



Research Article

Quality Control of Powdered Metformin Using Hyperspectral Imaging and Machine Learning Models

Fatemeh Hatefi, Zahra Bolhassani, Hadi Parastar* 
Department of Chemistry, Sharif University of Technology, Tehran, Iran

PAPER INFO

Article history:

Received: 25/May/2025

Revised: 30/Sep/2025

Accepted: 23/Nov/2025

Keywords:

Hyperspectral Imaging,
 Metformin,
 Quality Control,
 Chemometrics,
 PLS-DA Analysis,
 Artificial Neural Network
 (ANN)

ABSTRACT

Given the global growth in the production and consumption of generic drugs and the increasing risk of counterfeit or substandard pharmaceutical products, the development of novel, rapid, and non-destructive quality control methods has become critically important. In this study, hyperspectral imaging (HSI) in the visible to near-infrared range (Vis-NIR, 400–950 nm), combined with chemometric/machine learning techniques, was employed to assess the active pharmaceutical ingredient (API) content of metformin in powder-based samples. Samples were classified into three dosage groups: standard dose (SD), low non-standard dose (LD), and high non-standard dose (HD). Hyperspectral imaging data were processed using spectral averaging, principal component analysis (PCA), partial least squares-discriminant analysis (PLS-DA), and machine learning algorithms including artificial neural network (ANN) and support vector machines (SVM). Results demonstrated that chemometric models, particularly ANN and PLS-DA, could effectively differentiate between the three sample groups with high accuracy. The combination of Vis-NIR HSI and statistical modelling proved to be a powerful tool for detecting metformin dosage levels and distinguishing standard from non-standard pharmaceutical compositions.

DOI: <https://doi.org/10.22075/chem.2025.37869.2370>

© 2025 Semnan University.

This is an open access article under the CC-BY-SA 4.0 license. (<https://creativecommons.org/licenses/by-sa/4.0/>)

*.Corresponding author: Associate Professor of Analytical Chemistry. E-mail address: h.parastar@sharif.edu

How to cite this article: Hatefi, F., Bolhassani, Z., & Parastar, H. (2025). Quality Control of Powdered Metformin Using Hyperspectral Imaging and Machine Learning Models. *Applied Chemistry Today*, 20(76), 97-116. (in Persian)

کنترل کیفیت متفورمین در نمونه‌های پودری با استفاده از تصویربرداری ابرطیفی و

مدل‌های یادگیری ماشین

فاطمه هاتفی، زهرا بوالحسنی، هادی پرستار شهری*

دانشکده شیمی، دانشگاه صنعتی شریف، تهران، ایران

اطلاعات مقاله	چکیده
دریافت مقاله: ۱۴۰۴/۰۳/۰۴	با توجه به رشد فزاینده تولید و مصرف داروهای ژنریک در سطح جهانی و افزایش خطر ورود محصولات دارویی تقلبی و غیراستاندارد، ابداع روش‌های نوین، سریع و غیرمخرب برای کنترل کیفیت داروها از اهمیت ویژه‌ای برخوردار است. در این پژوهش، از فناوری تصویربرداری ابرطیفی در محدوده مرئی تا زیر قرمز نزدیک ^۱ (Vis-NIR HSI) به همراه روش‌های کمومتریکس ^۲ /یادگیری ماشین ^۳ برای ارزیابی دوز ماده مؤثره متفورمین در نمونه‌های پودری استفاده شد. نمونه‌ها در سه دسته با دوز استاندارد (SD)، غیر استاندارد پایین (LD) و غیر استاندارد بالا (HD) تهیه شدند. پس از جمع‌آوری داده‌های تصویری ابرطیفی، از میانگین‌گیری طیفی، تحلیل مؤلفه‌های اصلی ^۴ (PCA)، روش حداقل مربعات جزئی-تحلیل تمایزی ^۵ (PLS-DA) و دو الگوریتم یادگیری ماشین شامل شبکه عصبی مصنوعی ^۶ (ANN) و ماشین بردار پشتیبان ^۷ (SVM) برای تحلیل و طبقه‌بندی داده‌ها بهره گرفته شد. نتایج نشان داد که مدل‌های کمومتریکسی، به‌ویژه شبکه عصبی مصنوعی و حداقل مربعات جزئی-تحلیل تمایزی، توانسته‌اند سه گروه مورد نظر را با دقت و صحت بالا از یکدیگر تفکیک نمایند. تصویربرداری ابرطیفی در بازه طول موجی ۴۰۰ تا ۹۵۰ نانومتر، در کنار پردازش‌های آماری، ابزاری مؤثر و کارآمد در شناسایی کیفیت متفورمین و تمایز نمونه‌های استاندارد و غیراستاندارد فراهم کرده است.

کلمات کلیدی:

تصویربرداری ابرطیفی، متفورمین، کنترل کیفیت، کمومتریکس، یادگیری ماشین، شبکه عصبی مصنوعی.

DOI: <https://doi.org/10.22075/chem.2025.37869.2370>This is an open access article under the CC-BY-SA 4.0 license. (<https://creativecommons.org/licenses/by-sa/4.0/>)

۱- مقدمه

صنعت داروسازی علاوه بر نقش حیاتی در سلامت و درمان، تأثیر قابل توجهی در اقتصاد و تجارت جهانی دارد و دسترسی به داروهای باکیفیت و موثر اهمیت بالایی دارد. داروها به چند دسته تقسیم می‌شوند: داروهای اصیل که اثربخشی آنها در مطالعات بالینی اثبات شده و نمونه قبلی ندارند؛ داروهای ژنریک که مشابه داروی اصلی هستند اما پس از پایان انحصار تولید، توسط شرکت‌های دیگر ساخته می‌شوند؛ داروهای غیراستاندارد که کیفیت لازم را ندارند و ممکن است مواد مؤثره ناکافی یا بسته‌بندی

¹ Vis-NIR hyperspectral imaging² Chemometrics³ Machine learning⁴ Principal component analysis⁵ Partial least squares-discriminant analysis⁶ Artificial neural network⁷ Support vector machine

ضعیف داشته باشند؛ داروهای ثبت نشده که مورد تایید نهادهای نظارتی نیستند؛ و داروهای تقلبی که هویت یا ترکیب آنها عمداً اشتباه معرفی شده است [۲ و ۱].

همه‌گیری کووید-۱۹ با افزایش تقاضا برای محصولات پزشکی، اختلال در زنجیره تأمین و ضعف نظارت، زمینه را برای ورود داروهای بی کیفیت و تقلبی فراهم کرد. بیشترین داروهای تقلبی مربوط به داروهای دستگاه تناسلی، سیستم عصبی مرکزی، داروهای ضد عفونت، سیستم عضلانی-اسکلتی و داروهای کنترل تقسیم سلولی است. این پدیده علاوه بر آسیب به سلامت، موجب بی اعتمادی به نظام سلامت می شود [۴ و ۳].

برای مقابله با این مشکلات، آژانس‌های نظارتی استانداردهایی مانند اصول تولید خوب^۱ و فارماکوپه‌ها را تدوین کرده‌اند تا کنترل کیفیت از مواد اولیه تا توزیع دارو انجام شود. با توجه به رشد بازار دارویی و افزایش ریسک توزیع داروهای بی کیفیت، توسعه روش‌های تحلیلی سریع، دقیق و غیر تخریبی برای کنترل کیفیت ضروری است. اگرچه روش‌های پیچیده کروماتوگرافی همچنان پایه کنترل کیفیت هستند، اما استفاده از روش‌های ساده‌تر و جدیدتر برای کاهش هزینه و زمان و بهبود عملکرد مورد توجه قرار گرفته است. رعایت الزامات نظارتی برای پذیرش این روش‌ها ضروری است و توسعه رویکردهای نوین می‌تواند به ارتقای سلامت عمومی و تحقق اهداف شیمی سبز کمک کند [۵].

