l^{(i)}) = \frac{1}{k} \sum_{s=1}^k x_{sNN}^{(j)} \quad (1)$$

در این رابطه، $M_j(x_l^{(i)})$ میانگین محلی l امین نمونه از دسته i ام در دسته z ام، $x_{sNN}^{(j)}$ s امین نزدیک‌ترین همسایه به $x_l^{(i)}$ در دسته z ام و k تعداد نمونه‌های همسایه را مشخص می‌کند. با محاسبه میانگین محلی هر نمونه در تمامی دسته‌ها، ماتریس‌های پراکندگی با استفاده از روابط (۲) و (۳) محاسبه می‌شوند [8].

$$S_b^{NDA} = \sum_{i=1}^q P_i \sum_{j=1}^q \sum_{l=1}^{N_i} \frac{W_l^{(i,j)}}{N_i} \quad (2)$$

$$((x_l^{(i)} - M_j(x_l^{(i)}))(x_l^{(i)} - M_j(x_l^{(i)})))^T$$

$$S_w^{NDA} = \sum_{i=1}^q P_i \sum_{l=1}^{N_i} \frac{W_l^{(i,i)}}{N_i} \quad (3)$$

$$((x_l^{(i)} - M_i(x_l^{(i)}))(x_l^{(i)} - M_i(x_l^{(i)})))^T$$

در این روابط، P_i احتمال پیشین دسته i ام، N_i تعداد نمونه‌های برجسب‌دار در دسته i ام، W_l وزن نمونه l ام و q

علاوه‌بر روش‌های غیرپارامتریک ذکرشده، روش‌های دیگری نیز برای رفع مشکلات LDA معرفی شده‌اند. روش استخراج ویژگی با استفاده از نقطه جذب^۱ [14] و تحلیل ممیز فضای ویژگی^۲ [15] از نمونه این روش‌ها هستند. در FEAP، یک نقطه جذب برای هر دسته انتخاب می‌شود. با انتخاب نقطه جذب، نمونه‌های هر دسته به سمت نقطه جذب دسته‌شان، جذب شده و از طرف نقطه جذب دسته‌های دیگر دفع می‌شوند. در نتیجه، نمونه‌های دسته‌های مختلف در فضای انتقال یافته متمایز خواهند شد. در FSDA، علاوه‌بر تمایز دسته‌ها که با استفاده از ماتریس‌های پراکندگی درون‌دسته‌ای و بین‌دسته‌ای به دست می‌آید، تفاوت بین باندهای طیفی نیز در نظر گرفته می‌شود. ویژگی‌ها در FSDA به شکلی استخراج می‌شوند که بیشینه تفاوت را نسبت به هم داشته باشند و تمایز بین دسته‌ها افزایش یابد.

در این مقاله، به‌منظور رفع مشکلات روش NDA، روش تحلیل ممیز غیرپارامتریک بهبودیافته^۳ ارائه شده است. در این روش، تعداد همسایگی‌ها با توجه به موقعیت نمونه در فضای ویژگی تعیین می‌شود. همچنین با استفاده از توابع وزن جدید، تأثیر نمونه‌های مرزی در تشکیل S_b و تأثیر نمونه‌های نزدیک به میانگین دسته در تشکیل S_w افزایش یافته است. برای جلوگیری از تکین شدن ماتریس پراکندگی درون‌دسته‌ای، از فن تنظیم استفاده شده است. در روش‌های غیرپارامتریک، ماتریس‌های پراکندگی از مجموع وزن‌دار پراکندگی نمونه‌های آموزشی در اطراف میانگین‌های محلی آنها (پراکندگی‌های محلی) محاسبه می‌شود. در روش پیشنهادی، برای محاسبه پراکندگی درون‌دسته‌ای، وزن بیشتری به نمونه‌های نزدیک به میانگین دسته اختصاص می‌یابد. بنابراین تأکید بیشتری روی پراکندگی محلی نمونه‌های آموزشی نزدیک به مرکز تجمع نمونه‌های هر دسته شده است و با کمینه کردن این ماتریس، تراکم درون دسته‌ای در فضای انتقال یافته افزایش خواهد یافت. همچنین برای محاسبه ماتریس پراکندگی بین‌دسته‌ای، وزن بیشتری به نمونه‌های مرزی اختصاص می‌یابد؛ بنابراین تأکید بیشتری روی پراکندگی محلی نمونه‌های آموزشی نزدیک به مرز شده است و با بیشینه کردن این ماتریس، فاصله بین دسته‌ها در فضای انتقال یافته افزایش می‌یابد. در نتیجه،

¹ Feature Extraction using attraction points (FEAP)

² Feature space discriminant analysis (FSDA)

³ Modified Nonparametric Discriminant Analysis (MNDA)

⁴ Indian Pines

⁵ Salinas

روش NDA و راه‌حل‌های ارائه‌شده در روش MNDA. به شرح زیر است:

۱. میانگین محلی نمونه $x_i^{(i)}$ در NDA با استفاده از k نمونه همسایه محاسبه می‌شود. مقدار پارامتر k با استفاده از قانون خاصی به دست نمی‌آید و بهترین مقدار با انجام آزمایش محاسبه می‌شود. پس از محاسبه مقدار بهینه برای k ، این مقدار در تمامی محاسبات ثابت در نظر گرفته می‌شود. در MNDA، تعداد همسایگی‌ها با توجه به موقعیت نمونه در فضای ویژگی مشخص می‌شود. از آنجایی که تعداد همسایگی‌های درون‌دسته‌ای و بین‌دسته‌ای برای یک نمونه، در نتیجه به دست آمده تأثیرگذار است، تعداد همسایگی‌ها را به صورت متغیر می‌توان در نظر گرفت و با توجه به فشردگی یا تنگی داده‌ها، تعداد همسایگی‌ها را انتخاب کرد. ناحیه‌ای که در آن بیشترین تراکم برای نمونه‌های یک دسته وجود دارد و می‌تواند ساختار هندسی محلی داده‌های آن دسته را مشخص کند، تعداد بیشتری برای همسایگی از همان دسته در نظر گرفته شود. در نزدیکی مرز یک دسته که نمونه‌های دسته‌های مختلف می‌توانند ساختار مرز بین دسته‌ها را مشخص کنند، تعداد بیشتری برای همسایگی از دسته‌های دیگر در نظر گرفته شود. الگوریتم پیشنهادشده برای تعیین تعداد همسایگی‌ها در محاسبه میانگین محلی $x_i^{(i)}$ به شرح زیر است:

الف) محاسبه میانگین دسته‌ها با استفاده از نمونه‌های برچسب‌دار انتخاب‌شده برای هر دسته. ب) محاسبه فاصله اقلیدسی نمونه $x_i^{(i)}$ از میانگین هر دسته و مرتب‌کردن دسته‌ها با توجه به نزدیکی $x_i^{(i)}$ به میانگین دسته‌ها. پ) قراردادن دسته‌های مرتب‌شده در گروه‌های سه‌تایی و اختصاص برچسب نزدیکی Z ($Z = 1, 2, 3, \dots$) به هر گروه. به‌عنوان مثال سه‌دسته‌ای که نسبت به سایر دسته‌ها در فاصله کمتری از نمونه i قرار دارند، عضو گروه نخست بوده و برچسب $Z = 1$ را به خود اختصاص می‌دهند. ت) محاسبه k با استفاده از رابطه (۷).

$$k = N_i - (2Z - 1) \quad (7)$$

با توجه به الگوریتم پیشنهادی، مقدار k با تغییر در تعداد نمونه‌های آموزشی تغییر خواهد کرد. همچنین برای دسته‌های نزدیک‌تر به نمونه $x_i^{(i)}$ تعداد بیشتری از نمونه‌های همسایه در تشکیل میانگین محلی استفاده می‌شود. بنابراین میانگین محلی در دسته‌های نزدیک‌تر به نمونه $x_i^{(i)}$ با دقت بیشتری تخمین زده خواهد شد.

