

Analyzing Meat Samples: A Comparative Evaluation of Linear versus Nonlinear Machine Learning Techniques

Somaye Vali Zade¹, Hossein Rastegar²

¹ Halal Research Center of IRI, Food and Drug Administration, Ministry of Health and Medical Education, Tehran, Iran. smy.valizade@gmail.com

² Cosmetic Products Research Center, Food and Drug Administration, Ministry of Health and Medical Education, Tehran, Iran (**Corresponding author**).
mhrastegar2@yahoo.com

Abstract

Objective: The moisture, protein, and fat content in meat are among the most critical characteristics determining its quality. Traditional methods for measuring these attributes are typically time-consuming, destructive, and costly. Substituting these with spectroscopic methods could be a significant step toward developing rapid techniques in this field. Combining linear or nonlinear machine learning methods with near-infrared spectroscopy enables the quantitative determination of these components in meat samples. This study aims to compare the performance of linear and nonlinear machine learning methods in determining these three attributes in meat samples.

Materials and Methods: In this study, near-infrared (NIR) spectral data from 240 ground meat samples were collected using an NIR spectrometer in the wavelength range of 850 to 1050 nm. Of these, 170 samples were used for calibration and 70 for testing. Two machine learning methods, partial least squares (PLS) as a linear method and radial basis function artificial neural network (RBF-ANN) as a nonlinear method, were employed to model and predict the concentrations of moisture, protein, and fat. Modeling was performed using MATLAB 2018b with the MVC1 package.

Findings: Spectral data from 240 meat samples were analyzed using partial least squares (PLS) and radial basis function artificial neural network (RBF-ANN) methods. A total of 170 samples were used for calibration, and 70 samples were used for testing. The RMSEP values for predicting moisture, protein, and fat using RBF-ANN were 0.8500, 0.6538, and 0.6761, respectively, and for PLS, they were 2.3132, 2.8023, and 0.7615, respectively. The REP% values for RBF-ANN were 1.3494%, 3.5680%, and 3.8254% for moisture, protein, and fat, respectively, while for PLS, they were 3.6725%, 15.2922%, and 4.3083%, respectively.

Cite this article: Vali Zade, S. & Rastegar, H. (2023). Analyzing Meat Samples: A Comparative Evaluation of Linear versus Nonlinear Machine Learning Techniques. *Biannual journal of halal*, 6(1), p. 71-86. <https://doi.org/10.30502/H.2024.447237.1135>

Received: 2022/09/05 ; **Revised:** 2022/10/31 ; **Accepted:** 2022/01/07 ; **Published online:** 2023/04/09

Article type: Research Article

© the authors

Publisher: Halal Research Center of Islamic Republic of Iran



Conclusion: Partial least squares (PLS) and radial basis function artificial neural network (RBF-ANN) methods were applied to analyze spectral data from 240 meat samples. The variables of each machine learning method were optimized during the calibration phase using near-infrared spectra from 140 meat samples, and the corresponding linear and nonlinear models were used to predict the concentrations of three analytes (moisture, protein, and fat) in 70 independent meat samples. Although the nonlinear RBF-ANN model, as an artificial intelligence approach, demonstrated higher efficiency in predicting concentrations, the linear PLS model also provided acceptable performance in determining fat content in meat samples.

Keywords: Machine learning, Partial least squares, Artificial neural networks, Meat.

تجزیه و تحلیل نمونه‌های گوشت: ارزیابی مقایسه‌ای روش‌های یادگیری ماشین خطی و غیر خطی

سمیه ولی‌زاده^۱، حسین رستگار^۲

^۱ مرکز تحقیقات حلال جمهوری اسلامی ایران، سازمان غذا و دارو، وزارت بهداشت، درمان و آموزش پزشکی، تهران، ایران. smy.valizade@gmail.com

^۲ مرکز تحقیقات فرآورده‌های آرایشی و بهداشتی، سازمان غذا و دارو، وزارت بهداشت، درمان و آموزش پزشکی، تهران، ایران (نویسنده مسئول). mhrastegar2@yahoo.com