سنجش کیفیت و اصالت داروها از مراحل اولیه، یعنی آنالیز مواد خام، آغاز شده و تا مراحل پایانی مانند توزیع در داروخانه‌ها و بیمارستان‌ها ادامه می‌یابد. این کنترل کیفیت از سه جنبه بصری، فیزیکی و شیمیایی انجام می‌شود. در بررسی بصری، یکپارچگی بسته‌بندی و مطابقت دوز بررسی می‌شود. ارزیابی فیزیکی شامل بررسی هولوگرام و ضریب شکست برای اطمینان از اصالت است. در کنترل شیمیایی، ترکیبات و خواص شیمیایی دارو با استفاده از تکنیک‌های تجزیه‌ای مانند الکتروفورز موئینه [۶]، کروماتوگرافی مایع-طیف سنج جرمی [۷]، کروماتوگرافی سیال فوق بحرانی و کروماتوگرافی دو بعدی [۸]، کروماتوگرافی مایع با عملکرد بالا [۹] و طیف‌سنجی رامان [۱۰] مورد بررسی قرار می‌گیرند. این روش‌ها اگرچه دقت و قدرت بالایی دارند، اما هزینه‌بر، زمان‌بر و نیازمند مواد شیمیایی خاص و آماده‌سازی نمونه [۱۱] هستند.

کاربرد گسترده کروماتوگرافی مایع با عملکرد بالا در اندازه‌گیری و آنالیز داروها در مطالعات مختلف گزارش شده است؛ برای نمونه، در یک پژوهش از کروماتوگرافی مایع با عملکرد بالا پس از میکرواستخراج فاز جامد پخشی مبتنی بر اکسیدگرافن برای اندازه‌گیری دیازپام در نمونه‌های دارویی و بیولوژیکی استفاده شده است [۱۲]، و در مطالعه دیگری کروماتوگرافی مایع با کارایی بالا با آشکارساز مرئی-فرابنفش به عنوان روشی سریع، دقیق و کاملاً اعتبارسنجی شده برای تعیین بنیدپین در نمونه‌های دارویی، سرم و ادرار انسانی به کار گرفته شده است [۱۳]. این نمونه‌ها نقش کروماتوگرافی مایع با عملکرد بالا را به عنوان یکی از روش‌های

¹ Good manufacturing practice (GMP)

مرجع در تحلیل داروها تأیید می‌کنند؛ با این حال، چالش‌هایی مانند زمان آنالیز طولانی و آماده‌سازی نسبتاً پیچیده نمونه، نیاز به توسعه روش‌های سریع‌تر و غیرتخریبی را برجسته‌تر کرده است.

به همین دلیل، اخیراً استفاده از تکنیک‌های طیف‌سنجی ارتعاشی و تصویربرداری ابرطیفی در ناحیه مرئی-زیرقرمز نزدیک رواج یافته است. این روش‌ها بدون آسیب به نمونه و با حداقل آماده‌سازی، امکان تجزیه و تحلیل داروها را در اشکال مختلف (مایع، جامد، پودر) فراهم می‌کنند و حتی می‌توانند در مراحل مختلف تولید و توزیع مورد استفاده قرار گیرند. هرچند نتایج این دستگاه‌های قابل حمل باید با روش‌های مرجع مقایسه شوند، اما ترکیب آن‌ها با روش‌های پردازش داده و مدل‌سازی آماری، سنجش کیفیت داروها را سریع‌تر و کارآمدتر کرده است [۱۴].

متفورمین دارویی از خانواده بیگوانیدها و برگرفته از گیاه گالگا است که به‌عنوان خط اول درمان دیابت نوع ۲ تجویز می‌شود. این دارو معمولاً به‌صورت قرص‌های ۵۰۰ و ۱۰۰۰ میلی‌گرمی عرضه شده و دوز مصرفی آن در بزرگسالان تا ۳ گرم در روز می‌رسد. به دلیل نیمه‌عمر کوتاه، در فرآورده‌های پیوسته‌رهش از موادی مانند هیدروکسی‌پروپیل متیل سلولوز برای کنترل آزادسازی استفاده می‌شود [۱۵].

با افزایش شیوع دیابت، بازار متفورمین رشد چشمگیری داشته و همین امر خطر تولید محصولات تقلبی را افزایش داده است. بنابراین، کنترل کیفیت آن اهمیت بالایی دارد. روش‌هایی مانند کروماتوگرافی مایع با کارایی بالا^۱، طیف بینی رزونانس مغناطیسی هسته^۲، مرئی تا نزدیک فروسرخ^۳، و اخیراً طیف‌سنجی زیرقرمز نزدیک^۴ به عنوان روش‌های سریع، غیرمخرب و کم‌هزینه برای تحلیل متفورمین به کار گرفته می‌شوند و نقش مهمی در ارتقای کیفیت دارو ایفا می‌کنند [۱۶ و ۱۷].

طیف‌سنجی در ناحیه زیرقرمز نزدیک از اوایل قرن بیستم به تدریج توسعه یافت و با کشف پتانسیل‌های کمی آن توسط کارل نوریس، کاربردهای این روش گسترش یافت. در طیف‌سنجی کلاسیک، اطلاعات طیفی یکپارچه‌ای از نمونه به دست می‌آید که متناسب با ناحیه‌ای است که پرتو نور با آن تماس داشته است. نخستین نقشه شیمیایی با میکروسکوپ در ناحیه زیرقرمز میانه در سال ۱۹۹۸ توسط هارتکاک و اتکین ارائه شد که اطلاعات طیفی و فضایی از سطح نمونه را به‌صورت همزمان فراهم می‌کرد. پیشرفت‌های اپتیکی، مانند توسعه آشکارسازهای آرایه صفحه کانونی و نرم‌افزارهای پردازش داده، منجر به گسترش تصویربرداری شیمیایی یا تصویربرداری ابرطیفی شد که در آن علاوه بر تصویر، اطلاعات طیفی گسترده‌ای نیز ثبت می‌شود [۱۸ و ۱۹].

تصویربرداری ابرطیفی نسبت به تصاویر دوربین سه کاناله^۵ اطلاعات بسیار کامل‌تری از نمونه‌ها ارائه می‌دهد و به‌عنوان ابزاری هوشمند در حوزه‌های مختلفی مانند کشاورزی، مطالعات بافت بیولوژیک و کنترل کیفیت مواد دارویی و غذایی کاربرد دارد.

¹ High-performance liquid chromatography (HPLC)

² Nuclear magnetic resonance (NMR)

³ Visible near infrared (Vis-NIR)

⁴ Near infrared (NIR)

⁵ RGB

سنجش نور در این دستگاه‌ها به سه صورت عبوری، بازتاب شده و از طریق فیبرهای نوری انجام می‌شود که سنجش نور عبوری برای مواد شفاف مناسب است و برای نمونه‌های ضخیم مانند قرص‌ها کاربرد کمتری دارد. از آنجا که بسیاری از ترکیبات آلی، از جمله مواد دارویی، در ناحیه نور مرئی شفاف هستند و اطلاعات طیفی محدودی در این ناحیه دارند، ترکیب ناحیه مرئی با زیرقرمز نزدیک امکان دریافت اطلاعات کیفی و کمی بهتری را فراهم می‌کند [۲۰]. در ناحیه زیرقرمز نزدیک، باندهای ارتعاشی ناشی از پیوندهای شیمیایی مانند کشش باندهای S-N، S-O، X-C و X-H و همچنین ارتعاشات پیوندهای قوی‌تر مانند C=O مشاهده می‌شوند. این باندها معمولاً شدت ضعیف و همپوشانی زیادی دارند که تفسیر مستقیم طیف را دشوار می‌کند، بنابراین استفاده از روش‌های کمومتریکس برای تحلیل داده‌های حاصل از طیف‌سنجی در این ناحیه بسیار مفید است و به درک بهتر ساختار مولکولی کمک می‌کند [۲۱].