تعداد دسته‌ها هستند. رابطه (۲)، نشان‌دهنده نقش مهم فاصله هر نمونه از میانگین محلی‌اش در دسته دیگر ($x_i^{(i)}$) برای محاسبه S_b^{NDA} است. این بردار، جهتی به سمت دسته Z را نشان می‌دهد و می‌تواند نماینده ناحیه مرزی باشد. برای نمونه‌هایی که دور از مرز قرار دارند، بردار تشکیل‌شده دامنه بزرگ‌تری خواهد داشت و اطلاعات ساختاری مرز را می‌تواند منحرف کند. بنابراین از تابع وزن در تشکیل ماتریس‌های پراکندگی استفاده شده است. این تابع، طبق رابطه (۴) تعریف شده است [8].

$$W_i^{(i,j)} = \frac{\min\{d^\alpha(x_L^{(i)}, x_{kNN}^{(i)}), d^\alpha(x_L^{(i)}, x_{kNN}^{(j)})\}}{d^\alpha(x_L^{(i)}, x_{kNN}^{(i)}) + d^\alpha(x_L^{(i)}, x_{kNN}^{(j)})} \quad (4)$$

در این رابطه، α پارامتر کنترل و بین صفر و یک مقدار محدود است و $d(x_L^{(i)}, x_{kNN}^{(j)})$ فاصله اقلیدسی نمونه $x_i^{(i)}$ از k امین نزدیک‌ترین همسایه در دسته Z است. با استفاده از تابع وزن، نمونه‌های دور از مرز وزن کمتری را به دست خواهند آورد و تأثیر آنها در تشکیل S_b^{NDA} کاهش می‌یابد. با محاسبه ماتریس‌های پراکندگی، ماتریس انتقال (T) با استفاده از رابطه (۵) و حل مسأله مقدار ویژه توسعه‌یافته به دست می‌آید. ماتریس انتقال شامل بردارهای ویژه توسعه‌یافته (t) متناظر با r بزرگ‌ترین مقادیر ویژه در رابطه (۶) است [8].

$$T = \arg \max_{T \in \mathbb{R}^{d \times r}} \frac{T^* S_b^{NDA} T}{T^* S_w^{NDA} T} \quad (5)$$

$$S_b^{NDA} t = \lambda S_w^{NDA} t \quad (6)$$

در رابطه (۵)، علامت * ترانزپوز ماتریس را نشان می‌دهد.

۳- تحلیل ممیز غیر پارامتریک بهبودیافته

استفاده از ماتریس‌های پراکندگی غیر پارامتریک در NDA، تعدادی از معایب LDA را پوشش داده است؛ ولی این روش نیز محدودیت‌هایی را به همراه دارد. در این مقاله، برای رفع مشکلات NDA، راه‌حل‌هایی پیشنهاد شده است. روش پیشنهادی بر پایه NDA است و از میانگین محلی داده‌ها در تشکیل ماتریس‌های پراکندگی استفاده می‌کند. مشکلات

۴. با توجه به رابطه (۴)، مقدار وزن اختصاص یافته به نمونه $x_i^{(i)}$ که در نزدیکی دسته z قرار دارد، با کم شدن فاصله، کوچک خواهد بود و تأثیر این نمونه مرزی را در تشکیل S_b^{NDA} کم رنگ خواهد کرد. با پیشنهاد تابع وزن طبق رابطه (۹)، نمونه های مرزی در MNDA، از وزن بیشتری برخوردار خواهند شد. اگر نمونه $x_i^{(i)}$ به مرز دسته z نزدیک باشد، فاصله $M_i(x_i^{(i)})$ و $M_j(x_i^{(i)})$ نسبت به نمونه ای که دور از مرز قرار دارد، کمتر خواهد بود و وزن بیشتری خواهد داشت.

$$W_i^{(i,j)} = \frac{d(x_i^{(i)}, m_j)}{d(x_i^{(i)}, m_j) + d(M_i(x_i^{(i)}), M_j(x_i^{(i)}))} \quad (9)$$

۵. در شرایط نمونه آموزشی محدود و زمانی که دسته ها در هم فرورفتگی داشته باشند، مشکل تکین شدن S_W^{NDA} وجود خواهد داشت. برای رفع مشکل ذکر شده و به منظور کامل کردن رتبه ماتریس پراکندگی در شرایط نمونه آموزشی محدود، از تنظیم آن طبق رابطه زیر استفاده شده است.

$$S_W^{MNDA} = 0.5S_W^{NDA} + 0.5diag(S_W^{MNDA}) \quad (10)$$

در این رابطه، $diag(A)$ ماتریس قطری است که شامل عناصر موجود بر روی قطر اصلی A است.



(شکل-۱): روندنمای روش پیشنهادی
(Figure-1): Flowchart of the proposed method

۴- نتایج آزمایش ها

برای ارزیابی روش پیشنهادی، آزمایش ها روی دادگان واقعی ابرطیفی ایندیانا و سالیانس انجام شده است. دادگان ایندیانا مربوط به یک منطقه کشاورزی- جنگلی است که توسط سنجنده AVIRIS¹ از یک منطقه آزمایشی در شمال شرقی ایالت ایندیانا گرفته شده است. این دادگان شامل ۲۲۴ باند طیفی است که بعد از حذف ۲۴ باند نوفه ای، آزمایش ها روی دویست باند باقی مانده انجام می شود. اطلاعات هر باند، به صورت یک تصویر با ابعاد 145×145 قابل دستیابی بوده و شانزده دسته در این دادگان مشخص شده است.

¹ Airborne Visible/Infrared Imaging Spectromete

۲. پارامتر α در رابطه (۴) با استفاده از آزمایش برای رسیدن به بهترین نتیجه مشخص می شود که در روش پیشنهادی پارامتر آزاد در رابطه تابع وزن وجود ندارد.

۳. با توجه به رابطه (۴)، مشخص می شود که وزن تمامی نمونه ها در تشکیل S_W^{NDA} ثابت است ($W_i^{(i,i)} = 0.5$). در صورتی که نمونه های نزدیکتر به مرکز تجمع هر دسته یا میانگین هر دسته سهم بیشتری در محاسبه ماتریس پراکندگی درون دسته ای می توانند داشته باشند. در روش پیشنهادی، با در نظر گرفتن نکته ذکر شده، وزن نمونه $x_i^{(i)}$ در محاسبه ماتریس پراکندگی درون دسته ای طبق رابطه (۸) محاسبه می شود:

$$W_i^{(i,i)} = \frac{d(x_i^{(i)}, m_i)}{d(x_i^{(i)}, m_i) + d(x_i^{(i)}, M_i(x_i^{(i)}))} \quad (8)$$

در این رابطه، m_i میانگین دسته i ام را نشان می دهد. در صورتی که نمونه $x_i^{(i)}$ به m_i نزدیک باشد، فاصله آن از میانگین محلی اش در دسته i ام نسبت به نمونه ای که دورتر از m_i قرار دارد، کمتر خواهد بود؛ زیرا نمونه نزدیک تر به میانگین به مرکز تجمع نمونه ها نزدیک تر بوده و نمونه های همسایه آن در فاصله کمتری قرار دارند.

با توجه به راه حل های ارائه شده، الگوریتم روش پیشنهادی را به شکل زیر می توان خلاصه کرد. همچنین روندنمای روش پیشنهادی در شکل (۱) نشان داده شده است.

- ۱- تعیین مقدار k با استفاده از رابطه (۷) برای هر نمونه.
- ۲- محاسبه میانگین محلی نمونه ها در تمامی دسته ها با استفاده از رابطه (۱).
- ۳- محاسبه وزن نمونه ها با استفاده از رابطه های (۸) و (۹).
- ۴- محاسبه ماتریس های پراکندگی با استفاده از رابطه های (۲) و (۳).
- ۵- استفاده از رابطه (۱۰) برای تنظیم ماتریس پراکندگی.
- ۶- محاسبه ماتریس انتقال با استفاده از رابطه (۶) و جایگذاری ماتریس های پراکندگی در آن.

