Journal of Operational Research in its Applications
Vol.23, NO.1, Spring 2026
ISSN: 2251-7286
DOI: 10.71773/jamlu-2026-1-2325

Classification of Grey Streaming Data Using Integrated Data Envelopment Analysis and Multi-Attribute Decision Making: Application to Vegetation Clusterin

M. Davoudabadi[1], A. Mohtadi[2*]
[1] *Assistant Professor, Department of Remote Sensing, Hikmat Institute of Higher Education, Qom, Iran*
[2] *M.Sc. in Remote Sensing, Hikmat Institute of Higher Education, Qom, Iran*

Research Paper

Received: 20 August 2025
Accepted: 1 February 2026

Abstract: This study proposes a novel method for classifying grey streaming data by integrating Data Envelopment Analysis (DEA) with Multi-Attribute Decision Making (MADM) techniques. Streaming data are inherently dynamic and continuously evolving, and the presence of grey uncertainty further increases the complexity of their analysis. In this research, Landsat satellite imagery was used as a representative example of streaming data. Pixel brightness values were first normalized into fractional grey numbers within the interval [0, 1], and the Normalized Difference Vegetation Index (NDVI) was computed to provide an initial unsupervised separation between vegetated and non-vegetated areas. To determine an optimal threshold between these two classes, a permutation-based MADM algorithm was employed.
The proposed model was first evaluated using simulated data and then applied to real satellite imagery from a region in Tehran, Iran. Validation based on randomly selected reference points using Google Earth images demonstrated that the method achieved an overall vegetation-classification accuracy of approximately 80%. The main contribution of this research lies in developing a hybrid DEA–MADM framework capable of handling continuous and uncertain grey data, thereby expanding the analytical capacity of operations research techniques within remote sensing applications.

Introduction: Grey streaming data exhibit both temporal dynamism and uncertainty, making their classification a significant challenge in data analysis. Integrating mathematical programming approaches -particularly DEA- with multi-criteria decision-making techniques offers a promising solution for addressing these complexities. This study introduces a hybrid analytical framework designed to classify grey streaming data derived from multispectral satellite imagery.

Materials and Methods: Landsat imagery was utilized as the primary data source. Pixel values from selected spectral bands were normalized into grey fractional numbers, and NDVI was calculated to provide a primary indicator for vegetation clustering. To establish a reliable threshold between vegetated and non-vegetated classes, a permutation-based MADM procedure was implemented. The model was validated through both simulated data and real imagery from northern Tehran.

* *Corresponding Author: mohtadi1375@gmail.com*

Results and Discussion: The MADM algorithm successfully identified a threshold that maximized separability between the two vegetation classes. The hybrid DEA–MADM model demonstrated effective classification performance across both simulated and real datasets. Ground-truth validation conducted using ten randomly selected points showed that eight were correctly classified, yielding an overall accuracy of roughly 80%. These results highlight the model's robustness in handling uncertainty, as well as the interpretability gained from integrating DEA with multi-criteria decision making.

Conclusions: This research introduces an integrated algorithmic framework that combines DEA and MADM for classifying grey streaming data. The method demonstrates high potential for vegetation-cover classification and provides a flexible foundation for applications involving dynamic and uncertain data. Future studies may include integrating supervised learning methods, optimizing the thresholding process using evolutionary algorithms, and applying the framework to other environmental-monitoring domains.

Keywords: Data Envelopment Analysis (DEA), Grey Streaming Data, Multi-Attribute Decision Making (MADM).

مجله تحقیق در عملیات در کاربردهای آن، سال بیست و سوم، شماره اول (پیاپی ۸۸)،بهار ۰۵، ص ص ۹۲-۷۱
شاپا ۷۲۸۶-۲۲۵۱
DOI: 10.71773/jamlu-2026-1-2325

دسته‌بندی داده‌های جریانی خاکستری با تلفیق تحلیل پوششی داده‌ها و تصمیم‌گیری چندشاخصه: کاربرد در خوشه‌بندی پوشش گیاهی

میثم داودآبادی[1]، علی مهتدی[2*]
۱- استادیار، گروه سنجش از دور موسسه آموزش عالی حکمت قم، قم، ایران
۲- کارشناس ارشد، گروه سنجش از دور، موسسه آموزش عالی حکمت قم، قم، ایران

رسید مقاله: ۲۹ مرداد ۱۴۰۴
پذیرش مقاله: ۱۲ بهمن ۱۴۰۴

چکیده

در این پژوهش روشی نوین برای دسته‌بندی داده‌های جریانی خاکستری ارایه شده است که بر ترکیب تحلیل پوششی داده‌ها (DEA) و روش‌های تصمیم‌گیری چندشاخصه (MADM) استوار است. داده‌های جریانی به دلیل پویایی زمانی و حجم بالا، و نیز همراهی با عدم قطعیت، چالش‌برانگیزترین انواع داده برای طبقه‌بندی به شمار می‌آیند. در این مطالعه، داده‌های ماهواره‌ای لندست به‌عنوان نمونه‌ای از داده‌های جریانی انتخاب شد و مقادیر روشنایی پیکسل‌ها پس از نرمال‌سازی، به صورت داده‌های کسری در بازه صفر تا یک مدل‌سازی گردید. سپس با استفاده از شاخص نرمال‌شده تفاضلی پوشش گیاهی (NDVI)، پیکسل‌ها به‌صورت نظارت‌نشده در دو خوشه «دارای پوشش گیاهی» و «فاقد پوشش گیاهی» تفکیک شدند. برای تعیین آستانه تفکیک بین دو خوشه، الگوریتم تصمیم‌گیری چندشاخصه مبتنی بر جایگشت به کار گرفته شد تا بهترین خط جدایش بر اساس معیارهای از پیش تعیین‌شده انتخاب شود. مدل پیشنهادی ابتدا بر داده‌های شبیه‌سازی‌شده آزمایش و سپس بر داده‌های واقعی پوشش گیاهی در محدوده‌ای از شهر تهران پیاده‌سازی شد. نتایج اعتبارسنجی با استفاده از نقاط نمونه و تصاویر گوگل ارث نشان داد که روش ارایه‌شده قادر است با دقتی در حدود ۸۰ درصد، خوشه‌بندی قابل اتکایی از پوشش گیاهی ارایه کند. نوآوری اصلی تحقیق، توسعه الگوریتمی ترکیبی برای دسته‌بندی داده‌های جریانی خاکستری و بهره‌گیری از رویکردهای تحقیق در عملیات در تحلیل داده‌های سنجش از دور است؛ الگوریتمی که قابلیت تعمیم به دیگر حوزه‌های مواجه با داده‌های حجیم و نامطمئن را نیز دارد.

کلمات کلیدی: تحلیل پوششی داده‌ها، داده‌های جریانی خاکستری، تصمیم‌گیری چندشاخصه، خوشه‌بندی پوشش گیاهی، لندست.

۱ مقدمه

مساله‌ی دسته‌بندی به‌عنوان یکی از مسایل بنیادی در یادگیری ماشین و تحلیل داده، به معنای تخصیص یک

* عهده‌دار مکاتبات
آدرس الکترونیکی: mohtadi1375@gmail.com

کلاس یا دسته به یک داده بر اساس تابع یا مدلی از پیش تعیین‌شده تعریف می‌شود. چنین مدلی معمولا با مقایسه‌ی داده‌های آموزشی که کلاس و شناسه‌ی آن‌ها مشخص است و با بهره‌گیری از الگوریتم‌هایی همچون درخت تصمیم، شبکه‌های عصبی مصنوعی، ماشین بردار پشتیبان و رگرسیون لجستیک ساخته می‌شود [۱]. این مدل‌ها امکان تشخیص تعلق یک داده به یک کلاس خاص را فراهم کرده و در کاربردهای متعددی همچون کنترل کیفیت کامپیوتری، پایش وضعیت ماشین‌آلات، و طراحی فرآیندهای تولیدی مورد استفاده قرار گرفته‌اند [۲].

یکی از چالش‌های اساسی در این حوزه، ایجاد مدل‌های دسته‌بندی‌کننده برای داده‌های غیرقطعی و فازی است؛ چرا که در بسیاری مواقع داده‌های جمع‌آوری‌شده به دلایل مختلف قطعی نیستند [۳]. همچنین، در بسیاری کاربردها داده‌ها به‌صورت جریانی مشاهده می‌شوند و نه به‌طور ایستا، که این امر خود پیچیدگی مضاعفی ایجاد می‌کند. پرسش اساسی در چنین شرایطی آن است که چگونه می‌توان مدلی کارآمد برای تعیین کلاس داده‌های جریانی و غیرقطعی ارایه داد که ضمن توانایی به‌روزرسانی در طول زمان، دقت و پایداری خود را نیز حفظ کند.

در سال‌های اخیر، توجه پژوهشگران به توسعه چارچوب‌هایی جلب شده است که توانایی مواجهه با پویایی داده، تغییر مفهوم[1] و عدم قطعیت را دارند. مطالعات جدیدی دیگری نشان داده‌اند که در داده‌های جریانی، استفاده از مدل‌های انطباق‌پذیر برای مدیریت تغییر رفتار داده و حفظ دقت طبقه‌بندی ضروری است [۴]. همچنین در پژوهش‌های دیگر بر اهمیت طراحی الگوریتم‌هایی برای طبقه‌بندی جریان‌های داده با حداقل فرض‌های آماری و حداکثر انطباق‌پذیری تاکید کرده‌اند [۵].

از سوی دیگر، استفاده از برنامه‌ریزی ریاضی و به‌ویژه تحلیل پوششی داده‌ها[2] (DEA) به‌عنوان ابزاری مناسب پیشنهاد شده است. روش DEA که اساسا برای اندازه‌گیری کارایی نسبی واحدهای تصمیم‌گیری[3] (DMU) توسعه یافت [۶] به دلیل ماهیت خطی و قابلیت انطباق‌پذیری، پتانسیل بالایی در دسته‌بندی داده‌ها دارد [۷]. در سال‌های اخیر نیز این روش توسعه یافته و پژوهش‌های متعددی نشان داده‌اند که DEA قادر است با داده‌های نامطمئن، بازه‌ای و خاکستری کار کند [۸، ۹، ۱۰].