چکیده

سابقه و هدف: مقدار رطوبت، پروتئین و چربی در گوشت از مهم‌ترین ویژگی‌های این ماده غذایی در تعیین کیفیت آن است. روش‌های سنتی اندازه‌گیری این ویژگی‌ها معمولاً زمان‌بر، تخریبی و پرهزینه هستند. جایگزینی روش‌های طیف‌سنجی می‌تواند گام مهمی در ابداع روش‌های سریع در این حوزه باشد. ترکیب روش‌های یادگیری ماشین خطی یا غیرخطی با اندازه‌گیری‌های طیف‌سنجی زیرقرمز نزدیک، امکان اندازه‌گیری این مولفه‌ها را در نمونه‌های گوشت ایجاد می‌کند. هدف این پژوهش مقایسه کارآمدی‌های دو روش یادگیری ماشین خطی و غیرخطی در تعیین سه ویژگی در نمونه‌های گوشت است.

مواد و روش‌ها: در این پژوهش، داده‌های طیفی نزدیک به زیرقرمز (NIR) از ۲۴۰ نمونه گوشت چرخ‌شده با استفاده از طیف‌سنج NIR در محدوده طول موج ۸۵۰ تا ۱۰۵۰ نانومتر ثبت شد. از این تعداد، ۱۷۰ نمونه برای کالیبراسیون و ۷۰ نمونه برای آزمون استفاده شدند. دو روش یادگیری ماشین، شامل حداقل مربعات جزئی (PLS) به‌عنوان روش خطی و شبکه عصبی مصنوعی با تابع پایه شعاعی (RBF-ANN) به‌عنوان روش غیرخطی، برای مدل‌سازی و پیش‌بینی غلظت رطوبت، پروتئین و چربی به کار برده شدند. مدل‌سازی با نرم‌افزار MATLAB 2018b و بسته MCV1 انجام شد.

یافته‌ها: داده‌های طیفی ۲۴۰ نمونه گوشت با استفاده از روش‌های حداقل مربعات جزئی (PLS) و شبکه عصبی مصنوعی - تابع پایه شعاعی (RBF-ANN) تحلیل شد. برای کالیبراسیون، ۱۷۰ نمونه و برای آزمون، ۷۰ نمونه استفاده شد. مقادیر RMSEP برای پیش‌بینی رطوبت، پروتئین و چربی با RBF-ANN به ترتیب ۸۵۰۰/۰، ۶۵۳۸/۰ و ۶۷۶۱/۰ و با PLS به ترتیب ۳۱۳۲/۲، ۸۰۲۳/۲ و ۷۶۱۵/۰ محاسبه شد. مقادیر REP% برای RBF-ANN در رطوبت، پروتئین و چربی به ترتیب ۳۴۹۴/۱، ۵۶۸۰/۳ و ۸۲۵۴/۳ درصد و برای PLS به ترتیب ۶۷۲۵/۳، ۲۹۲۲/۱۵ و ۳۰۸۳/۴ درصد بود.

استناد به این مقاله: ولی‌زاده، سیمیه؛ رستگار، حسین (۱۴۰۲). تجزیه و تحلیل نمونه‌های گوشت: ارزیابی مقایسه‌ای روش‌های یادگیری ماشین خطی و غیرخطی. *پژوهشنامه حلال*، ۱۶(۱)، ص ۷۱-۸۶. <https://doi.org/10.30502/H.2024.447237.1135>

تاریخ دریافت: ۱۴۰۱/۰۶/۱۴؛ تاریخ بازنگری: ۱۴۰۱/۰۸/۰۹؛ تاریخ پذیرش: ۱۴۰۱/۱۰/۱۷؛ تاریخ انتشار: ۱۴۰۲/۰۱/۲۰

نوع مقاله: پژوهشی

© نویسندگان

ناشر: مرکز ملی تحقیقات حلال جمهوری اسلامی ایران



نتیجه‌گیری: روش‌های حداقل مربعات جزئی (PLS) و شبکه عصبی مصنوعی- تابع پایه شعاعی (RBF-ANN) برای تجزیه و تحلیل داده‌های طیفی از ۲۴۰ نمونه گوشت به کار برده شدند. متغیرهای هر کدام از ماشین‌های یادگیری فوق در مرحله کالیبراسیون طیف‌های زیرقرمز نزدیک ۱۴۰ نمونه گوشت، در موقعیت‌های بهینه خود تنظیم شدند و از مدل‌های خطی و غیرخطی متناظر برای پیش‌بینی غلظت سه آنالیت (محتوای رطوبت، پروتئین و چربی) در ۷۰ نمونه مستقل گوشت استفاده شد. هرچند ماشین غیرخطی RBF-ANN به عنوان یک مدل هوش مصنوعی، کارآمدی بیشتری را در پیش‌بینی غلظت نشان داد، ولی ماشین خطی PLS هم در تعیین چربی در نمونه‌های گوشت، بازده قابل قبولی را ارائه داد.