داده‌گان بعدی، مربوط به دره سالیانس در کالیفرنیا بوده و توسط سنجنده AVIRIS به دست آمده است. این داده‌گان شامل ۲۲۴ باند طیفی است که بعد از حذف بیست باند نوفه‌ای، آزمایش‌ها روی ۲۰۴ باند باقی مانده انجام می‌شود. این داده‌گان شامل ۱۶ دسته بوده و اطلاعات هر باند به صورت تصویری با ابعاد 512×217 در دسترس است.

$$Z_{12} = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \quad (11)$$

به منظور بررسی کارایی روش پیشنهادی در مسائل دسته‌بندی با نمونه‌های آموزشی محدود، از ۱۴ و ۱۸ نمونه برچسب‌دار از هر دسته استفاده شده است. نمونه‌های آموزشی با استفاده از نقشه درستی^۱ داده‌گان، ۱۰ بار و به صورت تصادفی انتخاب شده‌اند و متوسط نتایج به دست آمده از ۱۰ آزمایش گزارش شده است. به منظور مقایسه منصفانه، در هر آزمایش برای تمامی روش‌های استخراج ویژگی، از نمونه‌های برچسب‌دار یکسان استفاده شده است. روش‌های استخراج ویژگی مورد مقایسه، شامل روش پارامتریک LDA و روش‌های غیر پارامتریک NDA،

DA-PC1 و CNFE، NWFE می‌باشند. برای دسته‌بندی داده‌ها، از دسته‌بندی‌های پیشینه احتمال گاوسی^۲ و ماشین بردار پشتیبان^۳ [16] استفاده شده است. دسته‌بندی SVM برای دسته‌بندی داده‌های چنددسته‌ای با استفاده از تابع هسته شعاعی^۴، توسط کتابخانه LIBSVM [17] پیاده‌سازی شده است. پارامتر جریمه با جستجو از میان مجموعه $\{1, 10, \dots, 10000\}$ و بهترین مقدار برای پارامتر پهنای با جستجو از میان مجموعه $\{10^{-8}, \dots, 10^{-1}, 10, 100\}$ با استفاده از اعتبارسنجی تقاطعی پنج مرتبه‌ای^۵ انتخاب شده‌اند. تمامی پیاده‌سازی‌ها در نرم‌افزار MATLAB انجام شده است.

به منظور نشان دادن کارایی روش‌های استخراج ویژگی در دسته‌بندی تصاویر ابرطیفی و مقایسه آنها، از شاخص‌های صحت و اطمینان‌پذیری متوسط استفاده شده است که با استفاده از ماتریس سردرگمی محاسبه می‌شوند [15]. صحت متوسط^۶، میانگین صحت به دست آمده برای دسته‌ها است. صحت دسته‌بندی در دسته i ام برابر است با درصد تعداد نمونه‌هایی که به درستی در دسته i ام دسته‌بندی شده‌اند، نسبت به تعداد کل نمونه‌های آزمون در دسته i ام.

نتایج به دست آمده از دسته‌بندی داده‌گان ایندیانا و نتایج به دست آمده از دسته‌بندی داده‌گان سالیانس، در شکل‌های (۲) و (۳) نشان داده شده است. این شکل‌ها، صحت متوسط دسته‌بندی را برحسب تعداد ویژگی‌های استخراج شده از روش‌های استخراج ویژگی، نشان می‌دهند. برای داده‌گان ایندیانا، MNDA بهترین نتیجه را با استفاده از دسته‌بندی‌های GML و SVM کسب کرده است. برای داده‌گان سالیانس، با افزایش ویژگی‌ها به سه و بیشتر، روش پیشنهادی نتایج بهتری را کسب کرده و نتایج به دست آمده از DA-PC1 بسیار نزدیک به روش پیشنهادی است. این مطلب می‌تواند نشان‌دهنده محاسبه مناسب وزن‌ها در روش پیشنهادی نسبت به سایر روش‌ها باشد؛ زیرا در DA-PC1 و NDA، وزن‌دهی به نمونه‌ها تنها در تشکیل ماتریس درون‌دسته‌ای انجام و در CNFE و NWFE برای وزن‌دهی به نمونه‌ها تنها روی نمونه‌های مرزی تأکید می‌شود. در تمامی نمودارها، نتایج CNFE از NWFE بهتر و از DA-PC1 ضعیف‌تر است. این مطلب می‌تواند نشان‌دهنده انتخاب مناسب وزن‌ها در DA-PC1 نسبت به CNFE و NWFE باشد؛ زیرا در DA-PC1، از اطلاعات بازنمایی نمونه‌ها برای محاسبه وزن استفاده می‌شود. نزدیکی نتایج به دست آمده از NDA و LDA نشان می‌دهد که در مسائل با نمونه آموزشی محدود، NDA قادر به پوشش معایب LDA نیست. همچنین نتایج به دست آمده با چهارده نمونه آموزشی

Ground Truth Map (GTM)
Gaussian Maximum Likelihood (GML)
Support Vector Machine (SVM)
Radial basis function (RBF)
Fivefold Cross Validation
Average Accuracy

۱ Average Reliability

سال ۱۳۹۶ شماره ۲ پیاپی ۳۲

گزارش شده است. عدد داخل پرانتز در خانه‌های جدول نشان‌دهنده تعداد ویژگی‌هایی است که با استفاده از آنها بیشترین مقدار صحت متوسط به دست آمده است.