در حوزه داده‌های حجیم و کلان‌داده نیز، مطالعات جدید نشان می‌دهند که DEA می‌تواند در کنار روش‌های داده‌کاوی برای تحلیل جریان‌های عظیم داده مورد استفاده قرار گیرد. نشان داد که شبکه‌های DEA(NDEA) و مدل‌های مبتنی بر داده‌های پیوسته می‌توانند نقش مؤثری در تحلیل کارایی در محیط‌های کلان‌داده داشته باشند [۱۱]. در مطالعات داخلی در بررسی ترکیب داده‌های بازه‌ای و گسسته به توسعه مدل‌های DEA در شرایط داده‌های نادقیق پرداخته شده است [۱۲].

همچنین مطالعات سال ۲۰۲۴ در حوزه‌ی دسته‌بندی با داده‌های غیرقطعی و مدل‌های DEA مقاوم تحت عدم‌قطعیت بودجه‌ای[۸، ۱۳] بیانگر حرکت DEA به سمت کاربری گسترده‌تر در شرایط داده‌های پیچیده و خاکستری است. در ادبیات داخلی، رویکردهای استوار تحلیل پوششی داده‌ها نیز برای مواجهه با عدم‌قطعیت

[1] Concept drift
[2] Data Envelopment Analysis (DEA)
[3] Decision Making unit (DMU)

مطرح شده‌اند DEA را تحت داده‌های مبتنی بر سناریو بسط داده‌اند. [۱۴].

از سوی دیگر، موضوعاتی مانند تشخیص تغییر مفهوم[1]، مدیریت جریان‌های داده و تشخیص ناهنجاری نیز به طور فزاینده‌ای اهمیت یافته‌اند. برای مثال، شیا و همکاران در سال ۲۰۲۳ چارچوبی برای طبقه‌بندی داده‌های جریانی همراه با مدیریت تغییر مفهوم ارایه داده‌اند [۱۵]؛ رویکردی که می‌تواند پایداری مدل‌های دسته‌بندی را در محیط‌های پویای داده ارتقا دهد. همچنین، در حوزه سنجش از دور، پژوهشگران نشان داده‌اند که ترکیب یادگیری ماشین و تحلیل داده‌های جریانی می‌تواند دقت شناسایی پوشش گیاهی را بهبود دهد [۱۶].

مرور ادبیات نشان می‌دهد که اگرچه در هر یک از حوزه‌های داده‌های جریانی، داده‌های خاکستری، تحلیل پوششی داده‌ها و تصمیم‌گیری چندشاخصه پیشرفت‌های قابل توجهی صورت گرفته است، اما ترکیب این چهار حوزه در یک چارچوب واحد برای طبقه‌بندی داده‌های جریانی خاکستری همچنان یک خلأ پژوهشی محسوب می‌شود. پژوهش‌های موجود یا تنها به توسعه مدل‌های DEA برای داده‌های نامطمئن پرداخته‌اند یا به روش‌های MADM برای انتخاب گزینه‌ها، اما رویکردی که از توان مشترک این دو در تعیین مرزهای دسته‌بندی و انتخاب آستانه طبقه‌بندی استفاده کند، تاکنون ارایه نشده است.

بر همین اساس، هدف این تحقیق ارایه یک روش جدید برای دسته‌بندی داده‌های جریانی خاکستری با بهره‌گیری از تحلیل پوششی داده‌ها و الگوریتم جایگشتی تصمیم‌گیری چندشاخصه است. در این روش، داده‌های ماهواره‌ای لندست پس از تبدیل به داده‌های کسری، وارد چارچوب DEA می‌شوند و آستانه مناسب برای تفکیک دو کلاس «دارای پوشش گیاهی» و «فاقد پوشش گیاهی» با استفاده از روش MADM تعیین می‌گردد. صحت مدل ابتدا با داده‌های شبیه‌سازی‌شده و سپس با داده‌های واقعی پوشش گیاهی ارزیابی شده است تا کارایی آن در مواجهه با شرایط غیرقطعی و متغیر سنجیده شود.

۲ مواد و روش‌ها

با بررسی ادبیات مربوط به استفاده از تحلیل پوششی داده‌ها در دسته‌بندی، می‌توان دریافت که تحقیقات محدودی در این حوزه انجام شده است. بخشی از این تحقیقات به معرفی این رویکرد جدید پرداخته‌اند و بخشی دیگر به کاربرد آن برای داده‌های غیرقطعی توجه کرده‌اند. بر اساس بررسی‌های انجام‌شده، از مدل برنامه‌ریزی ریاضی مرتبط با DEA در دسته‌بندی داده‌های جریانی استفاده نشده است و هدف این تحقیق، پرداختن به این موضوع می‌باشد. برای این منظور، یک رویکرد جدید باید پیشنهاد شود تا بتواند مدل برنامه‌ریزی ریاضی جدید را برای دسته‌بندی داده‌های جریانی به‌کار گیرد. این مکانیزم باید قادر باشد با کنترل ورود و خروج داده‌های مؤثر بر مرز کارایی در طول زمان، بستر مناسب برای بهینه‌سازی را فراهم کند. در این خصوص می‌توان یک روش خلاصه‌سازی (فشرده‌سازی) داده‌ها، مانند روش‌های خوشه‌بندی[2] [۱۷]، پیشنهاد داد. این روش از یک سو بردارهای داده ورودی را دریافت می‌کند و آن‌ها را در قالب بردارهایی که می‌توانند مراکز خوشه‌ها باشند،

[1] Concept drift
[2] Clustering

خلاصه می‌نماید و برای در نظر گرفته شدن در دسته‌بندی ارسال می‌کند [۱۸، ۱۹]. فرض می‌شود هدف از مطالعه، شناسایی محدوده یک دسته به منظور انجام دسته‌بندی باشد. اگر هر داده به‌عنوان یک DMU در نظر گرفته شود، مقادیر مشخصه‌های هر داده به‌عنوان ورودی DMU و عدد یک به‌عنوان تنها خروجی آن در نظر گرفته می‌شوند.

با حل یک مساله DEA می‌توان داده‌هایی را که به‌عنوان نقاط مرزی شناخته می‌شوند، برای ترسیم محدوده دسته به‌کار برد و سپس از این محدوده‌ها برای پیش‌بینی دسته داده‌های جدید استفاده کرد. به عبارت دیگر، اگر n داده $x_i=(x_{i1},x_{i2},\ldots,x_{im})$ که $i=1,2,\ldots,n$ در یک دسته قرار داشته باشند، می‌توان با حل مجموعه مسئله‌های برنامه‌ریزی خطی در قالب یک مساله DEA، نقاط مرزی را شناسایی کرده و محدوده دسته را تعیین نمود.

مطالعات جدید نشان می‌دهد که DEA در حوزه داده‌های جریانی و غیرقطعی نیز قابلیت توسعه دارد و همین ویژگی امکان استفاده از آن در دسته‌بندی داده‌های خاکستری را فراهم می‌کند [۸-۱۰]

در این تحقیق، هدف توسعه روشی جدید برای دسته‌بندی داده‌های جریانی کسری با ترکیب روش‌های تصمیم‌گیری چندمعیاره و تحلیل پوششی داده‌ها است. داده‌های مورد استفاده در این مطالعه، داده‌های آموزشی هستند که به‌منظور سهولت پردازش توسط پژوهشگران ایجاد می‌شوند و می‌توانند به صورت جریانی، فازی، خاکستری یا کسری باشند. پس از ارزیابی روش پیشنهادی بر روی داده‌های آموزشی، کارایی آن بر روی داده‌های واقعی مورد بررسی قرار خواهد گرفت. داده‌های واقعی مورد مطالعه شامل تصاویر ماهواره‌ای هستند که برای هر پیکسل، مقداری عددی مرتبط با میزان روشنایی در بازه ۰ تا ۲۵۵ دارند. با تقسیم میزان روشنایی هر پیکسل بر ۲۵۵، مقدار روشنایی به صورت نرمال‌شده و کسری در بازه صفر تا یک نمایش داده خواهد شد.

استفاده از داده‌های کسری، که نوعی داده خاکستری به شمار می‌روند، امکان مدیریت عدم‌قطعیت و نوسانات ناشی از خطاهای سنجنده را فراهم می‌کند و همچنین موجب استانداردسازی ورودی‌های مدل می‌شود. این فرایند نرمال‌سازی یکی از پیش‌نیازهای مهم برای اجرای مدل DEA در شرایط داده‌های جریانی است، زیرا باعث مقیاس‌پذیری و مقایسه‌پذیری مقادیر در طول زمان می‌شود [۴].

۲-۱ تحلیل پوششی داده‌ها

در این تحقیق، برای به‌دست آوردن میزان کارایی و تعیین حدود تفکیک داده‌ها، از مدل تحلیل پوششی داده‌ها استفاده شده است. در چندین سال گذشته، در اکثر کشورهای جهان، کاربردهای مختلف DEA برای ارزیابی عملکرد نهادها و فعالیت‌های متنوع در زمینه‌های گوناگون مشاهده شده است و پژوهش‌های متعدد نشان داده‌اند که این روش قابلیت انطباق با مسائل مختلف ازجمله طبقه‌بندی، خوشه‌بندی و تحلیل داده‌های نامطمئن را دارد. علت مقبولیت گسترده روش تحلیل پوششی داده‌ها نسبت به سایر روش‌ها، امکان بررسی روابط پیچیده و اغلب نامعلوم بین چندین ورودی و چندین خروجی (که معمولا اندازه‌پذیر نیستند) است. این روش در حوزه‌های گوناگون، از ارزیابی عملکرد شهرها و سازمان‌ها گرفته تا تحلیل کارایی واحدهای تولیدی، به کار گرفته شده و نشان داده است که قادر است مرزهای کارایی را استخراج و وابستگی‌های میان ورودی و خروجی‌ها را مدل‌سازی

کند.