کاربرد ترکیبی طیف‌سنجی و روش‌های کمومتریکس در بسیاری از مطالعات دیگر نیز موفقیت‌آمیز بوده است؛ از جمله اندازه‌گیری هم‌زمان یون‌های Al(III)، Cu(II) و Fe(II) با استفاده از شناساگر اریوکروم‌سیانین R و مدل‌های حداقل مربعات جزئی برای غلبه بر هم‌پوشانی طیفی [۲۲]. تعیین هم‌زمان اوراسیل و ۵-فلوئورو اوراسیل با کالیبراسیون چندمتغیره حداقل مربعات جزئی در ناحیه ۲۲۰-۳۱۴ نانومتر [۲۳]، و اندازه‌گیری هم‌زمان نیوفوشین و ویکتوریا بلو با بهره‌گیری از حداقل مربعات جزئی و روش پیش‌پردازش تصحیح سیگنال عمودی به منظور بهبود قدرت پیش‌بینی مدل [۲۴]. این نمونه‌ها نشان می‌دهند که ترکیب روش‌های طیف‌سنجی و کمومتریکس ابزاری قدرتمند برای تحلیل مخلوط‌های پیچیده و سیستم‌هایی با هم‌پوشانی شدید طیفی است.

هدف اصلی این پژوهش بررسی عملکرد دستگاه تصویربرداری ابرطیفی در ترکیب با روش‌های کمومتریکس/یادگیری ماشین برای بررسی ماده مؤثره متفورمین در نمونه‌های پودر در مخلوط با مواد جانبی برای تشخیص ترکیب استاندارد این دارو از غیر استاندارد است. از آنجایی که اکثر مطالعات انجام شده با این تکنیک‌ها در بازه طول موجی ۱۰۰۰ تا ۲۵۰۰ نانومتر هستند سعی بر این است تا در این پژوهش کاربرد اطلاعات سودمند بازه طول موجی ۴۰۰ تا ۱۰۰۰ نانومتر را نیز بررسی کنیم.

۲- بخش تجربی

۲-۱- مواد و نمونه‌ها

مواد مورد استفاده در ساخت نمونه‌های متفورمین شامل ماده مؤثره متفورمین هیدرو کلراید^۱، مواد جانبی شامل هیدروکسی پروپیل متیل سلولز K100M^۲، پلی وینیل پیرولیدون K30^۳، آویسل^۴ و منیزیم استئارات^۵ می‌باشد. برای مجموعه کالیبراسیون

^۱ Met

^۲ HPMC K100M

^۳ PVP K30

^۴ Avicel

^۵ Magnesium stearate

۳۶ نمونه تهیه شد؛ هر نمونه ۱۰۰۰ mg وزن داشته و مقدار متفورمین در آن‌ها در ۳۰۰ تا ۶۵۰ میلی گرم توزیع شد. دسته‌بندی دوزها به سه گروه دوز غیر استاندارد پایین (LD) (۳۰۰-۴۴۰ میلی گرم)، دوز استاندارد (SD) (۴۵۰-۵۵۰ میلی گرم) و دوز غیر استاندارد بالا (HD) (۵۶۰-۶۵۰ میلی گرم) تقسیم شد. برای افزایش تنوع نمونه در هر گروه، از گام ۱۰ mg بین نمونه‌ها استفاده شد و مقادیر سایر اجزا (پلی وینیل پیرولیدون K30، منیزیم استئارات، آویسل) ثابت نگه داشته شد؛ مابقی وزن با هیدروکسی پروپیل متیل سلولز K100M تنظیم شد. جدول ۱ توزیع دقیق مقادیر هر جزء را نشان می‌دهد. لازم به ذکر است که در این پژوهش از قرص‌های تجاری استفاده نشد و تمامی نمونه‌ها به صورت مصنوعی و در شرایط آزمایشگاهی تهیه شدند. این موضوع اطمینان می‌دهد که اثرات مشاهده شده صرفاً ناشی از تغییر مقدار متفورمین و ترکیب اجزای فرمولاسیون بوده است. دوز استاندارد متفورمین بین ۹۰ تا ۱۱۰ درصد میلی گرم متفورمین در قرص مورد نظر است [۲۵].

جدول ۱. مشخصات دسته بندی نمونه ها و جرم مواد اولیه در هر گروه

ماده	جرم ماده های مختلف در سه طبقه بندی پودر ها (میلی گرم)		
	HD	LD	SD
متفورمین هیدرو کلراید	۵۶۰ تا ۶۵۰	۳۰۰ تا ۴۴۰	۴۵۰ تا ۵۵۰
هیدروکسی پروپیل متیل سلولز K100M	۱۸۰ تا ۲۷۰	۳۹۰ تا ۵۳۰	۲۸۰ تا ۳۸۰
پلی وینیل پیرولیدون K30	۱۳۰	۱۳۰	۱۳۰
آویسل	۳۰	۳۰	۳۰
منیزیم استئارات	۱۰	۱۰	۱۰

۲-۲- تصویربرداری ابرطیفی و نحوه تصویربرداری

تصویربرداری ابرطیفی با استفاده از دستگاه تصویربرداری ابرطیفی در محدوده مرئی تا نزدیک فروسرخ^۱ انجام شد. این دستگاه با نام HYSPIV Vis-NIR و مدل HSI-Vis-NIR، ۱۲ بیت ساخت شرکت ایرانی پرتو افزار صنعت (زنجان، ایران) است و در محدوده طول موجی ۴۰۰ تا ۹۵۰ نانومتر کار می‌کند. همان‌طور که در شکل ۱ مشاهده می‌شود، اجزای اصلی دستگاه شامل موارد زیر است: (۱) دوربین سه کاناله برای نمایش نمونه، که در بخش بالایی دستگاه قرار دارد، (۲) دوربین ابرطیفی، (۳) ستون نگهدارنده که دوربین‌ها روی آن نصب شده‌اند، (۴) منبع نور هالوژن، (۵) سطح متحرک (اسکتر خطی) برای قرار دادن نمونه که با حرکت نمونه در میدان دید دوربین، فرایند اسکن را انجام می‌دهد، (۶) درایور اسکتر که اطلاعات را به رایانه منتقل می‌کند، و (۷) یک رایانه مجهز به نرم‌افزار مربوطه جهت جمع‌آوری داده‌ها.

۲-۳- کمومتریکس و آنالیز داده های تصویربرداری ابرطیفی

تجزیه و تحلیل تصاویر ابرطیفی با وجود حجم بالای اطلاعات تشخیصی (به دلیل وضوح طیفی و فضایی بالا)، چالش‌هایی مانند همپوشانی باندها، تنوع امضاهای طیفی و سه‌بعدی بودن داده‌ها دارد. برای غلبه بر این پیچیدگی، از روش‌های کمومتریکس در سه دسته اصلی استفاده می‌شود:

^۱ Vis-NIR HSI



شکل ۱. اجزاء تشکیل دهنده دستگاه تصویربرداری ابرطیفی.

الف) روش‌های تشخیص الگو (طبقه‌بندی نظارت‌شده یا بدون نظارت): در این روش‌ها، داده‌ها بر اساس اطلاعات پیشینی (نظارت‌شده) یا بدون آن (بدون نظارت) دسته‌بندی می‌شوند.

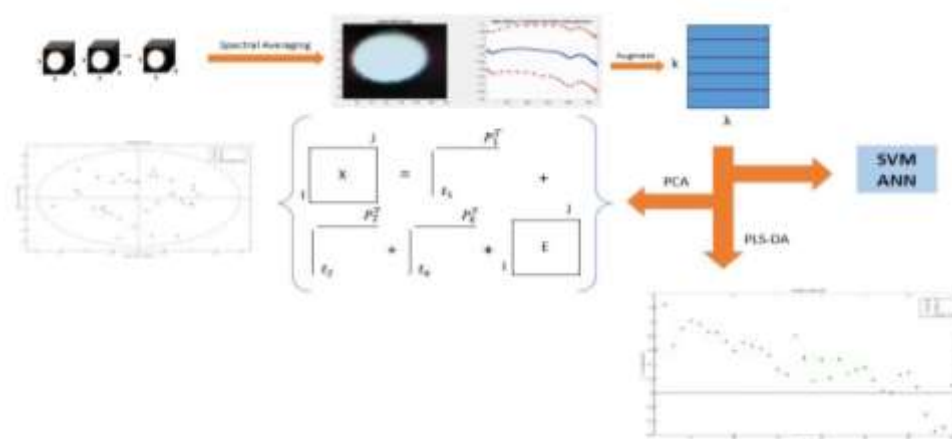
ب) کالبراسیون چندمتغیره: ایجاد ارتباط بین اندازه‌گیری‌ها و پارامترهای هدف (مانند غلظت ماده) با مدل‌های خطی یا غیرخطی برای پیش‌بینی مقادیر ناشناخته.