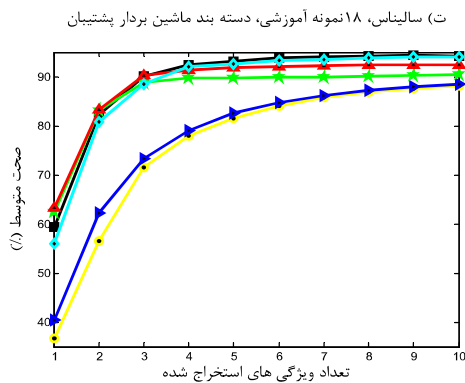
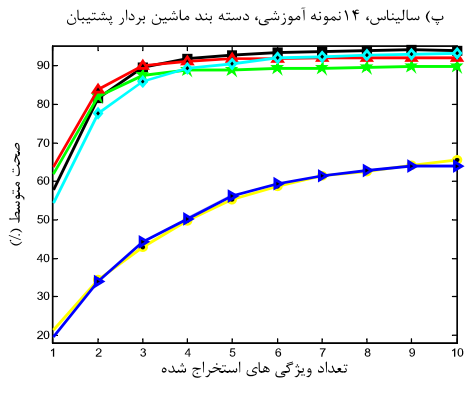
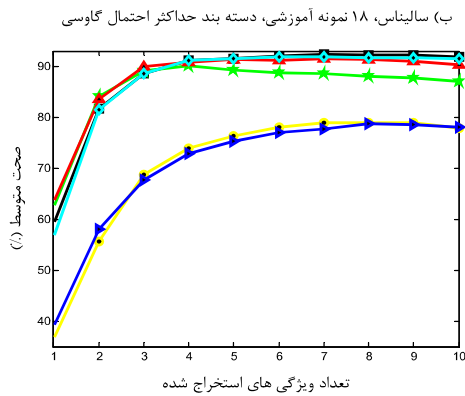
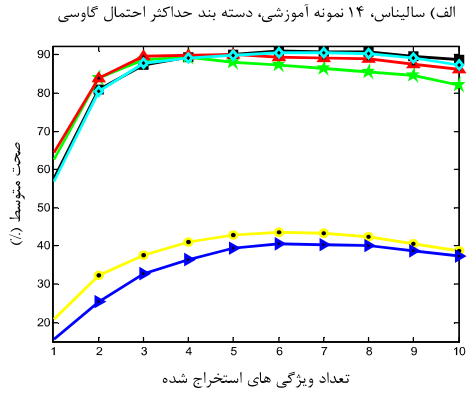
شکل (۶)، زمان محاسباتی لازم را برحسب ثانیه، برای روش‌های استخراج ویژگی با افزایش در تعداد نمونه‌های برچسب‌دار نشان می‌دهد. این نمونه‌ها از ۱۰ دسته با بیشترین تعداد نمونه، از دادگان ایندیانا انتخاب شده‌اند. نمونه‌های برچسب‌دار از ۵ تا ۵۰ و با گام ۵ افزایش یافته‌اند. شکل (۶) نشان می‌دهد که زمان محاسبات برای تمام روش‌های استخراج ویژگی غیرپارامتریک با افزایش در تعداد نمونه‌های آموزشی، افزایش می‌یابد. روش پارامتریک LDA، از اطلاعات کلی استفاده می‌کند. بنابراین، LDA سریع‌ترین روش است. در روش‌های غیرپارامتریک، زمان صرف‌شده برای محاسبه ماتریس‌های پراکندگی غیرپارامتریک، زمان محاسبات را افزایش می‌دهد. محاسبه ماتریس‌های پراکندگی غیرپارامتریک، با محاسبه میانگین‌های محلی، وزن نمونه‌ها و تشکیل بردارهای درون‌دسته‌ای و بین‌دسته‌ای وزن‌دار انجام می‌شود. در روش‌های پیشنهادی، NDA و CNFE، از k نمونه همسایه بدون در نظر گرفتن وزنی برای آنها، برای محاسبه میانگین محلی استفاده می‌شود. وزن در MNDA، تابعی از امضای طیفی، اطلاعات کلی و محلی نمونه‌ها و وزن در NDA، تابعی از امضای طیفی نمونه‌ها است. در CNFE، اندازه اطلاعات محلی روی اطلاعات کلی هر نمونه برای محاسبه وزن استفاده می‌شود. بنابراین، MNDA و NDA از CNFE کمی سریع‌تر هستند؛ و NDA نیز از MNDA کمی سریع‌تر است؛ زیرا مقدار k در آن ثابت است. در NWE، دو وزن متفاوت برای همه نمونه‌ها به دست می‌آید. ابتدا محاسبه وزن برای تعیین میانگین محلی وزن‌دار انجام می‌شود؛ سپس محاسبه وزن برای تشکیل ماتریس‌های پراکندگی انجام می‌شود. وزن در DA-PC1 با استفاده از اطلاعات نخستین مؤلفه اساسی به دست می‌آید. همچنین برای تشکیل ماتریس‌های پراکندگی در DA-PC1، بردارهای درون‌دسته‌ای و بین‌دسته‌ای وزن‌دار با استفاده از جفت نمونه‌های برچسب‌دار به دست می‌آیند. در صورتی که در سایر روش‌های غیرپارامتریک، بردارهای درون‌دسته‌ای و بین‌دسته‌ای وزن‌دار با استفاده از نمونه‌های برچسب‌دار و میانگین‌های محلی محاسبه می‌شوند. بنابراین DA-PC1 کندترین روش بوده و NWE سریع‌تر از DA-PC1 است.

نشان‌دهنده عدم کارایی این دو روش در مسائل با نمونه آموزشی محدود است. بیشترین مقدار برای میانگین صحت متوسط دسته‌بندی‌های GML و SVM با استفاده از ویژگی‌های استخراج‌شده از روش پیشنهادی با هجده نمونه آموزشی برای دادگان ایندیانا، به ترتیب با ۶ و ۱۰ ویژگی و برای دادگان سالیانس، به ترتیب با ۷ و ۹ ویژگی به دست آمده است. نتایج به دست آمده برای صحت و اطمینان‌پذیری دسته‌ها، و آزمون McNemars با تعداد ویژگی‌های مذکور در جدول‌های (۱) تا (۵) گزارش شده است و نقشه‌های دسته‌بندی در شکل‌های (۴) و (۵) نشان داده شده است.

نتایج گزارش‌شده در جدول‌های (۱) تا (۴) نشان می‌دهد که صحت دسته‌بندی برای بیشتر دسته‌ها با استفاده از ویژگی‌های استخراج‌شده از روش پیشنهادی، نسبت به سایر روش‌های استخراج ویژگی برتری دارد. صحت دسته‌بندی دسته‌ها با استفاده از روش DA-PC1 نزدیک به MNDA بوده و در بعضی موارد نسبت به آن برتری دارد. بررسی ماتریس سردرگمی به دست آمده از دسته‌بندی و جدول‌های (۱) تا (۴) نشان می‌دهد که دسته‌بندی بعضی از دسته‌ها در دادگان ایندیانا و سالیانس با مشکلاتی مواجه است. در دادگان ایندیانا، پیکسل‌ها در دسته‌های ۲ تا ۴، ۵ و ۶، ۱۰ تا ۱۲ و ۱۴ و ۱۵ امضای طیفی نزدیک به هم دارند. در دادگان سالیانس، پیکسل‌ها در دسته‌های ۸ و ۱۵ امضای طیفی نزدیک به هم دارند. بنابراین صحت و اطمینان‌پذیری دسته‌بندی در این دسته‌ها کاهش یافته است. البته نتایج به دست آمده از روش پیشنهادی برای این دسته‌ها نسبت به سایر روش‌ها بهبود داشته است. این مطلب نشان‌دهنده کارایی روش پیشنهادی در استخراج ویژگی‌های مناسب برای دسته‌بندی است. جدول (۵) نشان می‌دهد که دسته‌بندی دادگان ایندیانا و سالیانس با استفاده از ویژگی‌های استخراج‌شده از روش پیشنهادی نسبت به سایر روش‌ها، دقیق‌تر بوده و با خطای کمتری انجام شده است؛ البته نتایج روش DA-PC1 برای دادگان سالیانس با دسته‌بند SVM از روش پیشنهادی دقیق‌تر است. این مطلب می‌تواند به دلیل حفظ‌شدن ساختار غالب توزیع نمونه‌ها با استفاده از مؤلفه اساسی نخست در تابع وزن باشد.

بیشترین مقدار به دست آمده برای صحت متوسط دسته‌بندی با ۱۴ و ۱۸ نمونه آموزشی، در جدول (۶)

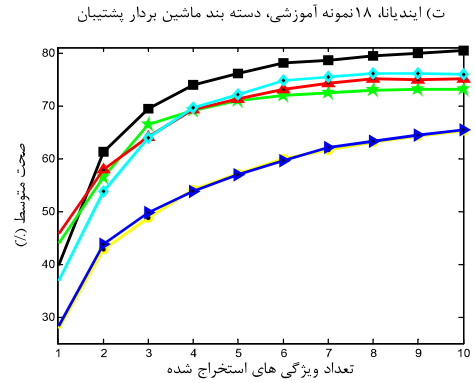
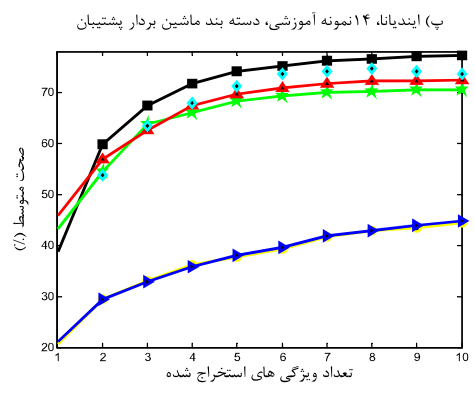
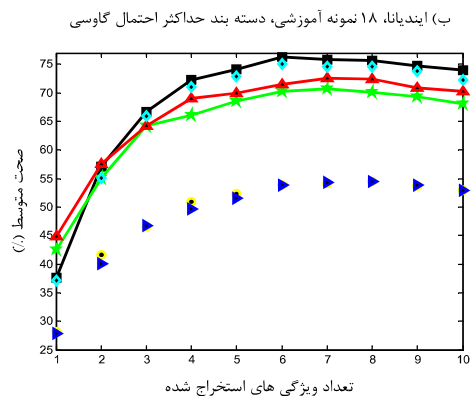
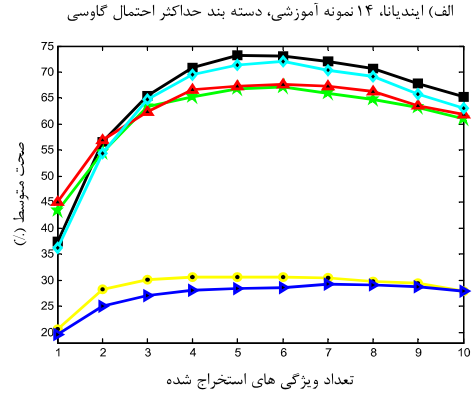
■ MNDA ● LDA ▲ NDA ★ NWFE ▲ CNFE ◆ DA-PCI