در این پژوهش نیز از همین قابلیت برای شناسایی نقاط مرزی دسته استفاده شده است. داده‌ها در قالب DMU مدل شده‌اند، به‌گونه‌ای که هر پیکسل یا نمونه داده یک واحد تصمیم‌گیری با مجموعه‌ای از ورودی‌ها و خروجی‌ها در نظر گرفته می‌شود. با حل مدل DEA، داده‌هایی که بر روی مرز کارایی قرار می‌گیرند شناسایی شده و از آن‌ها برای تعیین محدوده دسته‌ها و مرز جدایش داده‌های جدید استفاده می‌شود.

تجزیه‌وتحلیل پوششی داده‌ها که نخستین بار توسط چارنز و کوپر توسعه یافت [۶]، امروزه علاوه بر حوزه‌های مالی و اجرایی، در مسائل مرتبط با داده‌های جریانی، خاکستری و تصمیم‌گیری چندمعیاره نیز کاربرد گسترده‌ای یافته است و ابزار مؤثری برای بهبود تصمیم‌گیری و تحلیل ساختار داده‌ها محسوب می‌شود.

۲-۲ مبانی نظری و مفهوم‌سازی تحلیل پوششی داده‌ها

تحلیل پوششی داده‌ها یک روش برنامه‌ریزی ریاضی برای ارزیابی واحدهای تصمیم‌گیرنده است. منظور از «واحد تصمیم‌گیرنده» یک واحد سازمانی یا سازمان مجزا است که توسط فردی به نام مدیر، رئیس یا مسئول اداره می‌شود، به شرط آنکه این واحد دارای فرآیند سیستمی باشد. با توجه به اینکه سیستم مورد نظر می‌تواند شامل سیستم‌های تولیدی یا خدماتی، انتفاعی یا غیرانتفاعی، دولتی یا غیردولتی باشد، در ادبیات تحلیل پوششی داده‌ها به منظور جلوگیری از پراکنده‌کاری، به جای عوامل ورودی از مفهوم «نهاده[1]» و به جای محصولات خروجی سیستم از مفهوم «ستاده[2]» استفاده می‌شود.

تحلیل پوششی داده‌ها در ارزیابی واحدهای تصمیم‌گیرنده بر این فرض استوار است که این واحدها نهاده‌های مشابهی را برای تولید ستاده‌های مشابه به کار می‌گیرند. از این‌رو، این تکنیک ابزاری مناسب برای اندازه‌گیری کارایی نسبی واحدهایی است که دارای چندین ورودی و خروجی مشابه‌اند. همین ویژگی باعث شده است که در سال‌های اخیر، DEA علاوه‌بر حوزه‌های سنتی ارزیابی عملکرد، در تحلیل داده‌های نامطمئن، فازی، بازه‌ای و حتی جریانی نیز به‌کار گرفته شود [۸، ۹].

در علم اقتصاد، رابطه‌ی بین نهاده و ستاده‌ی یک سیستم با تابع تولید بیان می‌شود. این تابع نشان‌دهنده‌ی بیشترین میزان ستاده‌ای است که می‌توان از ترکیب‌های مختلف نهاده‌ها به دست آورد. در واقع، تلاش برای تابعی‌سازی رابطه‌ی نهاده‌ها و ستاده‌ها و تعیین حداکثر ستاده‌ی قابل حصول از نهاده‌ها، به معرفی توابع تولید پارامتری در مطالعات اقتصادی انجامید. توابعی مانند کاب-داگلاس[3]، لئونتیف[4] و کشش ثابت[5] از جمله نمونه‌هایی‌اند که در نظریه‌های اقتصاد خرد با همین انگیزه مطرح شده‌اند.

چنانچه این تابع مشخص و در دسترس باشد، بیان نسبت‌های نهاده-ستاده برای محاسبه‌ی «کارایی سیستم» ساده خواهد بود. کارایی هر واحد بر اساس فاصله‌ی آن از مرز مجموعه‌ی امکان تولید سنجیده می‌شود. شیوه‌ی

[1] Input
[2] Output
[3] Cobb Douglas
[4] Lean Tyef
[5] Strength Constant

نزدیک شدن هر واحد تصمیم‌گیرنده به این مرز، که اصطلاحا «حرکت» نامیده می‌شود، مبنای شکل‌گیری مدل‌های گوناگون ارزیابی است. اما در مساله طبقه‌بندی داده‌ها، از جمله داده‌های جریانی خاکستری در این پژوهش، تابع تولید مشخصی وجود ندارد و روابط میان نهاده و ستاده ناشناخته یا غیرخطی‌اند.

با این حال، در بسیاری از موارد تابع تولید در دسترس نیست و ارایه‌ی آن پیچیده یا حتی غیرممکن است. در چنین شرایطی، تحلیل پوششی داده‌ها با استفاده از مجموعه‌ای از مشاهدات، یک تابع تولید تجربی بر اساس داده‌های واقعی می‌سازد. نتیجه‌ی این فرایند، تابعی مرزی است که تمام داده‌ها را در بر می‌گیرد؛ از این رو به آن «تحلیل پوششی» یا «تحلیل فراگیر» گفته می‌شود. این ویژگی کلیدی همان چیزی است که DEA را برای مساله حاضر مناسب می‌سازد، زیرا می‌توان مرزهای کارایی را به عنوان مرزهای جدایش دسته‌ها در نظر گرفت و داده‌های جدید را بر اساس فاصله‌شان از این مرزها دسته‌بندی کرد [۱۱، ۱۶].

از سوی دیگر، چون تحلیل پوششی داده‌ها بر پایه‌ی مجموعه‌ای از مسایل بهینه‌سازی استوار است و در این مسایل هیچ پارامتری برای برآورد وجود ندارد، این روش در زمره‌ی روش‌های «غیرپارامتریک» قرار می‌گیرد. بنابراین، تحلیل پوششی داده‌ها یک روش برنامه‌ریزی خطی به شمار می‌رود که با استفاده از اطلاعات مربوط به سازمان‌ها و واحدهای تولیدی، به عنوان واحدهای تصمیم‌گیرنده، یا در این پژوهش، با استفاده از ویژگی‌های هر پیکسل یا داده به‌عنوان یک DMU، مرز کارایی را ترسیم می‌کند. این مرز با تکیه بر داده‌های نهاده و ستاده و بر اساس نتایج حاصل از حل متوالی مسایل برنامه‌ریزی خطی ساخته می‌شود. در واقع، درجه‌ی ناکارایی هر واحد تصمیم‌گیرنده به میزان فاصله‌ی آن از مرز کارایی بستگی دارد و همین فاصله، در این تحقیق، مبنای تعریف حدود دسته‌ها و ساخت مدل طبقه‌بندی است.

۲-۳ مدل‌سازی مجموعه امکان تولید با بازده به مقیاس متغیر

در این بخش، مدل‌سازی مجموعه امکان تولید[1] (PPS) با بازده به مقیاس متغیر ارایه می‌شود. یک مجموعه با n واحد تصمیم‌گیرنده را در نظر بگیرید که هر واحد DMU برای $(J=1,\dots,n)$ از m ورودی x_{ij} $(i=1,\dots,m)$ برای تولید s خروجی y_{rj} $(r=1,\dots,s)$ استفاده می‌کند.

فرض کنید بردارهای ورودی و خروجی برای هر DMU_j که به صورت (x_j , y_j) نمایش داده شود، به گونه‌ای که تمام مولفه‌های این بردارها نامنفی باشند و در هر بردار، حداقل یک مولفه‌ی ورودی و یک مؤلفه‌ی خروجی مقدار مثبت داشته باشد. مجموعهٔ تمامی (x,y)های شدنی را مجموعه‌ی امکان تولید نامیده شده و با T نشان می‌دهند که به صورت "T= {(x,y) |x می‌تواند خروجی y را تولید کند}" تعریف می‌شود [۲۰].

مجموعه‌ی امکان تولید با بازده به مقیاس متغیر (VRS) به صورت زیر تعریف می‌شود:

$$T_v=\{(x,y)\mid x\ge\sum_{j=1}^{n}x_j\lambda_j, y\le\sum_{j=1}^{n}y_j\lambda_j, \sum_{j=1}^{n}\lambda_j=1, \lambda_j\ge 0, j=1,\dots,n\}. \quad (1)$$

این تعریف بیان می‌کند که هر ترکیب محدب از واحدهای موجود می‌تواند معیاری برای کارایی یک واحد

[1] Production Possibility Set

جدید باشد. شرط $\sum_{j=1}^{n}\lambda_j = 1$ مشخص‌کننده‌ی بازده به مقیاس متغیر (VRS) است و باعث می‌شود مرز کارایی نسبت به مقیاس حساس باشد. در این پژوهش، همین تعریف به‌عنوان مبنای شناسایی مرزهای جداسازی دسته‌ها در داده‌های جریانی خاکستری استفاده شده است. با تشکیل مجموعه امکان تولید، می‌توان تشخیص داد کدام داده‌ها بر روی مرز کارایی قرار می‌گیرند و بنابراین، نقش «نقاط مرزی» را در فرایند دسته‌بندی ایفا می‌کنند.

تحلیل پوششی داده‌ها برای سنجش کارایی، نیازمند تعریف و توسعه‌ی «مجموعه امکان تولید است. این مجموعه شامل تمامی واحدهای تصمیم‌گیرنده‌ای است که تولید آن‌ها امکان‌پذیر باشد؛ به عبارت دیگر، واحدی با بردار ورودی x زمانی در مجموعه امکان تولید قرار می‌گیرد که قادر به تولید خروجی y باشد. در این تحقیق نیز، هر داده یا پیکسل به‌عنوان یک DMU در نظر گرفته شده و بر اساس فاصله‌ی آن از مرز مجموعه امکان تولید، موقعیت آن در دسته تعیین می‌شود.