کلیدواژه‌ها: یادگیری ماشین، حداقل مربعات جزئی، شبکه‌های عصبی، هوش مصنوعی، گوشت.

۱. مقدمه

رشد اقتصاد جهانی که منجر به افزایش تقاضای گسترده برای گوشت با کیفیت عالی شده، یکی از عوامل مهم در توسعه سریع صنعت گوشت در چند دهه اخیر است (۱). علاوه بر این، به دلیل فشارهایی که توسط جهانی شدن و افزایش ارتباطات به وجود آمده، ارائه محصولات با کیفیت به مصرف‌کنندگان ضروری شده است، به نحوی که مصرف‌کنندگان مایل به پرداخت قیمت‌های بالاتر برای محصولات با تضمین کیفیت بیشتر هستند (۲). آنالیز ترکیب شیمیایی گوشت به شکل سنتی توسط روش‌های شیمیایی انجام می‌شود که اغلب تخریبی و زمان‌بر بوده و البته برخی از آن‌ها روش‌های مرجع پذیرفته‌ای شده هستند. اهمیت جایگزینی روش‌های تجزیه‌ای سریع و غیرتخریبی مانند طیف‌سنجی زیر قرمز نزدیک^۱ یا رزونانس مغناطیسی هسته‌ای^۲ اخیراً مورد توجه ویژه قرار گرفته است (۳، ۴). محتوای رطوبت، پروتئین و چربی به‌عنوان مهم‌ترین ویژگی‌های فیزیکوشیمیایی گوشت شناخته شده است؛ زیرا نقش اساسی در ارزش تغذیه‌ای و مزه گوشت دارند. توسعه روش‌های تعیین غیرتخریبی و سریع این اجزا در گوشت می‌تواند اطلاعات بسیار مفیدی را در مورد کیفیت گوشت ارائه دهد. روش‌های مبتنی بر یادگیری ماشین^۳ و هوش مصنوعی^۴ در سال‌های اخیر مورد توجه فزاینده‌ای در زمینه‌های مختلف قرار گرفته‌اند. بسیاری از روش‌هایی که اخیراً در این زمینه‌ها مورد اقبال پژوهشگران قرار گرفته، سال‌ها پیش توسط محققین کمومتریکس ابداع و یا مورد استفاده بوده است. آنالیز مخلوط‌ها با روش‌های طیف‌سنجی که گزینش‌پذیری کمتری دارند، مانند داده‌های طیف‌سنجی UV-VIS و NIR، قدمتی بیش از ۵۰ سال دارد (۵، ۶). روش‌های کالیبراسیون چندمتغیره به منظور آنالیز مخلوط‌های پیچیده، در تحقیقات متعددی در حوزه کمومتریکس پیشنهاد، بررسی و به کار برده شده‌اند. ماشین‌های یادگیری مبتنی بر رگرسیون مانند حداقل مربعات جزئی^۵ توسط کمومتریکس دانه ابداع شده است (۷). در حالت ایده‌آل، ارتباط بین سیگنال‌های طیفی در روش‌های طیف‌سنجی با غلظت گونه‌های شیمیایی در مخلوط‌ها، از توابع خطی پیروی می‌کنند. در این حالت روش‌های یادگیری ماشین که بر مبنای رگرسیون‌های چندمتغیره خطی هستند، گزینه‌های مناسبی برای بنا کردن مدل‌های کالیبراسیون می‌باشند. دلایل متفاوتی می‌توانند باعث شوند که ارتباط بین شدت طیف‌ها و غلظت اجزاء، از معادلات خطی تبعیت نکنند. در این