(شکل-۳): صحت متوسط دسته‌بند در برابر ویژگی‌های استخراج شده برای دادگان سالیانس

(Figure-3): Average accuracy versus number of extracted features for Salinas data set

■ MNDA ● LDA ▲ NDA ★ NWFE ▲ CNFE ◆ DA-PCI



(شکل-۲): صحت متوسط دسته‌بند در برابر ویژگی‌های استخراج شده برای دادگان ایندیانا

(Figure-2): Average accuracy versus number of extracted features for Indian Pines data set

(جدول-۱): نتایج دسته‌بندی برای دادگان ایندیانا، به‌دست آمده توسط دسته‌بند GML با ۱۸ نمونه آموزشی و ۶ ویژگی استخراج شده
 (Table-1): Classification results for Indian Pines data set, obtained by GML classifier, 18 training samples and 6 extracted features

شماره دسته	نام دسته	تعداد نمونه‌ها	روش های استخراج ویژگی											
			MNDA		LDA		NDA		NWFE		CNFE		DA-PCI	
			صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری
1	یونجه	64	91.96	41.86	84.78	25.24	85.22	26.62	88.26	37.91	89.35	41.95	91.96	47.84
2	ذرت- بدون شخم	1428	58.02	58.38	22.95	30.87	20.06	32.14	43.03	52.04	45.49	56.27	54.69	56.69
3	ذرت- کم شخم	830	53.43	58.12	32.09	20.74	32.78	20.45	48.22	57.56	49.36	59.51	53.11	57.81
4	ذرت	237	82.49	39.8	32.49	13.35	38.31	12.23	68.95	31.11	67.51	31.44	82.57	39.02
5	سیزه- چمنزار	483	73.85	59.76	59.48	42.52	64.37	40.22	73.99	60.45	72.17	62.64	71.33	54.52
6	سیزه- درختان	730	78.74	93.14	55.41	68.95	60.21	70.58	72.3	88.56	72.12	86.03	77.36	92.36
7	سیزه- چمنزار کوتاه	28	96.79	73.57	94.29	24.12	91.79	26.77	94.65	76.54	95	78.27	96.07	67.87
8	یونجه- دسته شده	478	83.79	98.13	56.65	94.6	53.14	94.89	82.32	97.38	83.14	97.52	85.33	98.4
9	جو صحرایی	20	100	64.94	98	8.63	99	9.3	98.5	32.68	99.5	42.3	100	63.13
10	سویا- بدون شخم	972	70.34	54.96	28.98	25.21	29.68	25.28	61.54	36.9	61.36	38.87	69.17	52.42
11	سویا- کم شخم	2455	50.38	71.27	23.45	45.49	26.25	45.14	46.33	66.08	48.82	66.87	49.03	70.99
12	سویا- شخم کامل	593	60	45.61	30.49	20.94	27.66	23.89	38.57	31.79	45.72	35.63	54.2	39.23
13	گندم	205	88.68	98.98	83.41	72.54	80.1	76.15	90.59	97.32	88.39	98.07	89.51	98.61
14	بیشه	1265	70.36	91.21	44.68	85.42	47.31	86.78	74.29	91.1	74.22	89.91	69.38	89.07
15	ساختمان- سبزه	386	69.69	43.04	41.45	23.81	37.41	26.52	53.24	35.83	59.59	36.99	65.34	43.05
16	سنگ- برج	93	93.22	83.97	74.52	81.27	67.96	83.21	90.54	91.72	91.08	89.48	92.47	84.95
صحت و اطمینان پذیری متوسط			76.36 ±1.12	67.3 ±1.46	53.94 ±1.33	42.73 ±1.43	53.83 ±1.94	43.76 ±1.54	70.33 ±1.54	61.56 ±1.72	71.43 ±1.55	63.23 ±2.27	75.09 ±0.87	66 ±1.72

(جدول-۲): نتایج دسته‌بندی برای دادگان ایندیانا، به‌دست آمده توسط دسته‌بند SVM با ۱۸ نمونه آموزشی و ۱۰ ویژگی استخراج شده
 (Table-2): Classification results for Indian Pines data set, obtained by SVM classifier, 18 training samples and 10 extracted features

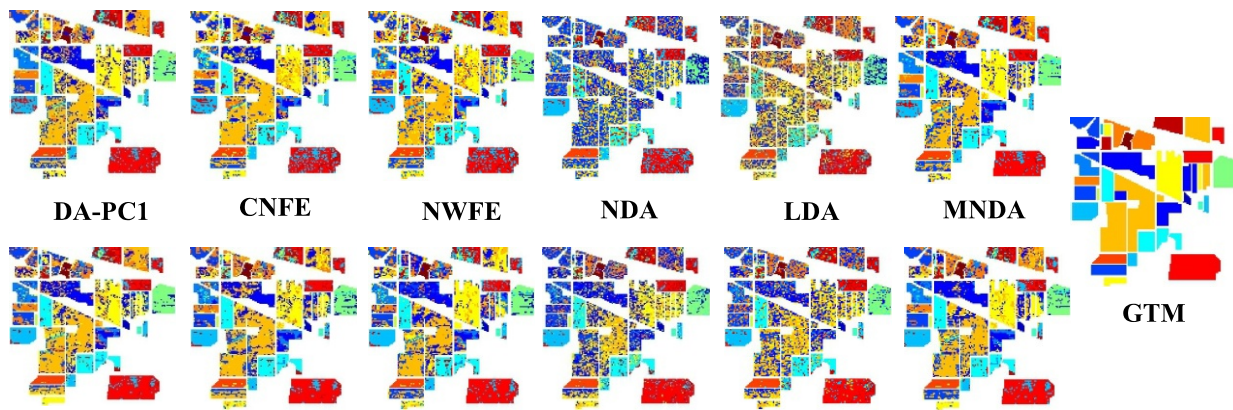
شماره دسته	نام دسته	تعداد نمونه‌ها	روش های استخراج ویژگی											
			MNDA		LDA		NDA		NWFE		CNFE		DA-PCI	
			صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری
1	یونجه	64	93.91	41.98	91.52	37.63	90.87	36.55	88.48	31.64	89.78	32.64	90.87	39.42
2	ذرت- بدون شخم	1428	66.8	61.03	36.49	42.29	38.56	42.83	49.15	46.91	55.05	54.84	56.2	61.28
3	ذرت- کم شخم	830	66.85	56.82	33.36	26.73	35.75	25.71	56	48.93	56.27	54.36	51.84	56.7
4	ذرت	237	81.22	45.79	45.99	19.11	42.49	22.05	66.92	33.77	73	35.51	76.75	39.54
5	سیزه- چمنزار	483	84.31	64.44	72.71	64.85	71.59	63.65	83.15	64.73	81.53	65.41	77.33	60.41
6	سیزه- درختان	730	86.03	91.09	76.27	83.92	77.04	84.62	82.22	82.93	81.07	87.99	83.31	87.77
7	سیزه- چمنزار کوتاه	28	98.57	73.25	98.93	37.01	98.22	32.53	97.86	37.72	97.5	50.72	97.86	77.63
8	یونجه- دسته شده	478	86.28	99.43	77.76	94.95	75.63	95.94	77.24	97.5	80.84	97.94	86.59	98.88
9	جو صحرایی	20	100	46.46	97.5	10.23	99	10.97	100	36.11	100	31.95	100	52.16
10	سویا- بدون شخم	972	74.58	60.77	39.35	29.55	38.27	31.73	63	52.4	64.41	53.65	70	52.35
11	سویا- کم شخم	2455	51.34	79.05	28.62	51.04	31.16	50.74	48.27	75.79	50.94	72.06	51.5	71.62
12	سویا- شخم کامل	593	66.26	58.1	45.35	31.75	43.58	35.14	49.17	37.39	51.01	41.76	50.57	37.15
13	گندم	205	97.02	90.88	93.12	77.28	93.56	82.99	95.22	89.4	95.37	87.33	97.17	86.89
14	بیشه	1265	76.09	92.67	67.6	90.53	73.06	90.34	75.5	93.96	78.51	92.69	78	91.19
15	ساختمان- سبزه	386	60.34	51.38	52.25	37.93	50.75	42.39	42.49	38.36	49.48	45.02	54.4	50.45
16	سنگ- برج	93	95.8	84.95	85.81	80.13	87.74	82.92	95.48	89.36	95.59	87.03	93.44	90.46
صحت و اطمینان پذیری متوسط			80.34 ±0.78	68.63 ±1.85	65.17 ±1.47	50.93 ±1.73	65.45 ±1.31	51.94 ±1.44	73.13 ±0.78	59.8 ±1.14	75.02 ±1.16	61.93 ±1.21	75.99 ±1.15	65.87 ±1.38