۲-٤ مدل CCR در تحلیل پوششی داده‌ها

مدل CCR نخستین الگوی تحلیل پوششی داده‌ها است که نام آن از حروف آغازین نام ابداع‌کنندگانش، یعنی چارنز، کوپر و رودز گرفته شده است. در این مدل، برای تعیین بیشترین نسبت کارایی و نیز دخالت دادن مقادیر نهاده‌ها و ستاده‌های سایر واحدهای تصمیم‌گیرنده در محاسبه‌ی اوزان بهینه‌ی واحد مورد بررسی، مدل پایه‌ی زیر ارایه شده است:

$$Max: \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \qquad (۲)$$

$$s.t.$$

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \le 1, j = 1,2,\ldots,n$$

$$u_r \ge 0 , v_i \ge 0$$

مدل برنامه‌ریزی کسری فوق به عنوان مدل کسری CCR شناخته می‌شود. در این مدل:

- u_r: وزن مربوط به ستاده‌ی r ام ،
- v_i وزن مربوط به نهاده i ام ،
- O: شاخص واحد تصمیم‌گیرنده تحت بررسی ($o \in \{1,2,\ldots,n\}$)
- y_{ro} و x_{io}: به ترتیب، مقادیر ستاده r ام و نهادهٔ i ام برای واحد تحت بررسی (واحد o)
- y_{rj} و x_{ij}: به ترتیب، مقادیر ستاده r ام و مقدار نهاده i ام برای واحد j ام هستند
- s، تعداد ستاده‌ها،
- m، تعداد نهاده‌ها
- n، تعداد کل واحدهای تصمیم‌گیرنده

تعریف کارایی در مدل کسری CCR به صورت زیر است:

(۳)

$$\text{کارایی} = \frac{\text{ترکیب وزنی ستاده‌ها}}{\text{ترکیب وزنی نهاده‌ها}}$$

با پیشنهاد چارنز و کوپر، شرط نرمال‌سازی زیر بر مدل برنامه‌ریزی کسری CCR اعمال شد:

(۴)

$$\sum_{i=1}^{m} v_i x_{jo} = 1$$

در مدل برنامه‌ریزی کسری CCR، این مدل به مدل برنامه‌ریزی خطی زیر تبدیل شد:

(۵)

$$Max \sum_{r=1}^{s} u_r y_{ro}$$

$$s.t. \sum_{i=1}^{m} v_i x_{io} = 1$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0 \ , j = 1,2,\ldots,n$$

$$u_r \ge 0 \ , v_i \ge 0$$

این مدل، مدل مضربی CCR ورودی‌محور (CCR-I) نامیده می‌شود. با این حال، برای تبدیل مدل کسری CCR به یک مدل برنامه‌ریزی خطی، می‌توان از رویکرد دیگری نیز بهره گرفت. در این روش، با اعمال محدودیت زیر:

$$\sum_{r=1}^{s} u_r y_{ro} = 1$$

و تبدیل به شکل زیر است که همان مدل مضربی CCR خروجی‌محور (CCR-O) با همان محدودیت‌های قبل است:

(۶)

$$Min \sum_{i=1}^{m} v_i x_{io}$$

در این پژوهش، از ساختار CCR برای استخراج مرزهای کارایی و تعیین نقاط مرزی دسته‌ها در داده‌های جریانی استفاده شده است. مدل CCR به دلیل ساختار خطی‌اش امکان پیاده‌سازی سریع در محیط داده‌های جریانی و خاکستری را فراهم می‌کند.

۲-۵ کاربرد سنجش از دور در بررسی پوشش گیاهی

شناخت ویژگی‌های پوشش‌های گیاهی و بررسی روابط میان گونه‌های مختلف گیاهی و عوامل محیطی همواره مورد توجه بوم‌شناسان بوده است. علت این توجه، اهمیت بالای پوشش گیاهی در ابعاد مختلف از جمله نقش آن در زیستگاه‌ها، تولید انرژی و دیگر خصوصیات حیاتی گیاهان بر روی کره زمین است. با این حال، پوشش‌های گیاهی در اثر عوامل طبیعی یا انسانی و در گذر زمان دچار تغییر می‌شوند که این امر شرایط و عملکرد اکوسیستم‌ها را تحت تأثیر قرار می‌دهد. همچنین، ارتباط میان پوشش گیاهی و خشکسالی با پدیده‌هایی نظیر وقوع

گرد و غبار در هر منطقه از اهمیت ویژه‌ای برخوردار است. این پویایی زمانی پوشش گیاهی به‌ویژه در داده‌های جریانی سنجش‌ازدور اهمیت پیدا می‌کند، زیرا تغییرات مداوم در سری‌های زمانی NDVI یکی از مهم‌ترین شاخص‌ها برای آشکارسازی تغییرات محیطی محسوب می‌شود [۱۳].

کسب اطلاعات دقیق درباره میزان و پراکنش پوشش‌های گیاهی اهمیت زیادی دارد. برای کاهش اثر عوامل ناخواسته در داده‌های مربوط به پوشش گیاهی و بهبود دقت این اطلاعات، می‌توان با ترکیب حداقل دو باند طیفی تصاویر ماهواره‌ای شاخص‌های مرکبی تحت عنوان شاخص‌های پوشش گیاهی ایجاد کرد. یکی از اهداف اصلی استفاده از این شاخص‌ها، پیش‌بینی و ارزیابی درصد پوشش گیاهی در یک منطقه است. در مطالعات جدید، این شاخص‌ها به‌صورت داده‌های زمانی پیوسته ثبت شده و برای تحلیل‌های جریانی[1] به‌کار می‌روند، امری که با موضوع این پژوهش نیز همسو است.

شاخص‌های پوشش گیاهی به طور گسترده به عنوان معیاری برای تحلیل تغییرات کاربری اراضی، پایش پوشش گیاهی و بررسی عوامل مرتبط مورد استفاده قرار می‌گیرند. شاخص‌های متنوعی مانند RVI، NDVI، SAVI، EVI و غیره برای این منظور معرفی شده‌اند. در این میان، شاخص تفاضلی نرمال‌شده پوشش گیاهی (NDVI) یکی از پرکاربردترین ابزارها برای پایش تغییرات پوشش گیاهی به شمار می‌رود. این شاخص از نسبت‌گیری بین باند قرمز و باند مادون‌قرمز نزدیک تصاویر ماهواره‌ای محاسبه می‌شود و سال‌هاست که به عنوان معیار استاندارد جهانی برای ارزیابی وضعیت پوشش گیاهی مورد استفاده قرار می‌گیرد.

مقادیر NDVI در بازه‌ی ۱- تا ۱+ قرار دارند. مقادیر نزدیک به ۱+ نشان‌دهنده‌ی پوشش گیاهی متراکم و سالم، و مقادیر نزدیک به ۱- معرف مناطق فاقد پوشش یا دارای تراکم گیاهی بسیار ضعیف هستند. محاسبه‌ی این شاخص از طریق رابطه‌ی ۷ امکان‌پذیر است:

$$\mathrm{NDVI} = \mathrm{NIR} - \mathrm{R}/\mathrm{NIR} + \mathrm{R} \tag{۷}$$

که در این رابطه NIR، میزان انعکاس تصویر در باند مادون قرمز نزدیک و R، میزان انعکاس تصویر در باند قرمز می‌باشد.

شاخص NDVI یکی از شناخته‌شده‌ترین، ساده‌ترین و کاربردی‌ترین شاخص‌ها در مطالعات پوشش گیاهی است. این شاخص از فرآیند محاسباتی ساده‌ای برخوردار بوده و در مقایسه با دیگر شاخص‌ها، بهترین توان دینامیکی را ارایه می‌دهد. NDVI بیشترین حساسیت را نسبت به تغییرات پوشش گیاهی داشته و در عین حال، به جز در شرایطی که تراکم پوشش گیاهی بسیار کم باشد، نسبت به اثرات جوی و ویژگی‌های زمینه‌ای خاک حساسیت کمتری نشان می‌دهد. همین ویژگی‌ها موجب شده‌اند که NDVI به‌عنوان داده‌ی ورودی اصلی در مدل‌های طبقه‌بندی جریانی، از جمله مدل پیشنهادی مبتنی بر DEA در این تحقیق، انتخاب شود؛ زیرا نوسانات زمانی NDVI می‌توانند شاخصی معتبر برای تشخیص تغییر، دسته‌بندی وضعیت پوشش گیاهی و شناسایی الگوهای مرزی باشند.

[1] Stream Data Analysis

۳ الگوریتم پیشنهادی و شبیه‌سازی

۳-۱مدل‌سازی و پیاده‌سازی روی داده‌ی آزمایشی

برای پیاده‌سازی مدل پیشنهادی و ارزیابی عملکرد آن پیش از اعمال بر داده‌های واقعی، ابتدا مجموعه‌ای از داده‌ی آزمایشی به صورت تصادفی تولید شد. این داده‌ها شامل مقادیر روشنایی پیکسل‌ها در دو باند ۳ و ۴ تصاویر ماهواره‌ای لندست بودند. شکل ۱ داده‌های تولید شده برای باند ۳ و شکل ۲ داده‌ها متناظر برای باند ۴ را نمایش می‌دهد. مقادیر هر پیکسل در این ماتریس‌ها در بازه‌ی ۰ تا ۲۵۵ قرار دارند و بیانگر شدت بازتاب طیفی در دو باند مذکور هستند.