ج) تفکیک منحنی چندمتغیره (مانند MCR-ALS): استخراج سهم خالص گونه‌های شیمیایی در داده‌های مخلوط. در این پژوهش، تمرکز بر روش تشخیص الگو برای تحلیل تصاویر ابرطیفی است. این روش به استخراج اطلاعات معنادار از داده‌های پیچیده و بهبود دقت تشخیص در حوزه‌هایی مانند پزشکی و داروسازی کمک می‌کند.

۲-۴- آنالیز داده‌های تصویرهای ابرطیفی

در این مطالعه، داده‌ها از تصویربرداری ابرطیفی ۳۶ نمونه دارویی با ابعاد اولیه $۵۴۰ \times ۳۶۵ \times ۷۰$ به دست آمد. در این ساختار سه‌بعدی، دو بعد اول (۳۶۵ و ۷۰) ابعاد تصویر هر نمونه هستند و بعد سوم (۵۴۰) مربوط به مقادیر جذب در طول موج‌های مختلف برای هر پیکسل است. برای تمرکز بر ناحیه حاوی اطلاعات مؤثر، یک منطقه مورد علاقه^۱ با ابعاد $۵۴۰ \times ۱۴۱ \times ۱۴$ برای هر نمونه انتخاب شد. سپس با میانگین‌گیری طیفی روی تمام پیکسل‌های این ناحیه، برای هر نمونه یک بردار طیفی با ابعاد ۵۴۰×۱ استخراج گردید که به عنوان ورودی مدل‌های تحلیل کمومتریکس به کار گرفته شد. نمای شماتیک این روند در شکل ۴ آورده شده است.

^۱ Region of interest (ROI)



شکل ۴. نمودار مراحل مختلف آنالیز تصاویر ابرطیفی

برای تحلیل ساختار داده‌ها و کاهش ابعاد، از تحلیل مؤلفه‌های اصلی^۱ استفاده شد. تحلیل مؤلفه‌های اصلی روشی بدون نظارت و مبتنی بر تجزیه ماتریس کوواریانس داده‌هاست که هدف آن یافتن ترکیب‌های خطی از متغیرها (مؤلفه‌های اصلی) است که بیشترین واریانس داده‌ها را حفظ کرده و در عین حال نسبت به یکدیگر متعامد^۲ باشند. خروجی این روش شامل ماتریس امتیازها^۳ به‌عنوان نمایش داده‌ها در فضای جدید، ماتریس لودینگ‌ها^۴ به‌عنوان ضرایب ترکیب خطی متغیرها و ماتریس باقی‌مانده‌ها است که تغییرات غیرقابل توضیح توسط مدل را دربر می‌گیرد [۲۶ و ۲۷].

در گام بعدی، به‌منظور انجام طبقه‌بندی نظارت‌شده، از تحلیل تمایزی حداقل مربعات جزئی^۵ استفاده شد. این روش با هدف یافتن مؤلفه‌هایی طراحی شده که هم‌زمان بیشترین واریانس را در متغیرهای مستقل (X) حفظ کرده و بیشترین همبستگی را با متغیرهای وابسته (Y) داشته باشند. در این مدل، هر دو ماتریس X و Y به فضای امتیازها و لودینگ‌ها تبدیل می‌شوند و مدل به‌صورت زوج معادلات ماتریسی $Y = UQ^t + F$ و $X = TP^t + E$ تعریف می‌گردد. این ساختار امکان استخراج متغیرهای پنهان^۶ را فراهم می‌کند که در آن دسته‌بندی کلاس‌ها با بیشینه‌سازی کوواریانس میان X و Y انجام می‌گیرد [۲۸].

طبق جدول ۲ مقادیر حساسیت، ویژه بودن و سایر پارامترهای سنجش مدل را براساس مقادیر مثبت واقعی (TP)، منفی واقعی (TN)، مثبت کاذب (FP) و منفی کاذب (FN) که در معادلات نشان داده شده است، بیان کرد [۲۹].

جدول ۲. نگرشی دیگر به مفاهیم حساسیت و ویژه بودن [۲۹].

مقدار پیش‌بینی شده	مقدار واقعی		
		Positive	Negative
	Positive	TP	FP
	Negative	FN	TN

¹ PCA

² Orthogonal

³ Scores

⁴ Loadings

⁵ PLS-DA

⁶ Latent Variables

$$\begin{aligned} \text{حساسیت} &= \frac{TP}{TP + FN} \\ \text{ویژه‌گزینی} &= \frac{TN}{TN + FP} \\ \text{صحت} &= \frac{TP + TN}{TP + FN + TN + FP} \end{aligned}$$

علاوه بر روش‌های آماری، از دو الگوریتم یادگیری ماشین نیز برای بررسی ساختار دسته‌بندی استفاده شد. شبکه عصبی مصنوعی^۱ به عنوان یک مدل داده‌محور با قابلیت یادگیری روابط غیرخطی، شامل لایه‌هایی از نورون‌های مصنوعی است که با وزن‌دهی ورودی‌ها، آن‌ها را از طریق توابع فعال‌سازی به خروجی تبدیل می‌کنند. در این مدل، با تنظیم تعداد لایه‌ها، تعداد نورون‌ها، نوع تابع فعال‌سازی و میزان بایاس، می‌توان مدل را برای شناسایی الگوهای پیچیده در داده‌های طیفی آموزش داد [۳۰]. همچنین، از ماشین بردار پشتیبان^۲ با تابع کرنل خطی برای طبقه‌بندی استفاده شد. این روش بر پایه یافتن ابرصفحه‌ای در فضای ویژگی‌ها عمل می‌کند که بیشترین فاصله ممکن^۳ را میان کلاس‌های مختلف حفظ کند. در ساختار خطی، ماشین بردار پشتیبان نیازی به تبدیل داده‌ها به فضای ویژگی غیرخطی ندارد و برای مسائل با مرزهای تصمیم ساده، عملکرد مؤثری دارد. مدل ماشین بردار پشتیبان با استفاده از تنظیم پارامترهای جریمه (C) و نوع تابع کرنل قابل سفارشی‌سازی است. در این مطالعه، نسخه خطی برای ایجاد مدلی ساده و قابل تفسیر انتخاب شد [۳۱ و ۳۲].

تمامی روش‌ها با استفاده از داده‌های طیفی میانگین‌گیری‌شده (با ابعاد ۵۴۰×۱) و با پیش‌پردازش‌های اولیه شامل هموارسازی^۴ با الگوریتم Savitzky-Golay پنجره ۱۵ نقطه و درجه صفر و مرکزسازی نسبت به میانگین^۵ اجرا شدند. اعتبارسنجی مدل‌ها با استفاده از روش اعتبارسنجی متقاطع^۶ انجام گرفت. تحلیل نتایج عددی حاصل از این روش‌ها در بخش‌های بعدی مورد بررسی قرار خواهد گرفت.