(جدول-۳): نتایج دسته‌بندی برای دادگان سالیانس، به‌دست آمده توسط دسته‌بند GML با ۱۸ نمونه آموزشی و ۷ ویژگی استخراج‌شده
(Table-3): Classification results for Salinas data set, obtained by GML classifier, 18 training samples and 7 extracted features

شماره دسته	نام دسته	تعداد نمونه‌ها	روش های استخراج ویژگی											
			MNDA		LDA		NDA		NWFE		CNFE		DA-PC1	
			صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری
1	بروکلی-سیزه- علف هرز ۱	2009	98.9	98.9	92.6	88.7	86.3	81.5	93.92	97.58	96.91	98.06	98.86	97.93
2	بروکلی-سیزه- علف هرز ۲	3726	98.82	99.41	92.78	95.85	86.8	91.87	95.72	96.43	97.38	98.34	98.21	99.38
3	زمین شخم شده	1976	96.61	91.55	71.73	64.01	72.52	62.45	86.04	82.9	94.55	92.08	96.49	88.7
4	زمین شخم شده زیر	1394	99.62	98.12	88.1	76.96	91.68	78.52	98.84	97.59	99.54	98.4	99.56	97.73
5	زمین شخم شده هموار	2678	91.37	98.74	64.79	78.81	59.21	79.27	84.97	94.55	91.58	98.11	89.28	98.61
6	کاه بن	3959	99.55	99.46	99.25	99.59	99.48	99.06	99.13	99.91	99.26	99.82	99.53	99.6
7	کرفس	3579	97.66	99.98	98.98	99.63	98.47	99.44	96.47	99.77	96.11	99.6	97.45	99.97
8	انگور	11271	59.86	71.78	57.16	63.83	55.44	63.65	55.14	70.28	55.93	72.14	60.54	71.24
9	خاک- تاکستان	6203	92.7	99.62	70.19	90.63	66.66	92.59	90.4	99.5	93.13	99.55	90.92	99.63
10	ذرت-سیزه- علف هرز	3278	94.35	78.38	65.66	71.1	68.39	59.26	88.25	63.73	92.97	72.04	93.9	75.88
11	کاهو- ۴ هفته	1068	98.3	92.92	86.38	74.85	85.64	74.24	96.83	83.34	97.87	93.22	97.52	92.71
12	کاهو- ۵ هفته	1927	96.66	98.22	66.77	47.56	63.88	51.97	92.32	97.34	96.81	98.84	96.57	98.57
13	کاهو- ۶ هفته	916	97.34	97.93	73.58	62.17	72.53	63.71	92.14	95.73	96.3	97.37	97.67	97.67
14	کاهو- ۷ هفته	1070	95.09	92.76	88.66	78.19	88.98	74.76	90.43	90.68	93.28	93.62	95.05	92.43
15	تاکستان	7268	62.64	51.71	50.03	44.45	50.87	43.35	59.42	48.84	63.57	50.04	61.09	51.6
16	تاکستان- شاغولی- داربست	1807	98.13	95.83	96.83	97.04	96.56	97.34	97.69	85.94	98.44	90.67	98.11	95.06
صحت و اطمینان پذیری متوسط			92.35 ±1.02	91.58 ±0.98	78.97 ±1.67	77.08 ±1.78	77.71 ±2.15	75.81 ±1.9	88.61 ±0.7	87.76 ±1.07	91.48 ±0.93	90.74 ±0.84	91.92 ±1.04	91.05 ±1.15

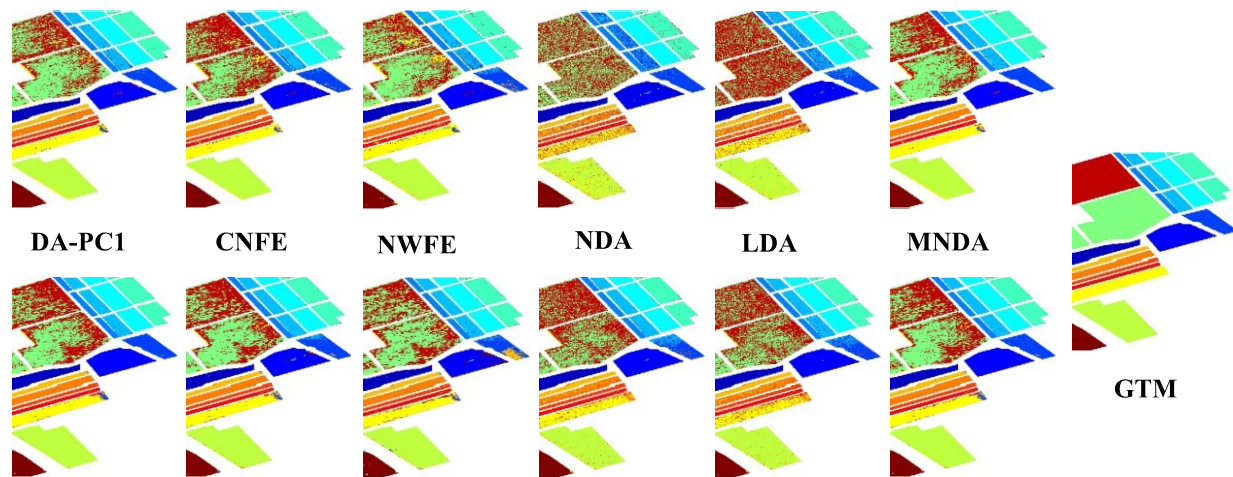
(جدول-۴): نتایج دسته‌بندی برای دادگان سالیانس، به‌دست آمده توسط دسته‌بند SVM با ۱۸ نمونه آموزشی و ۹ ویژگی استخراج‌شده
(Table-4): Classification results for Salinas data set, obtained by SVM classifier, 18 training samples and 9 extracted features