۱۸۲	۲۴۲	۲۰۵	۸۹	۱۸
۱۹۰	۷۵	۲۳۷	۲۹	۲۳۴
۳۹	۲۲۸	۱۰۶	۸	۲۴۲
۲۶	۷۱	۲۱۸	۱۵۶	۵۹
۱۴۳	۱۷۷	۱۶۷	۱۴۷	۱۰۲

شکل ۱. داده‌ی آزمایشی تولیدشده به صورت تصادفی برای میزان روشنایی در باند ۳

۶۷	۲۵۴	۱۳	۱۷۲	۸۸
۱۶۲	۳۰	۲۴۰	۲۳۷	۴۳
۲۷	۲۵۲	۱۸۶	۱۱۴	۱۸۴
۷۷	۹۹	۲۲۰	۸۳	۱۳۰
۲۱۲	۶۷	۲۴۶	۵۳	۱۲۲

شکل ۲. داده‌ی آزمایشی تولیدشده به صورت تصادفی برای میزان روشنایی در باند ۴

مطابق مباحث مطرح‌شده در بخش پیشین، برای آشکارسازی پوشش گیاهی از شاخص نرمال‌شده تفاضلی پوشش گیاهی (NDVI) استفاده می‌شود. این شاخص برای هر پیکسل بر مبنای رابطه‌ی (۱۲) محاسبه گردید:

$$\mathrm{NDVI} = \frac{B_۴ - B_۳}{B_۴ + B_۳} \qquad (۱۲)$$

در این رابطه، $B_۳$ و $B_۴$ به‌ترتیب بیانگر مقدار روشنایی پیکسل در باند قرمز و مادون‌قرمز نزدیک هستند. چند نکته اساسی درباره این شاخص از منظر تحلیل پوششی داده‌ها (DEA) قابل توجه است:

۱. محدوده‌ی مقادیر NDVI: مقادیر NDVI همواره در بازه‌ی صفر تا یک قرار می‌گیرند و ماهیت کسری و نرمال‌شده دارند. این موضوع رویکرد DEA مبتنی بر داده‌های کسری را توجیه‌پذیرتر می‌سازد.

۲. تفسیر NDVI در چارچوب DEA: در چارچوب DEA، مخرج کسر ($B_۴ + B_۳$) به عنوان ورودی‌ها و صورت کسر ($B_۴ - B_۳$) به عنوان خروجی‌ها در نظر گرفته می‌شود. پیکسل‌هایی که مقدار خروجی نسبت به ورودی آن‌ها بیشتر باشد، دارای کارایی بالاتر و از نظر پوشش گیاهی غنی‌تر تلقی می‌شوند.

۳. فرایند خوشه‌بندی پیکسل‌ها: با محاسبه NDVI برای ماتریس آموزشی ۵×۵، تصویر شاخص پوشش گیاهی مطابق شکل ۳ به دست آمد. در این ماتریس، پیکسل‌هایی با مقدار NDVI بالاتر، پوشش گیاهی مناسب‌تری را نشان می‌دهند. در این تصویر، میزان روشنایی هر پیکسل به صورت عددی در بازه‌ی ۰ تا ۱ نمایش داده شده است.

۰/۴۶	۰/۰۲	۰/۸۸	۰/۳۲	۰/۶۶
۰/۰۸	۰/۴۳	۰/۰۱	۰/۷۸	۰/۶۹
۰/۱۸	۰/۰۵	۰/۲۷	۰/۸۷	۰/۱۴
۰/۵۰	۰/۱۶	۰/۰۱	۰/۳۱	۰/۳۸
۰/۱۹	۰/۴۵	۰/۱۹	۰/۴۷	۰/۰۹

شکل ۳. تصویر شاخص نرمال‌شده تفاضلی پوشش گیاهی برای داده‌های آزمایشی

۴. تشکیل DMUها و استخراج ورودی/خروجی‌ها: برای هر یک از ۲۵ پیکسل، مقادیر ورودی و خروجی DEA مطابق توضیحات فوق محاسبه و در قالب جدول شکل ۴ سامان‌دهی شد. در این مرحله، مسئله به صورت یک مدل تک‌ورودی-تک‌خروجی تعریف می‌شود و هر پیکسل (DMU) بر اساس نسبت ورودی و خروجی مربوطه در مدل CCR تحلیل می‌گردد.

۱۱۵	۱۲	۱۹۲	۸۳	۷۰
۲۸	۴۵	۳	۲۰۸	۱۹۱
۱۲	۲۴	۸۰	۱۰۶	۵۸
۵۱	۲۸	۲	۷۳	۷۱
۶۹	۱۱۰	۷۹	۹۴	۲۰

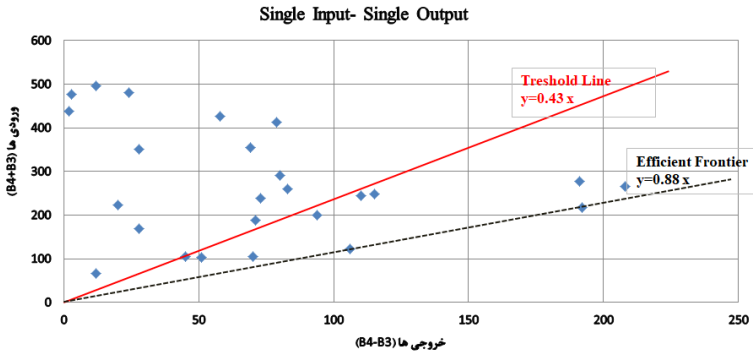
(الف) ورودی‌ها

۲۴۹	۴۹۶	۲۱۸	۲۶۱	۱۰۶
۳۵۲	۱۰۵	۴۷۷	۲۶۶	۲۷۷
۶۶	۴۸۰	۲۹۲	۱۲۲	۴۲۶
۱۰۳	۱۷۰	۴۳۸	۲۳۹	۱۸۹
۳۵۵	۲۴۴	۴۱۳	۲۰۰	۲۲۴

(ب) خروجی‌ها

شکل ٤. مقادیر ورودی ها و خروجی ها برای هر DMU

۵. اجرای مدل CCR و استخراج مرز کارایی: در گام بعدی، مدل CCR ورودی‌محور برای ۲۵ واحد تصمیم‌گیرنده اجرا شد. نتایج حاکی از تشکیل یک مرز کارایی محدب برای پیکسل‌ها است. شکل ۵ موقعیت DMUها را نمایش داده و مرز کارایی با رنگ مشکی در نمودار مشخص شده است. مرز مذکور بیانگر رفتاری است که در آن ترکیب ورودی-خروجی پیکسل‌ها بهینه بوده و مبنای تفکیک پیکسل‌های با پوشش گیاهی قوی از ضعیف قرار می‌گیرد.



شکل ۵. دیاگرام بیست‌وپنج واحد در مدل CCR الگوریتم تحلیل پوششی داده‌ها

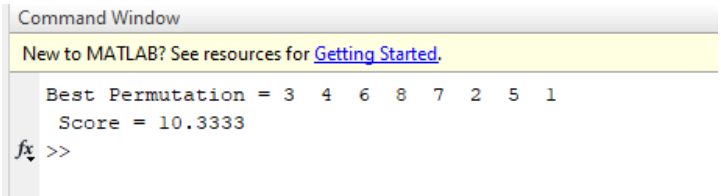
۶. تعیین حد آستانه[1] با استفاده از مدل جایگشتی: به‌منظور تعیین خط جدایش میان دو خوشه (پیکسل‌های دارای پوشش گیاهی مناسب و نامناسب)، مجموعه‌ای از خطوط با شیب‌های مختلف تولید و ارزیابی شدند. این ارزیابی در قالب یک ماتریس تصمیم (جدول ۱) شامل دو معیار صورت گرفت:

- حداقل ۲۰ درصد پیکسل‌ها در خوشه‌ی پوشش گیاهی قرار گیرند
- حداکثر ۵۰ درصد پیکسل‌ها در این خوشه قرار گیرند

جدول ۱. ماتریس تصمیم برای انتخاب حد آستانه‌ی مناسب (جهت استفاده در روش جایگشتی)

شیب / وزن	۰/۱	۰/۲
	حداقل بیست درصد پوشش	حداکثر پنجاه درصد پوشش
۰ تا ۰/۲	۹	۷
۰/۲ تا ۰/۳	۸	۸
۰/۳ تا ۰/۴	۷	۹
۰/۴ تا ۰/۵	۶	۸
۰/۵ تا ۰/۶	۵	۸
۰/۶ تا ۰/۷	۲	۹
۰/۷ تا ۰/۸	۱	۹
۰/۸ تا ۰/۹	۱	۹

هر بازه شیب یک گزینه محسوب شد و برای ارزیابی گزینه‌ها از روش تصمیم‌گیری چندشاخصه به شیوه جایگشت استفاده گردید. وزن معیارها به ترتیب ۰/۱ و ۰/۲ لحاظ شد. خروجی MATLAB (شکل ۶) نشان می‌دهد که بهترین گزینه، خط با شیب ۰/۴ تا ۰/۵ است.



شکل ٦. خروجی نرم‌افزار متلب برای تصمیم‌گیری چندشاخصه به روش جایگشتی

۷. ترسیم خط حد آستانه و دسته‌بندی نهایی: خط حاصل‌شده با معادله تقریبا y=0.43x به‌عنوان خط آستانه‌ای روی نمودار DEA ترسیم شد (شکل ۵). پیکسل‌هایی که در پایین این خط قرار گرفتند، به‌عنوان خوشه‌ی دارای

[1] Tresholding

پوشش گیاهی طبقه‌بندی شدند و سایر پیکسل‌ها در خوشه‌ی فاقد پوشش قرار گرفتند. نتیجه دسته‌بندی برای داده آزمایشی در شکل ۷ ارایه شده است.