۲-۵- نرم افزارهای مورد استفاده

داده‌های تصویربرداری ابرطیفی با استفاده از نرم‌افزار MATLAB نسخه R2021b (محصول شرکت MathWorks، ایالات متحده) پردازش شدند و برای مدل‌سازی با استفاده از شبکه عصبی مصنوعی و ماشین بردار پشتیبان هم از این نسخه MATLAB به همراه برنامه یادگیرنده طبقه‌بندی^۷ موجود در MATLAB استفاده شد. برای محاسبه میانگین طیفی از ناحیه نمونه در تصاویر، جعبه‌ابزار HYPER-Tools نسخه ۳ به کار گرفته شد. برای مدل‌سازی با استفاده از روش حداقل مربعات جزئی-آنالیز تمایزی و آنالیز مؤلفه اصلی، از نرم‌افزار MATLAB نسخه R2013a و جعبه‌ابزار PLS نسخه ۷٫۸ استفاده شد.

^۱ ANN

^۲ SVM

^۳ Margin

^۴ Smoothing

^۵ Mean-Centering

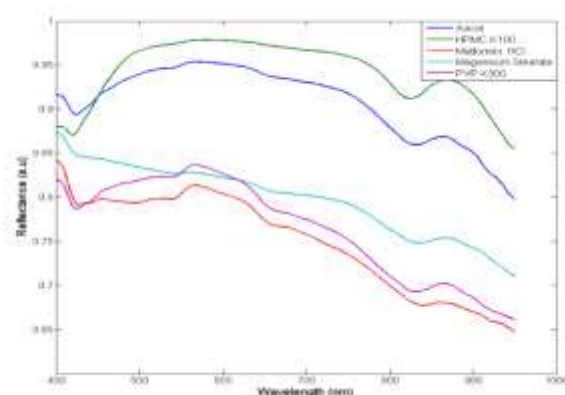
^۶ Cross-Validation

^۷ Classification Learner

۳- بحث و نتیجه گیری

۳-۱- اطلاعات طیفی مواد اولیه

مواد اولیه مورد استفاده شامل متفورمین هیدروکلراید، هیدروکسی پروپیل متیل سلولز K100M، پلی وینیل پیرولیدون K30، آویسل و منیزیم استئارات بودند. تمامی مواد اولیه با خلوص ۹۸٪ مورد استفاده قرار گرفتند. این ترکیبات همگی سفیدرنگ هستند که این ویژگی باعث محدودیت تحلیل کیفی آن‌ها در ناحیه مرئی طیف می‌شود. با این حال، ترکیب ناحیه مرئی و مادون قرمز نزدیک امکان استخراج اطلاعات کیفی مفیدی را فراهم می‌کند. در این ناحیه، جذب‌های ضعیف اما مشخصی ناشی از اورتون‌ها و باندهای ترکیبی ارتعاشات کششی پیوندهای X-H مانند C-H و N-H دیده می‌شود. در محدوده ۷۰۰ تا ۹۰۰ نانومتر (سومین اورتون)، گروه‌های C-H موجود در اغلب مواد اولیه جذب ضعیفی ایجاد می‌کنند. در حدود ۸۰۰ نانومتر نیز جذب مشخصی ناشی از گروه آمین اولیه موجود در متفورمین دیده می‌شود. همچنین در ناحیه ۴۰۰ تا ۵۰۰ نانومتر، حضور پیوندهای دوگانه‌ای مانند C=O و N=O منجر به جذب قوی‌تری می‌شود که در طیف برخی از مواد جانبی نیز مشاهده می‌شود [۳۳]. همان‌طور که در شکل ۵ نشان داده شده، طیف مواد اولیه در برخی نواحی با یکدیگر همپوشانی دارد اما تفاوت‌هایی در شدت جذب دیده می‌شود.



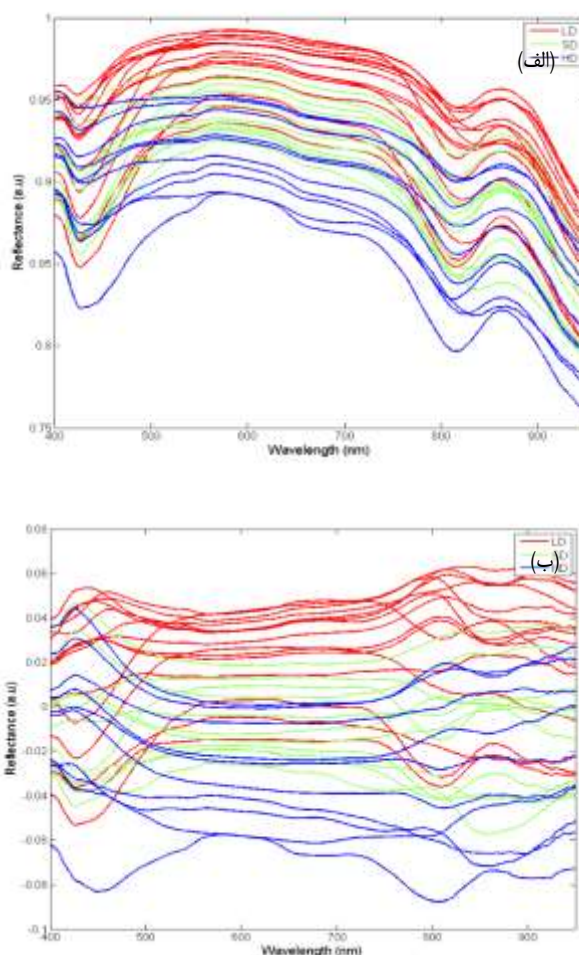
شکل ۵. میانگین طیفی مواد خام مورد استفاده در ساخت نمونه‌ها حاصل از تصویربرداری ابرطیفی با پیش پردازش هموارسازی.

این تفاوت‌های طیفی به‌ویژه در شدت جذب در طول‌موج‌های مشخص، مبنای تحلیل کیفی برای دسته‌بندی نمونه‌ها بر اساس دوزهای مختلف متفورمین قرار گرفته‌اند. در ادامه، از این تفاوت‌ها برای مدلسازی و تفکیک سه گروه مختلف قرص‌ها استفاده خواهد شد.

۳-۲- تشخیص الگو بدون نظارت با استفاده از تحلیل مؤلفه‌های اصلی

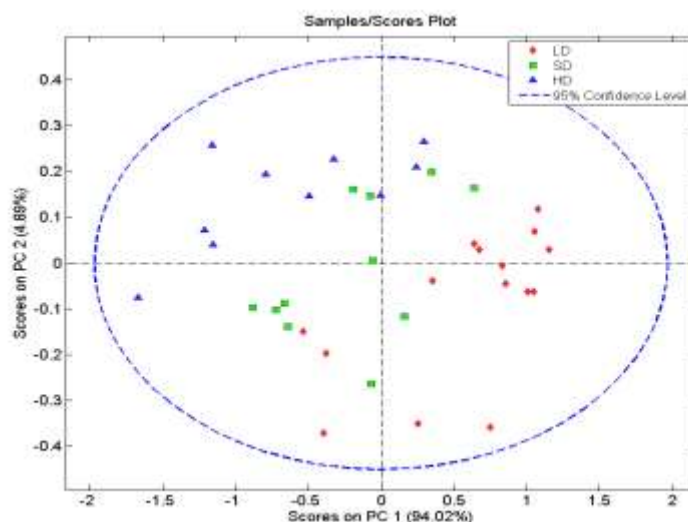
میانگین طیفی داده‌ها در شکل ۶ نمایش داده شده است. تفاوت در شدت جذب‌ها در میان ۳۶ نمونه، که دارای مقادیر متفاوتی از ماده مؤثره و ماده جانبی هیدروکسی پروپیل متیل سلولز K100M هستند، کاملاً مشهود است. پیش از انجام تحلیل مؤلفه‌های

اصلی، طیف‌ها تحت پیش‌پردازش‌هایی شامل هموارسازی و مرکزسازی نسبت به میانگین قرار گرفتند (شکل ۶ ب) تا نویز موجود در داده‌ها کاهش یابد. آنالیز تحلیل مؤلفه‌های اصلی با استفاده از سه مؤلفه اصلی انجام شد که به ترتیب 94.02، 4.89 و 0.83 درصد از واریانس داده‌ها را تبیین کردند. مقادیر RMSEC و RMSECV نیز به ترتیب برابر با 0.0017 و 0.0035 به دست آمد.