شماره دسته	نام دسته	تعداد نمونه‌ها	روش های استخراج ویژگی											
			MNDA		LDA		NDA		NWFE		CNFE		DA-PC1	
			صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری	صحت	اطمینان پذیری
1	بروکلی-سیزه- علف هرز ۱	2009	98.2	99.82	98.4	98.42	97.49	97.41	96.37	96.97	97.52	99.01	98.36	99.94
2	بروکلی-سیزه- علف هرز ۲	3726	99.81	99	98.71	99.14	98.12	98.64	95.74	97.74	97.6	98.64	99.85	99.09
3	زمین شخم شده	1976	98.7	93	81.95	85.97	86.58	89.52	91.11	89	96.18	91.5	97.82	91.99
4	زمین شخم شده زیر	1394	99.67	96.58	95.09	87.84	96.44	92.12	99.22	93.41	99.69	96.03	99.73	96.93
5	زمین شخم شده هموار	2678	97.49	98.23	86.67	85.47	90.74	88.42	95.07	96.45	95.96	96.73	97.38	97.78
6	کاه بن	3959	99.54	99.98	99.22	99.97	99.34	99.97	99.46	99.87	99.47	99.97	99.69	99.99
7	کرفس	3579	99.66	99.51	99.57	99.72	99.5	99.64	99.35	97.11	99.41	98.13	99.74	99.45
8	انگور	11271	69.88	78.09	54.79	71.3	56.07	72.85	59.62	74.42	66.38	77.14	64.18	77.01
9	خاک- تاکستان	6203	99.3	98.66	89.88	93.5	92.37	94.59	98.35	98.59	98.79	98.85	99.43	98.61
10	ذرت-سیزه- علف هرز	3278	91.06	89.06	76.73	86.71	79.16	88.49	84.18	84.83	88.88	87.03	90.83	90.69
11	کاهو- ۴ هفته	1068	96.43	92.3	92.57	92.94	93.07	92.63	94.12	83.32	97.05	91.81	95.99	94.57
12	کاهو- ۵ هفته	1927	99.53	97.57	85.41	65.59	90.33	71.08	99.13	93.98	99.72	96.48	99.85	96.61
13	کاهو- ۶ هفته	916	98.72	95.5	91.7	75.5	93.01	79.34	98.05	88.1	98.42	93.76	98.9	96.59
14	کاهو- ۷ هفته	1070	94.89	94.38	90.86	89.13	91.09	91.48	91.52	90.72	92.82	95.05	95.49	94.52
15	تاکستان	7268	68.19	62.04	65.21	49.55	66.74	50.79	65.7	53.58	68.05	59.41	69.07	57.92
16	تاکستان- شاغولی- داربست	1807	97.82	98.53	98.05	99.55	98.31	99.49	95.43	92.31	97.31	94.73	98.2	99
صحت و اطمینان پذیری متوسط			94.31 ±0.42	93.27 ±0.76	87.64 ±1.04	86.31 ±1.12	87.93 ±1.22	87.39 ±1.17	90.23 ±0.73	88.89 ±1.24	92.41 ±0.37	91.69 ±0.52	94.03 ±0.57	93.17 ±0.72



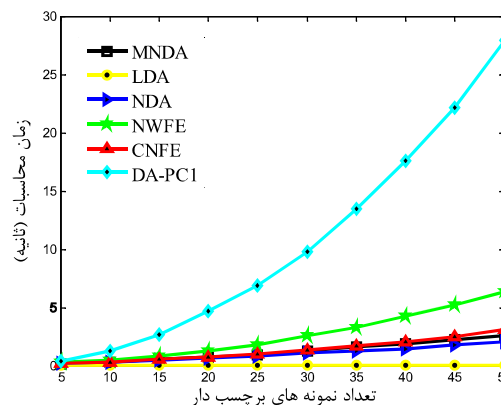
(شکل-۴): نقشه درستی و نقشه‌های دسته‌بندی برای دادگان ایندیانا با ۱۸ نمونه آموزشی، به‌دست آمده توسط دسته‌بند GML و ۶ ویژگی استخراج‌شده در بالا، و دسته‌بند SVM و ۱۰ ویژگی استخراج‌شده در پایین

(Figure-4): The GTM and classification maps of the Indian Pines data set by 18 training samples, obtained by GML classifier and 6 extracted features in the top, and SVM classifier and 10 extracted feature in the bottom



(شکل-۵): نقشه درستی و نقشه‌های دسته‌بندی برای دادگان سالیانس با ۱۸ نمونه آموزشی، به‌دست آمده توسط دسته‌بند GML و ۷ ویژگی استخراج‌شده در بالا، و دسته‌بند SVM و ۹ ویژگی استخراج‌شده در پایین

(Figure-5): The GTM and classification maps of the Salinas data set by 18 training samples, obtained by GML classifier and 7 extracted features in the top, and SVM classifier and 9 extracted feature in the bottom



(شکل-۶): زمان محاسبات در برابر تعداد نمونه‌های برچسب‌دار برای دادگان ایندیانا
 (Figure-6): Computational times versus number of labelled samples for Indian Pines data set

(جدول-۵): نتایج آزمون McNemars با ۱۸ نمونه آموزشی، خانه‌های جدول نشان‌دهنده Z_{op} هستند که o سطر و p ستون را نشان می‌دهد (Table-5): The McNemars test results by 18 training samples, each case of the table shows Z_{op} where o is the row and p is the column

این‌دیانا، دسته‌بند GML، ۶ ویژگی استخراج شده						
	MNDA	LDA	NDA	NWFE	CNFE	DA-PCI
MNDA	0	41.84	40.76	12.77	10.27	5.1
LDA	-41.84	0	-1.81	-32.59	-34.73	-39.42
NDA	-40.76	1.81	0	-31.54	-33.61	-38.28
NWFE	-12.77	32.59	31.54	0	-4.45	-9.84
CNFE	-10.27	34.73	33.61	4.45	0	-7
DA-PCI	-5.1	39.42	38.28	9.84	7	0
این‌دیانا، دسته‌بند SVM، ۱۰ ویژگی استخراج شده						
	MNDA	LDA	NDA	NWFE	CNFE	DA-PCI
MNDA	0	34.18	32.25	16.41	12.11	10.48
LDA	-34.18	0	-3.47	-21.11	-25.41	-27.07
NDA	-32.25	3.47	0	-19.03	-23.39	-25.15
NWFE	-16.41	21.11	19.03	0	-5.41	-6.58
CNFE	-12.11	25.41	23.39	5.41	0	-2.02
DA-PCI	-10.48	27.07	25.15	6.58	2.02	0
سالیانس، دسته‌بند GML، ۷ ویژگی استخراج شده						
	MNDA	LDA	NDA	NWFE	CNFE	DA-PCI
MNDA	0	3.14	3.15	1.05	0.46	0.13
LDA	-3.14	0	0.13	-2.13	-2.95	-3.04
NDA	-3.15	-0.13	0	-2.24	-2.85	-3.03
NWFE	-1.05	2.13	2.24	0	-0.98	-0.91
CNFE	-0.46	2.95	2.85	0.98	0	-0.31
DA-PCI	-0.13	3.04	3.03	0.91	0.31	0
سالیانس، دسته‌بند SVM، ۹ ویژگی استخراج شده						
	MNDA	LDA	NDA	NWFE	CNFE	DA-PCI
MNDA	0	2.87	3.37	3.68	0.85	-0.73
LDA	-2.87	0	-1.33	1.01	-1.59	-2.22
NDA	-3.37	1.33	0	1.89	-0.83	-1.82
NWFE	-3.68	-1.01	-1.89	0	-3.32	-4.42
CNFE	-0.85	1.59	0.83	3.32	0	-1.42
DA-PCI	0.73	2.22	1.82	4.42	1.42	0

(جدول-۶): بیشترین مقدار به دست آمده برای صحت متوسط دسته‌بندها با استفاده از ۱۴ و ۱۸ نمونه آموزشی (Table-6): The highest average classification accuracy by using 14 and 18 training samples