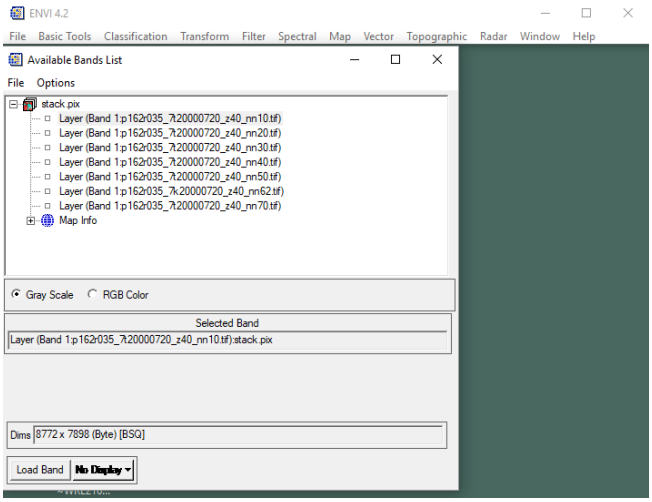
۰/۴۶	۰/۰۲	۰/۸۸	۰/۳۲	۰/۶۶
۰/۰۸	۰/۴۳	۰/۰۱	۰/۷۸	۰/۶۹
۰/۱۸	۰/۰۵	۰/۲۷	۰/۸۷	۰/۱۴
۰/۵۰	۰/۱۶	۰/۰۱	۰/۳۱	۰/۳۸
۰/۱۹	۰/۴۵	۰/۱۹	۰/۴۷	۰/۰۹

شکل ۷. نقشه‌ی نهایی پوشش گیاهی دسته‌بندی‌شده برای داده‌های آزمایشی

۳-۲ مدلسازی و پیاده‌سازی روی داده‌ی واقعی

در این تحقیق، داده‌ی ماهواره‌ای لندست از سنجنده‌ی TM مربوط به بخشی از محدوده‌ی شهر تهران به‌عنوان داده‌ی اصلی انتخاب شد. این داده دارای قدرت تفکیک مکانی ۳۰ متر و قدرت تفکیک زمانی ۴۲ روز است. ویژگی اخیر، یعنی برداشت مستمر داده از سطح زمین در بازه‌های زمانی منظم، خاصیت جریانی بودن داده‌ها در این تحقیق را تأیید می‌کند. به دلیل ماهیت زمانی-مکانی این داده‌ها، امکان تحلیل پویا و مقایسه وضعیت پوشش گیاهی در بازه‌های مختلف زمانی فراهم شد.

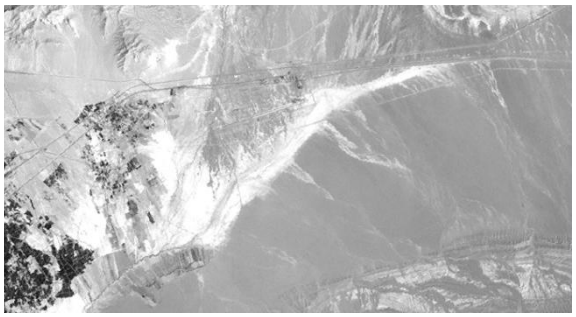
شکل ۸ تصویر هفت‌باندی از منطقه‌ی مورد مطالعه را در محیط نرم‌افزار پردازش داده‌های ماهواره‌ای ENVI نشان می‌دهد. استفاده‌ی هم‌زمان از این باندها به دلیل پوشش طیفی گسترده، امکان تحلیل دقیق‌تر پوشش گیاهی و تمایز بهتر میان سطوح مختلف کاربری زمین را فراهم می‌سازد. در این پژوهش، تمرکز اصلی بر باندهای ۳ و ۴ است؛ با این حال وجود سایر باندها در مرحله تفسیر و کنترل کیفیت داده‌ها نقش مهمی دارد.



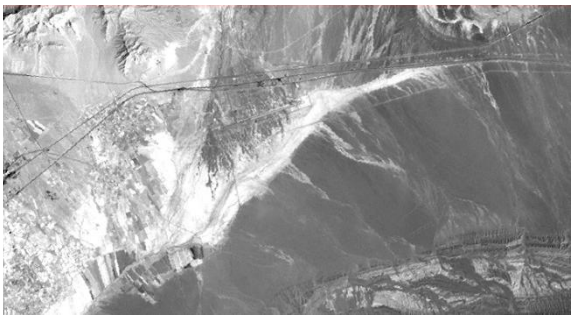
شکل ۸. تصویر هفت‌باندی از منطقه‌ی مورد مطالعه در محیط نرم‌افزار پردازش داده‌های ماهواره‌ای ENVI

برای تحلیل تفصیلی پوشش گیاهی، باند ۳ (قرمز) و باند ۴ (مادون‌قرمز نزدیک) که بیشترین حساسیت را به کلروفیل و بازتاب پوشش گیاهی دارند استخراج شد. شکل ۹ باند ۳ و شکل ۱۰ باند ۴ را نمایش می‌دهد. محدوده‌ی موردمطالعه شامل یک قاب با ابعاد ۸۰۰×۴۰۰ پیکسل است که مساحتی معادل ۱۲×۲۴ کیلومتر را پوشش می‌دهد. تعداد پیکسل‌های این تصویر برابر با ۳۲۰ هزار پیکسل بوده و هر پیکسل مشابه یک DMU در مدل تحلیل پوششی داده‌ها تلقی می‌شود.

اهمیت این دو باند به دلیل رفتار طیفی متفاوت آن‌هاست: باند ۳ (ناحیه‌ی قرمز طیف مرئی) بیشترین جذب را در پوشش گیاهی دارد و باند ۴ (مادون‌قرمز نزدیک) بیشترین بازتاب را از پوشش گیاهی نشان می‌دهد. همین تضاد طیفی میان دو باند، مبنای محاسبه‌ی شاخص NDVI است که امکان تفکیک سطوح مختلف پوشش گیاهی را فراهم می‌کند.

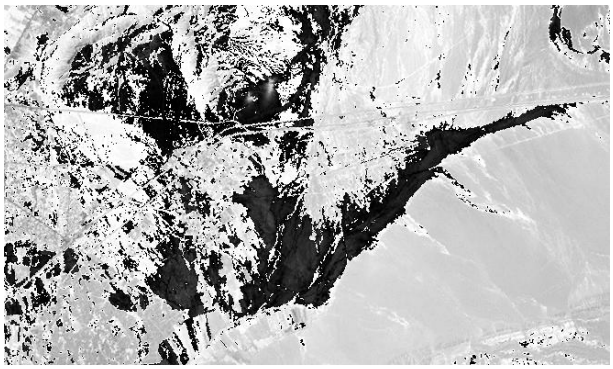


شکل ۹. نمایش باند ۳ تصویر ماهواره‌ای



شکل ۱۰. نمایش باند ۴ تصویر ماهواره‌ای

برای ایجاد تصویر اولیه پوشش گیاهی، شاخص NDVI برای تمام پیکسل‌ها محاسبه و نقشه‌ی دسته‌بندی‌نشده‌ی پوشش گیاهی مطابق شکل ۱۱ تولید شد. این نقشه وضعیت خام پوشش گیاهی را قبل از اعمال مدل خوشه‌بندی مبتنی بر DEA و حد آستانه‌ای نشان می‌دهد و در واقع گام اولیه برای مقایسه و تمایز بین مناطق دارای پوشش گیاهی و فاقد پوشش گیاهی محسوب می‌شود.



شکل ۱۱. نقشه‌ی دسته‌بندی‌نشده‌ی پوشش گیاهی

با توجه به شناخت میدانی از محدوده و به منظور انتخاب حد آستانه‌ای مناسب برای تفکیک دو خوشه «پوشش گیاهی» و «فاقد پوشش»، ماتریس تصمیم توسط کارشناس منطقه تهیه و وارد نرم افزار متلب شد (جدول ۲). لازم به ذکر است که به دلیل تعداد بسیار زیاد واحدها (۳۲۰ هزار پیکسل) ترسیم دیاگرام مشابه داده‌های آزمایشی برای مدل DEA امکان‌پذیر نبود و تحلیل به طور مستقیم بر اساس شاخص‌های تصمیم‌گیری

چندشاخصه انجام شد.

در ماتریس تصمیم، شیب‌های مختلف به‌عنوان گزینه و دو معیار کلیدی به‌عنوان شاخص ارزیابی تعریف شدند:

- حداقل ۲۰٪ پیکسل‌ها باید در خوشه‌ی دارای پوشش قرار گیرند
- حداکثر ۵۰٪ پیکسل‌ها باید در این خوشه قرار گیرند

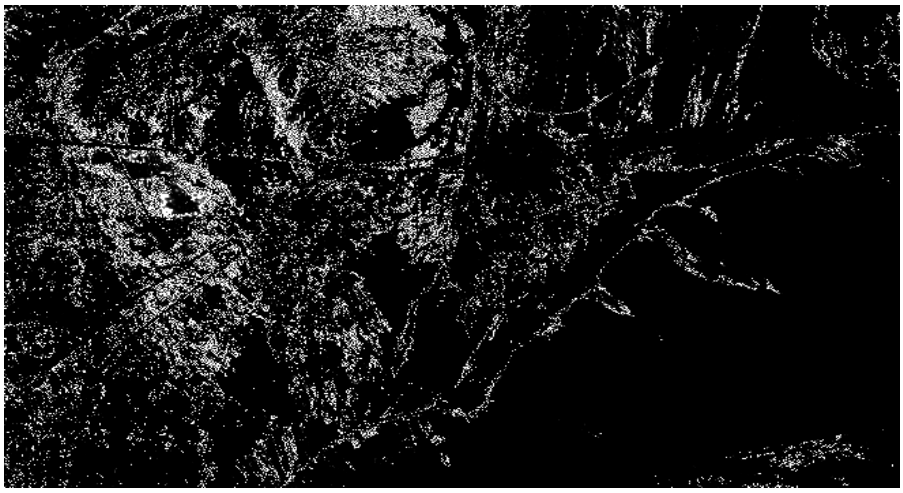
وزن معیارها نیز مطابق جدول ۲ برابر با ۰/۱ و ۰/۲۵ در نظر گرفته شده است.