شکل ۶ الف) میانگین طیفی ۳۶ نمونه کالیراسیون حاصل از تصویر برداری ابر طیفی با پیش‌پردازش هموارسازی (ب) با پیش‌پردازش هموارسازی و مرکزسازی نسبت به میانگین.

بر اساس شکل ۷ که نمودار امتیازهای دو مؤلفه اصلی اول را نشان می‌دهد، تفکیک نمونه‌های غیراستاندارد از یکدیگر به خوبی قابل مشاهده است. همچنین، برخی از نمونه‌های دسته استاندارد که میزان ماده مؤثره آن‌ها به دو گروه غیراستاندارد نزدیک‌تر بود، در مجاورت این گروه‌ها قرار گرفتند. به طور کلی، با توجه به نتایج مطلوب حاصل از آنالیز تحلیل مؤلفه‌های اصلی، تحلیل‌های طبقه‌بندی نظارت‌شده به منظور جداسازی دقیق این سه دسته در ادامه مورد بررسی قرار گرفته‌اند.



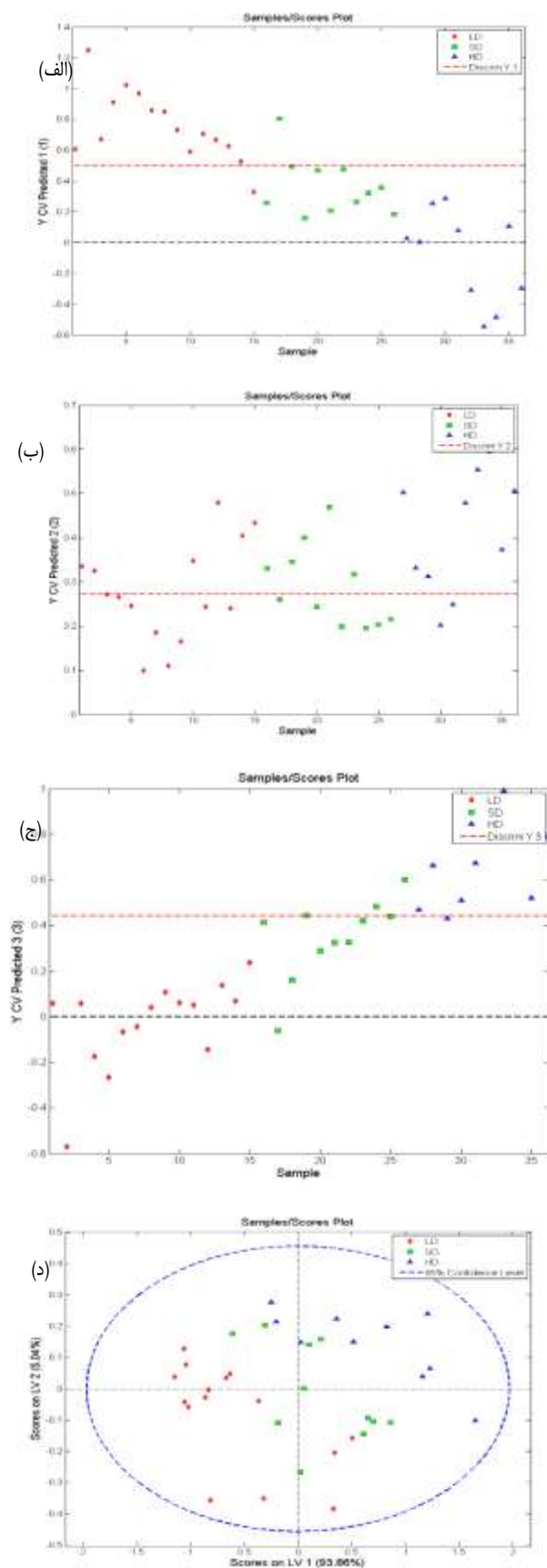
شکل ۷. نمودار امتیاز نمونه های میانگین طیفی حاصل از تصویربرداری ابرطیفی بر اساس دو PC اول.

۳-۳- تشخیص الگو با نظارت با استفاده از روش تحلیل تمایزی حداقل مربعات جزئی

پس از تحلیل مؤلفه‌های اصلی، به منظور ایجاد تمایز مشخص بین سه گروه LD، SD و HD، تحلیل تمایزی حداقل مربعات جزئی انجام شد. مشابه تحلیل پیشین، طیف‌های به دست آمده از میانگین‌گیری طیفی، پیش از آنالیز تحت پیش‌پردازش‌هایی شامل هموارسازی و مرکزسازی نسبت به میانگین قرار گرفتند تا اثر نویز کاهش یابد. در آغاز آنالیز تحلیل تمایزی حداقل مربعات جزئی، ماتریس X با ابعاد 36×540 شامل داده‌های طیفی و ماتریس Y متشکل از ۳۶ مقدار عددی تعریف شد، به گونه‌ای که عدد ۱ برای نمونه‌های LD، عدد ۲ برای SD و عدد ۳ برای HD در نظر گرفته شد. ارزیابی متقاطع به روش پرده کرکره‌ای^۱ انجام شد، به طوری که داده‌ها به ۶ دسته تقسیم و از هر دسته یک نمونه برای ارزیابی خارج گردید. نتیجه این تحلیل در شکل ۸ ارائه شده است. آنالیز با ۳ متغیر پنهان^۲ انجام شد. همان‌طور که در شکل مشاهده می‌شود، مدل موفق شده است تا حد قابل قبولی سه دسته مورد نظر را از یکدیگر تفکیک کند، به ویژه جداسازی دسته LD و HD که با وضوح بیشتری نسبت به سایر گروه‌ها انجام شده است.

^۱ Venetian Blind

^۲ LVs



شکل ۸. نمودار پیش بینی ارزیابی مدل تحلیل تمایزی حداقل مربعات جزئی برای (الف) جدایی گروه LD (ب) جدایی گروه SD (ج) جدایی گروه HD و (د) نمودار امتیاز نمونه‌های میانگین طیفی حاصل از تصویربرداری ابرطیفی بر اساس دو LV اول

ارزیابی شاخص‌های شایستگی مدل تحلیل تمایزی حداقل مربعات جزئی بر حسب درصد در جدول ۳ ارائه شده است. همان‌طور که مشاهده می‌شود، دقت و صحت مدل در هر دو مجموعه کالیبراسیون و اعتبارسنجی متقاطع، به‌ویژه در طبقه‌بندی دو کلاس غیراستاندارد LD و HD، در سطح بالایی قرار دارد که نشان‌دهنده قابلیت اطمینان بالای مدل در تشخیص این دو گروه است. با این حال، به‌طور کلی نیز مدل از دقت مناسبی در طبقه‌بندی کل نمونه‌ها برخوردار بوده است.

جدول ۳. ارقام شایستگی مدل تحلیل تمایزی حداقل مربعات جزئی

ارقام شایستگی	LD	SD	HD
حساسیت	93.3	72.7	100
ارزیابی متقاطع	93.3	36.4	90.0
ویژه گزینی	95.2	48.0	92.3
ارزیابی متقاطع	95.2	56.0	92.3
کالیبراسیون		91.6	
ارزیابی متقاطع		83.0	

۳-۴- تشخیص الگو با استفاده از روش‌های غیرخطی

به‌منظور بررسی این‌که آیا استفاده از روش‌های پیشرفته طبقه‌بندی می‌تواند به بهبود نتایج تفکیک سه گروه مورد نظر کمک کند، از دو روش شبکه عصبی مصنوعی و ماشین بردار پشتیبان استفاده شد. هر دو تحلیل با استفاده از جعبه‌ابزار برنامه یادگیرنده طبقه‌بندی در نرم‌افزار MATLAB انجام شدند.