دادگان	تعداد نمونه‌های آموزشی	دسته‌بند	روش‌های استخراج ویژگی					
			MNDA	LDA	NDA	NWFE	CNFE	DA-PCI
این‌دیانا	14	ML	73.3±1.98 (5)	30.7±2.03 (6)	29.21±1.65 (7)	67.12±2.51 (6)	72.03±1.86 (6)	58.12±1.8 (7)
		SVM	77.2±1.69 (10)	44.45±2.75 (10)	44.74±2.24 (10)	70.47±1.39 (9)	72.35±1.91 (10)	74.67±1.6 (8)
	18	ML	76.36±1.12 (6)	54.43±2.12 (8)	54.54±1.8 (8)	70.74±1.51 (7)	72.55±0.97 (7)	75.09±0.87 (6)
		SVM	80.34±0.78 (10)	65.17±1.47 (10)	65.45±1.31 (10)	73.13±0.78 (10)	75.06±0.8 (8)	76.02±0.75 (8)
سالیانس	14	ML	91.05±1.11 (6)	43.54±4.63 (6)	40.65±5.43 (6)	89.31±0.77 (4)	90.07±0.89 (5)	90.57±1.04 (6)
		SVM	94.1±0.57 (9)	65.51±4.09 (10)	64.02±3.45 (9)	89.94±0.97 (10)	92.2±0.49 (10)	93.19±0.63 (10)
	18	ML	92.35±1.02 (7)	78.97±1.67 (7)	78.77±2.2 (8)	90.18±0.41 (4)	91.48±0.93 (7)	91.92±1.04 (7)
		SVM	94.31±0.42 (9)	88.15±1.03 (10)	88.44±1.03 (10)	90.38±0.62 (10)	92.43±0.39 (10)	94.03±0.57 (9)

مناسب و کافی شده است. همچنین در این روش، پارامتر آزاد وجود ندارد و برای تعیین تعداد نمونه‌های همسایه، از الگوریتم ارائه شده، استفاده می‌شود. ارزیابی روش پیشنهادی با استفاده از دادگان واقعی ابرطیفی این‌دیانا و سالیانس انجام شده است. نتایج آزمایش‌ها نشان می‌دهد که صحت و

۵- نتیجه‌گیری

در این مقاله، روشی برای بهبود NDA به منظور استخراج ویژگی تصاویر ابرطیفی ارائه شده است. استفاده از توابع وزن پیشنهادی در تشکیل ماتریس‌های پراکندگی و تنظیم ماتریس پراکندگی درون‌دسته‌ای، باعث استخراج ویژگی‌های

- [10] M. Kamandar, and H. Ghassemian, "Linear feature extraction for hyperspectral images based on information theoretic learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, pp. 702-706, 2013.
- [11] B. C. Kuo, and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification" *IEEE Transaction on Geoscience and Remote Sensing*, vol. 42, pp. 1096-1105, 2004.
- [12] J. Yang, P. Yu, and B. C. Kuo, "A nonparametric feature extraction and its application to nearest neighbor classification for hyperspectral image data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 48, pp. 1279-1293, 2010.
- [13] M. Imani, and H. Ghassemian, "Feature reduction of hyperspectral images: discriminant analysis and the first principal component," *journal of AI and Data Mining*, vol. 3, pp. 1-9, 2015.
- [14] M. Imani, and H. Ghassemian, "Feature extraction using attraction points for classification of hyperspectral images in a small sample size situation," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, pp. 1325-1329, 2014.
- [15] M. Imani, and H. Ghassemian, "Feature space discriminant analysis for hyperspectral data feature reduction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 102, pp. 1-13, 2015.
- [16] R. O. Duda, P. E. Hart, and D. G. Stock, *Pattern classification*, 2nd ed. New York, Wiley, 2001.
- [17] C. Chang, and C. Lin, "LIBSVM : A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 1-27, 2011.
- [18] G. M. Foody, "Thematic map comparison: evaluating the statistical significance of differences in classification accuracy," *Photogrammetric Engineering and Remote Sensing*, vol. 70, pp. 627-633, 2004.

اطمینان‌پذیری دسته‌بندی با استفاده از ویژگی‌های استخراج‌شده از روش پیشنهادی نسبت به سایر روش‌ها، برتری داشته است. روش پیشنهادی علاوه بر رفع مشکلات NDA، کارایی مناسبی در شرایط نمونه آموزشی محدود دارد.

6-References

۶- مراجع

- [1] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, pp. 55-63, 1968.
- [2] M. Imani, and H. Ghassemian, "Binary coding based feature extraction in remote sensing high dimensional data," *Information Sciences*, vol. 342, pp. 191-208, 2016.
- [3] S. A. Hosseini, and H. Ghassemian, "Rational function approximation for feature reduction in hyperspectral data," *Remote Sensing Letters*, vol. 7, pp. 101-110, 2015.
- [4] S. A. Hosseini, and H. Ghassemian, "Hyperspectral data feature extraction using rational function curve fitting," *Signal and Data Processing*, vol. 13, pp. 3-16, 2016.
- [5] M. Imani, and H. Ghassemian, "Feature reduction of hyperspectral data for increasing of class separability and preserving of data structure," *Signal and Data Processing*, vol. 14, pp. 71-82, 2017.
- [6] W. Liao, A. Pizurica, P. Scheunders, W. Philips, and Y. Pi, "Semisupervised local discriminant analysis for feature extraction in hyperspectral images," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 51, pp. 184-198, 2013.
- [7] Z. Feng, Sh. Yang, Sh. Wang, and L. Jiao, "Discriminative spectral-spatial margin-based semisupervised dimensionality reduction of hyperspectral data," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, pp. 224-228, 2015.
- [8] K. Fukunaga, *Introduction to statistical pattern recognition*. San Diego, CA, USA, Academic, 1990.
- [9] X. He, and P. Niyogi, "Locality preserving projections," *Advances in Neural Information Processing Systems*, vol. 16, pp. 153-160, 2004.



آزاده کیانی سرکله متولد ۱۳۶۲،

تحصیلات متوسطه خود را در شهرستان رودسر سپری و در سال ۱۳۸۴ مدرک کارشناسی خود را در رشته مهندسی پزشکی از دانشگاه اصفهان اخذ کرد؛

سپس تحصیلات خود را در رشته مهندسی برق- الکترونیک در دانشگاه گیلان ادامه داد و در سال ۱۳۸۸ موفق به اخذ مدرک کارشناسی ارشد شد. در سال ۱۳۸۹ در مقطع دکترای رشته مهندسی برق- الکترونیک در دانشگاه آزاد اسلامی واحد علوم و تحقیقات پذیرفته شد و در سال ۱۳۹۵ موفق به اخذ مدرک دکترا شد. ایشان هم‌اکنون عضو هیئت علمی دانشگاه آزاد اسلامی واحد رشت است. پردازش تصویر، پردازش سیگنال‌های پزشکی، شناسایی الگو و طراحی مدارهای مجتمع دیجیتال از جمله سرفصل‌های پژوهشی مورد علاقه ایشان است.

نشانی رایانامه ایشان عبارت است از:

azadeh.kiani@iaurasht.ac.ir



محمد حسن قاسمیان یزدی تحصیلات

کارشناسی خود را در رشته مهندسی مخابرات در سال ۱۳۸۵ در دانشکده مخابرات ایران به پایان رساند. ایشان مدارک کارشناسی ارشد و دکترای خود را در رشته

مخابرات از دانشگاه پردو آمریکا به ترتیب در سال ۱۳۶۹ و ۱۳۶۷ اخذ کرد؛ سپس به‌عنوان عضو هیئت علمی دانشگاه تربیت مدرس به تدریس و پژوهش پرداخت. ایشان در حال حاضر، استاد دانشگاه تربیت مدرس است. آنالیز و پردازش تصاویر از منابع چندگانه، پردازش اطلاعات، شناسایی الگو، سنجش از دور و پردازش سیگنال‌های پزشکی از جمله سرفصل‌های پژوهشی مورد علاقه ایشان است.

نشانی رایانامه ایشان عبارت است از:

ghassemi@modares.ac.ir