جدول ۲. ماتریس تصمیم برای انتخاب حد آستانه‌ی مناسب در داده‌ی واقعی

شیب / وزن	۰/۱	۰/۲۵
	حداقل بیست درصد پوشش	حداکثر پنجاه درصد پوشش
۰ تا ۰/۲	۹	۵
۰/۲ تا ۰/۳	۸	۶
۰/۳ تا ۰/۴	۷	۷
۰/۴ تا ۰/۵	۶	۸
۰/۵ تا ۰/۶	۵	۸
۰/۶ تا ۰/۷	۲	۹
۰/۷ تا ۰/۸	۱	۹
۰/۸ تا ۰/۹	۱	۹

خروجی روش جایگشتی در متلب نشان داد که مناسب‌ترین گزینه، خطی با شیب بین ۰/۲ تا ۰/۳ است. ازاین‌رو مقدار آستانه‌ی ۰/۲۵ برای دسته‌بندی نهایی انتخاب شد. این مقدار، تعادلی منطقی بین دو معیار اصلی برقرار کرده و از نظر ساختار توزیع NDVI منطقه نیز مناسب‌ترین تفکیک را ایجاد می‌کند.

در نهایت، نقشه نهایی پوشش گیاهی به‌صورت دو خوشه مجزا (دارای پوشش / فاقد پوشش) تولید و در شکل ۱۲ نمایش داده شد. این نقشه خروجی نهایی مدل پیشنهادی در سطح داده‌های واقعی بوده و قابلیت استفاده در تحلیل‌های بعدی، مقایسه زمانی و شناسایی تغییرات محیطی را فراهم می‌سازد.



شکل ۱۲. نقشه‌ی نهایی خوشه‌بندی‌شده‌ی پوشش گیاهی در محدوده‌ی مورد مطالعه

٤ یافته‌ها

مساله دسته بندی عبارتست از تخصیص یک کلاس یا دسته به یک داده براساس یک تابع یا مدل از پیش تعیین شده. این مدل‌ها معمولاً با استفاده از داده‌های آموزشی و از طریق روش‌هایی مختلفی همچون درخت تصمیم، شبکه های عصبی مصنوعی، ماشین بردار پشتیبان، رگرسیون لجستیک ساخته می‌شوند.

با این حال، بخشی از چالش‌های دسته‌بندی زمانی بروز می‌کند که داده‌ها غیر قطعی، فازی یا ناقص باشند. در بسیاری از کاربردهای واقعی، داده‌های جمع‌آوری‌شده به‌دلایل مختلف از قطعیت برخوردار نیستند و از سوی دیگر، داده‌هایی که باید دسته‌بندی شوند اغلب به‌صورت تدریجی و در قالب داده‌های جریانی در طول زمان مشاهده می‌شوند.

در چنین شرایطی این پرسش مطرح است که چگونه می‌توان مدلی که هم توانایی دسته‌بندی داده‌های جریانی را داشته باشد و هم بتواند با داده‌های غیرقطعی و فازی سازگار شود و در طول زمان با تغییر وضعیت سیستم قابل به‌روزرسانی باشد. برای حل مسئله در این حالات استفاده از قابلیت‌های برنامه ریزی خطی درشرایط وجود داده‌های فازی و به‌کارگیری روش تحلیل پوششی داده‌ها (DEA) است. هرچند که روش DEA دراصل برای اندازه‌گیری کارایی نسبی واحدهای مشابه طراحی شده است، اما با تغییر تفسیر اجزای مدل از آن برای دسته‌بندی داده‌ها نیز استفاده کرد. در این رویکرد، هر داده به عنوان یک DMU در نظر گرفته می شود؛ ورودی‌ها ویژگی‌های داده و خروجی نشان‌دهنده کلاس آن است. در نتیجه، فضای نقاطی که به یک کلاس تعلق دارند با مرز کارایی DEA مشخص می‌شود و حصول تابع دسته‌بندی‌کننده از حل یک مدل برنامه‌ریزی خطی حاصل می‌شود. این ویژگی، امکان تحلیل حساسیت و به‌روزرسانی سریع مدل در صورت تغییر ضرایب را فراهم می‌سازد و همچنین امکان مواجهه با داده‌های فازی را فراهم می‌کند.

با وجود این مزایا، پژوهش‌های کمی به‌صورت مشخص به استفاده از DEA در مسئله دسته‌بندی پرداخته‌اند و بخشی از این پژوهش‌ها نیز با تمرکز بر داده‌های غیرقطعی انجام شده‌اند. از سوی دیگر، مدل‌های کلاسیک DEA همچون CCR فرض می‌کنند که ورودی‌ها و خروجی‌ها مقادیر دقیق دارند؛ در حالی که در بسیاری از محیط‌های واقعی، ورودی‌ها و خروجی‌ها ممکن است تنها به شکل داده‌های کراندار، ترتیبی یا بازه‌ای شناخته شوند. همین محدودیت سبب شده است که مدل‌های فازی، تصادفی و شبکه‌ای در DEA مطرح و توسعه یابند.

برای نمونه، مدل تحلیل پوششی داده‌های شبکه‌ای (NDEA) با تفکیک یک واحد تصمیم‌گیرنده به زیرفرآیندها امکان تحلیل دقیق‌تر تعاملات داخلی و شناسایی نقاط ناکارایی را فراهم می‌آورد. در کنار این توسعه‌ها، یکی از رویکردهای جدید که برای مواجهه با داده‌های ناقص یا مبهم مطرح شده تئوری خاکستری[1] است. این نظریه با تعریف داده‌ها به‌صورت بازه‌ای[2] (خاکستری) امکان تحلیل شرایطی را فراهم می‌کند که بخشی از اطلاعات شناخته‌شده و بخشی ناشناخته است. در مدل پیشنهادی این تحقیق نیز ورودی‌ها و خروجی‌ها به‌صورت بازه‌ای در نظر گرفته شده‌اند تا بتوان عدم قطعیت موجود در داده‌های سنجش از دور را مدل‌سازی کرد.

[1] Grey Theory
[2] Interval Grey Numbers

بدین ترتیب، سوال اصلی تحقیق این است که آیا می‌توان با ترکیب ویژگی‌های DEA، داده‌های خاکستری و روش‌های تصمیم‌گیری چندشاخصه روشی کارآمد برای دسته‌بندی داده‌های جریانی غیرقطعی ارایه کرد. در این پژوهش، در ادامه مطالعات پیشین در زمینه مدل‌سازی فازی و خاکستری، روشی ترکیبی ارایه شده است که امکان دسته‌بندی داده‌هایی را فراهم می‌کند که در طول زمان به‌صورت جریانی وارد می‌شوند و رفتار آن‌ها تغییر می‌کند.

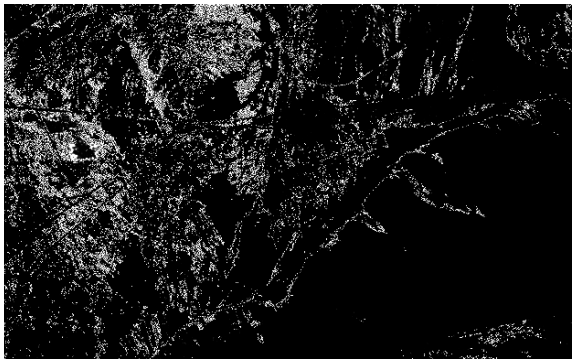
از سوی دیگر، مدل‌های تصمیم‌گیری چندشاخصه (MCDM) که یکی از شاخه‌های تصمیم‌گیری چندمعیاره‌اند، ابزار مناسبی برای انتخاب بهترین گزینه از میان چند گزینه پیشنهادی محسوب می‌شوند. این مدل‌ها، که به دو نوع جبرانی و غیرجبرانی تقسیم می‌شوند، با استفاده از ماتریس تصمیم، امکان تلفیق شاخص‌های مختلف و انتخاب بهترین گزینه را فراهم می‌کنند [۲۱]. علاوه بر این، تئوری خاکستری در سال‌های اخیر در کنار MCDM کاربرد چشمگیری در تحلیل داده‌های ناقص و کمّی داشته است [۲۲].

٤-۱ نوآوری انجام شده

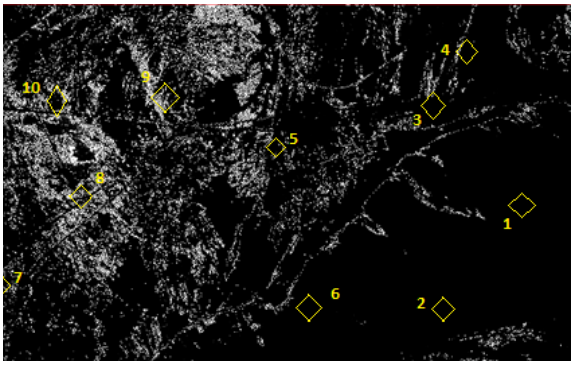
نوآوری انجام‌شده در این پژوهش، تلفیق روش‌های تصمیم‌گیری چندمعیاره و تحلیل پوششی داده‌ها برای دسته‌بندی داده‌های جریانی کسریِ خاکستری و همچنین به‌کارگیری داده‌های تصویری ماهواره‌ای به‌عنوان داده‌ی جریانی بزرگ برای دسته‌بندی ویژگی‌های تصویری است. بدین منظور، ابتدا الگوریتمی طراحی شد که طی آن داده‌های ماهواره لندست به داده‌های کسری تبدیل گردید و سپس با استفاده از روش جایگشتی، یک حد آستانه بهینه برای تمایز دو کلاس مختلف تعیین شد. این الگوریتم ابتدا بر داده‌های آزمایشی پیاده‌سازی و پس از تایید صحت نتایج، بر داده‌های اصلی اعمال شد.