طیف‌های به‌دست‌آمده از میانگین‌گیری طیفی، در دو حالت به کار گرفته شدند: (۱) بدون هیچ‌گونه پیش‌پردازش و (۲) با پیش‌پردازش شامل هموارسازی با پنجره ۱۵ نقطه و درجه ۰ و مرکزسازی نسبت به میانگین. داده‌های ورودی به این مدل‌ها شامل ماتریسی با ابعاد 540×36 بودند که در سه کلاس مختلف (دوز پایین، استاندارد و بالا) دسته‌بندی شدند. برای ارزیابی عملکرد مدل‌ها از اعتبارسنجی متقاطع به روش پرده کرکره‌ای با تقسیم داده‌ها به ۷ دسته استفاده شد، به‌گونه‌ای که از هر دسته یک نمونه برای آزمون انتخاب گردید. برای پیاده‌سازی و انتخاب بهترین ساختار، از قابلیت بهینه‌سازی هایپرپارامترها^۱ در MATLAB استفاده شد. در این روش، محدوده وسیعی برای پارامترهای مختلف تعریف شد و سپس مدل نهایی بر اساس اعتبارسنجی متقاطع انتخاب گردید. جزئیات بازه‌های جست‌وجو و مدل‌های انتخاب‌شده در جدول ۴ آمده است.

همان‌طور که در جدول ۴ مشخص است، فرآیند بهینه‌سازی منجر به انتخاب یک شبکه عصبی مصنوعی بسیار کوچک با ساختار زیر شد: یک لایه ورودی با ۵۴۰ نورون (مطابق با ۵۴۰ متغیر طیفی)، سه لایه پنهان با تعداد نورون‌های ۱، ۱ و ۳ و یک لایه خروجی با ۳ نورون مطابق با سه کلاس LD، SD و HD می‌باشد. همچنین تابع فعال‌سازی غیرخطی سیگموئیدی انتخاب شد. این معماری ظرفیت نمایش محدودی دارد و عملاً عملکردی مناسب با منظم‌سازی L2 نشان می‌دهد. همچنین، بهترین مدل ماشین بردار پشتیبان

^۱ Hyperparameter Optimization

یک ماشین بردار پشتیبان با کرنل خطی و روش چندکلاسه یکی در مقابل همه بود. این دو مدل نهایی به‌عنوان مدل‌های بهینه برای مقایسه با سایر روش‌های طبقه‌بندی در نظر گرفته شدند.

جدول ۴. محدوده جست‌وجوی هایپرپارامترها و مدل بهینه انتخاب‌شده برای شبکه عصبی مصنوعی و ماشین بردار پشتیبان			
مدل	هایپرپارامتر	بازه جست‌وجو ^۲	مدل انتخاب‌شده ^۱
شبکه عصبی مصنوعی	تعداد لایه‌های کاملاً متصل ^۳	۱-۳	۳
	اندازه لایه‌ها ^۴	۱-۳۰۰	[3,1,1]
	تابع فعال‌سازی ^۵	ReLU, Tanh, Sigmoid	Sigmoid
	قدرت منظم‌سازی ^۶ (λ)	2.7778e-07-2777.7778	5.255e-0.5
ماشین بردار پشتیبان	استانداردسازی داده‌ها ^۷	Yes/No	No
	روش چندکلاسه ^۸	One-vs-All, One-vs-One	One-vs-All
	تابع کرنل ^۹	Linear, Gaussian, Quadratic, Cubic	Linear
	سطح محدودیت جعبه ^{۱۰}	0.001 – 1000	2.6647
پشتیبان	مقیاس کرنل ^{۱۱}	0.001 – 1000	1
	استانداردسازی داده‌ها ^{۱۲}	Yes/No	Yes

میزان صحت دسته‌بندی ارزیابی مقاطع برای هر روش، در دو حالت با و بدون پیش‌پردازش، در جدول ۵ ارائه شده است. آنالیز شبکه عصبی مصنوعی در حالت پیش‌پردازش‌شده عملکرد بسیار مطلوبی داشت؛ به‌طوری‌که تمامی نمونه‌های کلاس HD به‌درستی شناسایی شدند و در کلاس LD نیز ۱۴ نمونه از مجموع ۱۵ نمونه به‌درستی طبقه‌بندی شدند و زمانی که روش تحلیل تمایزی حداقل مربعات جزئی در کلاس SD تنها توانسته بود ۶ نمونه از ۱۱ نمونه را به درستی طبقه‌بندی کند، این روش توانست ۹ نمونه از ۱۱ نمونه را درست طبقه‌بندی کند. در مقابل، نتایج حاصل از تحلیل ماشین بردار پشتیبان نشان داد که صحت طبقه‌بندی در دو حالت با و بدون پیش‌پردازش تفاوت قابل توجهی نداشته و عملکرد نسبتاً یکسانی ارائه شده است. این موضوع می‌تواند ناشی از استفاده از تابع کرنل خطی در ساختار مدل باشد؛ زیرا در صورت وجود مرزهای نسبتاً خطی بین کلاس‌ها، مدل ماشین بردار پشتیبان با تابع کرنل خطی به‌طور ذاتی قادر به جداسازی مناسب داده‌ها است و در چنین شرایطی، پیش‌پردازش‌هایی نظیر هموارسازی و مرکزسازی تأثیر چشم‌گیری بر عملکرد مدل نخواهند داشت.

جدول ۵. نتایج صحت آنالیز ماشین بردار پشتیبان و شبکه عصبی مصنوعی

صحت

¹ Optimized Model² Search Range³ Fully Connected Layers⁴ Layer Size⁵ Activation Function⁶ Regularization Strength⁷ Standardization⁸ Multiclass Method⁹ Kernel Function¹⁰ Box Constraint Level¹¹ Kernel Scale¹² Standardization

آنالیز	با پیش پردازش	بدون پیش پردازش
شبکه عصبی مصنوعی	۹۱/۷٪	۸۳/۳٪
ماشین بردار پشتیبان	۸۰/۶٪	۸۰/۶٪

۴- نتیجه گیری

نتایج این پژوهش نشان داد که استفاده از فناوری تصویربرداری ابرطیفی^۱ در بازه طول موجی ۴۰۰ تا ۹۵۰ نانومتر، در ترکیب با روش‌های تحلیل کمومتریکس نظیر تحلیل مؤلفه‌های اصلی و تحلیل تمایزی حداقل مربعات جزئی و همچنین الگوریتم‌های یادگیری ماشین شبکه عصبی مصنوعی و ماشین بردار پشتیبان، ابزاری بسیار مؤثر برای شناسایی دوز ماده مؤثره متفورمین در نمونه‌های پودری است. مدل تحلیل تمایزی حداقل مربعات جزئی توانست دقت قابل قبولی به‌ویژه در شناسایی نمونه‌های غیراستاندارد (LD و HD) نشان دهد، در حالی که روش شبکه عصبی مصنوعی با پیش‌پردازش مناسب بهترین عملکرد را از نظر صحت طبقه‌بندی ارائه کرد.

همچنین مشخص شد که ناحیه طیفی مرئی-نزدیک فروسرخ اطلاعات ارزشمندی در زمینه تمایز ترکیبات دارویی فراهم می‌کند. این یافته با نتایج پژوهش‌های پیشین نیز هم‌راستا است، جایی که استفاده از تصویربرداری ابرطیفی در ترکیب با تحلیل‌های کمومتریکسی برای شناسایی متفورمین و سایر ترکیبات دارویی کارایی بالایی نشان داده است [۳۵ و ۳۴]. بنابراین، روش پیشنهادی می‌تواند به‌عنوان جایگزینی سریع، غیرمخرب و دقیق در کنترل کیفیت دارویی، به‌ویژه در مراحل تولید و غربالگری داروهای ژنریک، به کار گرفته شود و از ورود محصولات غیراستاندارد به بازار جلوگیری نماید.

۵- تقدیر و تشکر

بدین‌وسیله مراتب سپاس و قدردانی خود را از معاونت پژوهشی دانشگاه صنعتی شریف جهت حمایت مالی این پروژه تشکر و قدردانی می‌نمایم. همچنین نویسندگان مقاله از جناب آقای مهندس علی آقایی عضو هیات مدیره موسسه علمی آموزشی نور به جهت در اختیار قرار دادن مواد دارویی مورد نظر تشکر می‌نمایند.