بر اساس خروجی روش جایگشتی، مقدار آستانه ۰/۲۵ به‌عنوان بهترین حد برای جدایش دو کلاس انتخاب گردید. بر مبنای این آستانه، نقشه نهایی خوشه‌بندی پوشش گیاهی محدوده‌ی مطالعه تهیه شد (شکل ۱۳). از آنجا که اعتبار روش‌های سنجش‌ازدور برای تفکیک نواحی دارای پوشش گیاهی در ادبیات علوم جغرافیایی و علوم زمین به‌طور گسترده تأیید شده است، می‌توان بیان کرد که نقشه تولیدشده تطابق قابل قبولی با واقعیت زمینی دارد. در این بخش، هدف اصلی ارایه راهکاری عملی برای تفکیک کلاس‌های «دارای پوشش گیاهی» و «فاقد پوشش گیاهی» بوده است.

برای ارزیابی دقت مدل، تعداد ۱۰ نقطه از محدوده به‌صورت تصادفی انتخاب و وضعیت پوشش گیاهی آن‌ها با تصاویر گوگل‌ارث بررسی شد. از میان این ده نقطه، هشت نقطه به‌طور کامل با طبقه‌بندی ارایه‌شده توسط الگوریتم مطابقت داشتند. بدین ترتیب، دقت نهایی مدل برابر با ۸۰ درصد برآورد شد (شکل ۱۴ و جدول ۳).



شکل ۱۳. نقشه نهایی خوشه‌بندی شده پوشش گیاهی



شکل ۱٤. موقعیت نقاط اعتبارسنجی

جدول ۳. صحت‌سنجی عملکرد الگوریتم در دسته‌بندی پوشش گیاهی

شماره نقطه	دسته بندی ارایه شده توسط الگوریتم	وضعیت پوشش گیاهی روی زمین	صحیح یا غلط بودن دسته‌بندی
۱	بدون پوشش گیاهی	بدون پوشش گیاهی	صحیح
۲	بدون پوشش گیاهی	بدون پوشش گیاهی	صحیح
۳	بدون پوشش گیاهی	بدون پوشش گیاهی	صحیح
۴	بدون پوشش گیاهی	دارای پوشش گیاهی	غلط
۵	بدون پوشش گیاهی	دارای پوشش گیاهی	غلط
۶	بدون پوشش گیاهی	بدون پوشش گیاهی	صحیح
۷	دارای پوشش گیاهی	دارای پوشش گیاهی	صحیح
۸	دارای پوشش گیاهی	دارای پوشش گیاهی	صحیح
۹	دارای پوشش گیاهی	دارای پوشش گیاهی	صحیح
۱۰	بدون پوشش گیاهی	بدون پوشش گیاهی	صحیح

۵ نتیجه‌گیری

یافته‌های این پژوهش نشان داد که چارچوب پیشنهادی مبتنی بر تحلیل پوششی داده‌ها، همراه با داده‌های جریانی کسریِ خاکستری، توانایی استخراج مرزهای کارایی و جداسازی دو کلاس اصلی پوشش گیاهی و عدم پوشش را داراست. اجرای الگوریتم بر داده‌های واقعی لندست و اعتبارسنجی میدانی ده‌نقطه‌ای، دقتی در حدود ۸۰ درصد را نشان داد که بیانگر کارایی اولیه مدل در شرایط داده‌های غیرقطعی و جریانی است. این نتایج عددی، صحت عملکرد فرایند تعیین آستانه و طبقه‌بندی را تأیید می‌کند، هرچند همچنان نیازمند تکمیل و تقویت در مراحل بعدی پژوهش است [۲۳،۲۴].

با وجود این دستاوردها، محدودیت‌هایی نیز در پژوهش مشاهده شد؛ از جمله حجم محدود نقاط میدانی برای صحت‌سنجی، استفاده از دسته‌بندی دوسویه (دو کلاس) و اتکا به روش نظارت‌نشده برای تعیین آستانه طبقه‌بندی. این محدودیت‌ها نشان می‌دهند که مدل، اگرچه در داده‌های آزمایشی و عملی عملکرد قابل قبولی داشته، اما برای تعمیم‌پذیری در مقیاس‌های بزرگ‌تر نیازمند توسعه است.

در ادامه، با توجه به یافته‌های عددی پژوهش، چند مسیر پیشنهادی برای گسترش و تعمیق مطالعات آینده ارایه می‌شود.

نخست آنکه پیشنهاد می‌شود تعیین آستانه طبقه‌بندی، به‌جای اتکای صرف بر ماتریس تصمیم کارشناسی، با بهره‌گیری از روش‌های هوشمند و فراابتکاری همچون الگوریتم ژنتیک، بهینه‌سازی ازدحام ذرات و مدل‌های شبکه عصبی انجام گیرد تا امکان بهینه‌سازی دقیق‌تر آستانه فراهم شود. دوم آنکه استفاده از روش‌های دسته‌بندی نظارت‌شده مانند جنگل تصادفی، ماشین بردار پشتیبان یا شبکه‌های عصبی در کنار DEA می‌تواند امکان مقایسه، ترکیب و ارزیابی دقت مدل را افزایش دهد.

در گام بعد، گسترش تعداد نقاط میدانی و استفاده از فناوری‌های نوین گردآوری داده مثل پهپاد، سنجش زمینی و تصاویر با قدرت تفکیک بالاتر، می‌تواند خطای مدل را کاهش و قابلیت اطمینان آن را در مقیاس‌های گسترش‌یافته بهبود بخشد. همچنین توسعه چارچوب برای بیش از دو کلاس طبقه‌بندی، زمینه را برای شناسایی تنوع پوشش زمین، مطالعات زیست‌محیطی، تحلیل کیفیت خاک و پایش تغییرات کاربری فراهم می‌کند.

با توجه به ماهیت پویای داده‌های جریانی، بهره‌گیری از معماری‌های کلان‌داده و مدل‌های یادگیری عمیق برای پردازش آنلاین می‌تواند سرعت و مقیاس‌پذیری تحلیل را افزایش دهد. در نهایت، توسعه نسخه‌هایی از مدل که بتوانند خود را با تغییرات توزیع داده در طول زمان سازگار کنند، گامی اساسی برای ارتقای پایداری، قابلیت اعتماد و کاربردپذیری بلندمدت این رویکرد خواهد بود.

منابع

[1] Ghazanfari, M., Alizadeh, S., & Teymourpour, B. (2008). *Data Mining and Knowledge Discovery*. Iran University of Science and Technology. (In persian)

[2] Yin, S., & Kaynak, O. (2015). Big Data for Modern Industry: Challenges and Trends [Point of View]. *Proceedings of the IEEE*, 103(2), 143–146.

[3] Mena-Torres, D., & Aguilar-Ruiz, J. S. (2014). A similarity-based approach for data stream classification. *Expert Systems with Applications*, 41(9), 4224–4234.

[4] Shyaa, M. A., Zainol, Z., Abdullah, R., Anbar, M., Alzubaidi, L., & Santamaría, J. (2023). Enhanced intrusion detection with data stream classification and concept drift. *Sensors*, 23(7), 3736.

[5] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2020). A survey on concept drift adaptation. *ACM Computing Surveys*.

[6] Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.

[7] Yan, H., & Wei, Q. (2011). Data envelopment analysis classification machine. *Information Sciences*, 181(22), 5029–5041.

[8] Garner, C., & Holder, A. (2024). Classifying with uncertain data envelopment analysis. *EURO Journal on Computational Optimization*, 12, 100090.

[9] Jiang, B., Chi, E., & Li, J. (2022). Uncertain DEA for cross-efficiency with imprecise data. *Mathematics*, 10(13), 2161.

[10] Arabmaldar, A. & Hatami-Marbini, A. & Loske, D0 & Hammerschmidt, M. & Klumpp, M., 2024. "Robust data envelopment analysis with variable budgeted uncertainty," European Journal of Operational Research, Elsevier, vol. 315(2), pages 626-641. DOI: 10.1016/j.ejor.2023.11.043

[11] Zhu, J. (2020). DEA under big data: data-enabled analytics and network DEA. *Annals of Operations Research*.

[12] Khoshandam L. An Extension to Imprecise Data Envelopment Analysis. jor 2019; 16 (1) :117-129 (In persian)http://jamlu.liau.ac.ir/article-1-1616-fa.html

[13] Matyukira, A., Hatami-Marbini, A., Loske, D., Hammerschmidt, M., & Klumpp, M. (2024). Robust DEA with variable budgeted uncertainty. *European Journal of Operational Research*, 315(2), 626–641.
[14] Peykani P, Gheidar-Kheljani J. Modeling Scenario-Based Robust Data Envelopment Analysis Approaches: An Application to Research and Development Projects. jor 2022; 19 (2) :23-37 (In persian) dx.doi.org/10.52547/jamlu.19.2.23
[15] Shyaa, M. A., Zainol, Z., Abdullah, R., Anbar, M., Alzubaidi, L., & Santamaría, J. (2023). Enhanced intrusion detection with data stream classification and concept drift. *Sensors*, 23(7), 3736.
[16] Matyukira, C., & Mhangara, P. (2024). Advances in vegetation mapping through remote sensing and ML. *European Journal of Remote Sensing*.
[17] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666.
[18] Abonyi, J., & Feil, B. (2007). *Cluster analysis for data mining and system identification*. Springer.
[19] Lee, S., Kim, G., & Kim, S. (2011). Self-adaptive and dynamic clustering for online anomaly detection. *Expert Systems with Applications*, 38(12), 14891–14898.
[20] Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078–1092.
[21] Colson, G., Bruyn, C., & Rodin, E. Y. (1989). *Models and Methods in Multiple Criteria Decision Making*.
[22] Lewis, R. W., & Zheng, Y. (1993). On the optimization concept of grey systems. *Applied Mathematical Modelling*.
[23] Wu, H., Yan, W., & Liu, S. (2007). Analysis of the efficiency of regional electricity input-output for China based on grey DEA model. *International Conference on Systems, Man and Cybernetics*, 2200–2204.
[24] Lewis, H. F., & Sexton, T. R. (2004). Network DEA: efficiency analysis of organizations with complex internal structure. *Computers & Operations Research*, 31, 1365–1410.