۶- فهرست منابع

- [1] Dunne, S., Close, H., & Shanahan, E. (2013). A review of the differences and similarities between generic drugs and their originator counterparts, including economic benefits associated with usage of generic medicines, using Ireland as a case study. *BMC Pharmacology and Toxicology*, 14, 1–19.
- [2] World Health Organization. (2017). *WHO global surveillance and monitoring system for substandard and falsified medical products*.

^۱ HSI

- [3] Ammassari, A., Trotta, M. P., Shalev, N., Marconi, V. C., Antinori, A., & d'Arminio Monforte, A. (2021). Comparison of demand for drugs used for COVID-19 treatment and other drugs during the early phase of the COVID-19 pandemic in Italy. *JAMA Network Open*, 4(2), e2037060.
- [4] Ranney, M. L., Griffeth, V., & Jha, A. K. (2020). Critical supply shortages—the need for ventilators and personal protective equipment during the Covid-19 pandemic. *The New England Journal of Medicine*, 382(18), e41.
- [5] Mohamed, H. M. (2015). Green, environment-friendly, analytical tools give insights in pharmaceuticals and cosmetics analysis. *Trends in Analytical Chemistry*, 66, 176–192.
- [6] Przybylska, A., Gackowski, M., & Koba, M. (2021). Application of capillary electrophoresis to the analysis of bioactive compounds in herbal raw materials. *Molecules*, 26(8), 2135.
- [7] Gu, X., Zhang, L., Wang, Y., & Chen, H. (2023). Fast screening of biomembrane-permeable compounds in herbal medicines using bubble generation technology. *Journal of Pharmaceutical Analysis*, 15(3), 123-135.
- [8] Dispas, A., Sacré, P., Ziemons, E., & Hubert, P. (2022). Emerging analytical techniques for pharmaceutical quality control: Where are we in 2022? *Journal of Pharmaceutical and Biomedical Analysis*, 221, 115071.
- [9] Seo, C.-S., & Shin, H.-K. (2022). Simultaneous analysis of 19 marker components for quality control of Oncheong-Eum using HPLC-DAD. *Molecules*, 27(9), 2992.
- [10] Zhao, X., Li, Y., Wang, J., & Zhang, R. (2023). Fast quantification of diclofenac sodium content in commercially available tablets by transmission Raman spectroscopy with PLS. *Journal of Chemometrics*, 37(2), e3462.
- [11] Wu, X., Chen, B., Wang, Y., & Li, S. (2012). Optimization of a solid phase extraction and hydrophilic interaction liquid chromatography-tandem mass spectrometry method for the determination of metformin in dietary supplements and herbal medicines. *Food Chemistry*, 133(2), 482-488.
- [12] Deidda, R., Schembri, S., Bucaria, F., & Calà, P. (2019). Vibrational spectroscopy in analysis of pharmaceuticals: Critical review of innovative portable and handheld NIR and Raman spectrophotometers. *Trends in Analytical Chemistry*, 114, 251-259.
- [13] Mirzaei, N., Rezaei, V., & Aibaghi, B. (2018). Preconcentration and determination of diazepam in pharmaceutical and biological samples by graphene oxide based on dispersive solid phase microextraction. *Applied Chemistry Today*, 14(51), 21-34. (in Persian)
- [14] Malarsoli, F. (2025). Development and Step-by-Step Optimization of the HPLC Method with Dilute-and-Shoot Sample Preparation Technique for Determination of Benidipine Hydrochloride in Pharmaceutical and Biological Samples. *Applied Chemistry Today*, 20(74), 139-160. (in Persian)
- [15] Triggle, C. R., Mohammed, I., Bshesh, K., Marei, I., & Ye, K. (2022). Metformin: Is it a drug for all reasons and diseases? *Metabolism*, 133, 155223.

- [16] Mubeen, G., & Noor, K. (2009). Spectrophotometric method for analysis of metformin hydrochloride. *Indian Journal of Pharmaceutical Sciences*, 71(1), 100-102.
- [17] Pyzowski, J., Brzezińska, E., & Kościelniak, P. (2017). Fast and convenient NIR spectroscopy procedure for determination of metformin hydrochloride in tablets. *Journal of Applied Spectroscopy*, 84(4), 710-715.
- [18] Norris, K. (1996). History of NIR. *Journal of Near Infrared Spectroscopy*, 4(1), 31-37.
- [19] Harthcock, M., & Atkin, S. (1988). Imaging with functional group maps using infrared microspectroscopy. *Applied Spectroscopy*, 42(3), 449-455.
- [20] Mehta, N., Shaik, A., & Parker, A. (2018). Single-cell analysis using hyperspectral imaging modalities. *Journal of Biomechanical Engineering*, 140(2), 020802.
- [21] Tan, M. I. S. H., Abdullah, A. Z., Rahman, N. A., & Osman, R. (2021). Ganoderma boninense disease detection by near-infrared spectroscopy classification: A review. *Sensors*, 21(9), 3052.
- [22] Sharifizadeh, G., Asghari, A., & Rajabi, M. (2019). Simultaneous analysis of Al(III), Cu(II) and Fe(II) ions in aqueous solutions: using partial least squares (PLS) method-based on UV-Vis spectrophotometry. *Applied Chemistry Today*, 14(50), 193–208. (in Persian)
- [23] Khaje Sharifi, H., & Soleimanzadegan, S. (2013). Partial least squares method for simultaneous spectrophotometric determination of uracil and 5-fluorouracil in spiked biological samples. *Applied Chemistry Today*, 7(25), 31–38. (in Persian)
- [24] Niazi, A., & Bashiri far, A. (2017). Simultaneous Determination Spectrophotometric of New fushin and Victoria blue using chemometric methods. *Applied Chemistry Today*, 12(43), 137–150. (in Persian)
- [25] The United States Pharmacopeia, Metformin Hydrochloride Extended-Release Tablets, https://www.uspnf.com/sites/default/files/usp_pdf/EN/USPNF/revisions/metformin-hcl-ert-sept18-pending-nitr.pdf , 2018 (accessed April 2024).
- [26] Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, 6(9), 2812-2831.
- [27] Oliveira, A. D. D., Silva, L. E., Castro, E. V. R., & Filho, J. C. (2018). Use of infrared spectroscopy and near infrared hyperspectral images to evaluate effects of different chemical agents on PET bottle surface. *Materials Research*, 21(5), e20170949.
- [28] Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 4-37.
- [29] Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Analytical Methods*, 5(16), 3790-3798.
- [30] Kumar, K. (2017). Principal component analysis: Most favourite tool in chemometrics. *Resonance*, 22(8), 747-759.
- [31] Brereton, R. G., & Lloyd, G. R. (2010). Support vector machines for classification and regression. *Analyst*, 135(2), 230-267.

- [32] Cogdill, R. P., & Dardenne, P. (2004). Kernel chemometrics: An introduction to support vector machines. In *Near Infrared Spectroscopy: Proceedings of the 11th International Conference* (pp. 165-170). IM Publications.
- [33] Pavia, D. L., Lampman, G. M., Kriz, G. S., & Vyvyan, J. R. (2015). *Introduction to spectroscopy* (5th ed., Chapter 2, pp. ۳۱–۸۴). Cengage Learning.
- [34] Bolhassani, Z., Aghaei, A., & Parastar, H. (2025). Quantification of metformin in pharmaceutical formulations using Vis-SWNIR hyperspectral imaging combined with multivariate curve resolution. *Chemometrics and Intelligent Laboratory Systems*, 250, 104918.
- [35] Xia, Y., Cen, H., He, Y., & Xu, L. (2021). Rapid quantification and visualization of metformin hydrochloride tablets using hyperspectral imaging. *Chemometrics and Intelligent Laboratory Systems*, 212, 104287.