1. NIR

2. NMR

3. ML

4. AI

5. PLS

صورت نیاز به روش‌های یادگیری ماشین غیرخطی ایجاد می‌شود. شبکه‌های عصبی مصنوعی^۱، روش‌های مدل‌سازی غیرخطی هستند که براساس نحوه یادگیری مغز انسان شبیه‌سازی شده و در حوزه روش‌های هوش مصنوعی^۲ قرار می‌گیرند (۸، ۹، ۱۰). در این پژوهش دو روش یادگیری ماشین متداول و بسیار پر کاربرد، یعنی رگرسیون حداقل مربعات جزئی^۳ و شبکه عصبی مصنوعی - تابع پایه شعاعی^۴ برای تعیین کمی رطوبت، پروتئین و چربی در نمونه‌های گوشت، مورد مقایسه و بررسی قرار گرفته‌اند. PLS به عنوان روش یادگیری ماشین خطی، RBF-ANN به عنوان روش یادگیری ماشین غیرخطی مبتنی بر هوش مصنوعی و داده‌های NIR نمونه‌های گوشت، به عنوان مجموعه داده‌های طیف‌سنجی با رفتار غیرخطی انتخاب شده‌اند. موضوعی که در این پژوهش مورد توجه قرار گرفته، مقایسه کارآمدی روش خطی PLS با نتایج مدل آموزش دیده RBF-ANN است.

۲. پیش‌زمینه تئوری

۲-۱. شبکه‌های عصبی مصنوعی با توابع پایه شعاعی (RBF-ANN)

شبکه‌های عصبی مصنوعی^۵ می‌توانند به عنوان الگوریتم‌های غیرپارامتریک چندمتغیره رگرسیون تعریف شوند. بعد از آموزش مناسب، آن‌ها قادر به یادگیری یک قاعده هستند که یک مجموعه از متغیرهای مستقل را با یک پاسخ مطلوب مرتبط می‌کند، به همین دلیل، در زمینه کالیبراسیون چندمتغیره، این الگوریتم‌ها به طور گسترده در سیستم‌های غیرخطی که رابطه‌ای غیرخطی بین سیگنال و غلظت یا ویژگی آنالیت وجود دارد، استفاده شده‌اند. تاکنون انواع زیادی از شبکه‌های عصبی مصنوعی توسعه یافته‌اند که می‌توانند براساس معیارهای مختلف مربوط به اهداف، معماری و الگوریتم‌های یادگیری به طور سیستماتیک، دسته‌بندی شوند. شبکه‌های عصبی تابع پایه شعاعی^۶ یکی از الگوریتم‌های متداول در این زمینه هستند. در شبکه‌های عصبی مصنوعی، واحدهای عملیاتی به نام نورون‌ها (گره‌ها) در سه لایه ورودی، پنهان و خروجی قرار دارند. نورون‌های ورودی و خروجی به ترتیب متناظر با سیگنال‌های اندازه‌گیری شده در ماتریس داده و پاسخ‌های قابل

1. ANN

2. AI

3. PLS

4. RBF-ANN

5. ANN

6. RBF

پیش‌بینی توسط مدل هستند، در حالی که نورون‌های پنهان در مرکز الگوریتم محاسباتی قرار دارند. در طول آموزش، هر نورون پنهان و خروجی، ورودی‌های وزن‌دار را دریافت می‌کند، سپس وارد یک تابع غیرخطی (تابع انتقال) می‌شوند، تا خروجی غیرخطی تولید شود. این شبکه‌های عصبی مصنوعی با تنظیم وزن‌های مربوط به اتصالات بین نورون‌ها، رابطه بین سیگنال‌ها و غلظت‌ها را یاد می‌گیرند، به گونه‌ای که خروجی نهایی، به مقدار غلظت معلوم برای آنالیت نزدیک باشد. از آنجا که سیگنال‌های طیفی مورد استفاده معمولاً در تعداد زیادی طول موج ثبت شده‌اند، به طور معمول لایه ورودی حاوی تعدادی نورون بوده که با تعداد اسکورها‌های مرتبط پس از اعمال PCA روی ماتریس کالیبراسیون برابر است.

شبکه‌های عصبی مصنوعی با توابع پایه شعاعی (RBF-ANN) از یک تابع پایه شعاعی به عنوان تابع انتقال در نورون‌های لایه پنهان استفاده می‌کنند، که به طور معمول از توابع گوسی استفاده می‌شود. پس از تبدیل غیرخطی ورودی‌ها، یک نورون خروجی مجموع وزن‌داری از خروجی‌های نورون‌های پنهان را تولید می‌کند، تا خروجی نهایی شبکه به دست آید. شبکه‌های عصبی مصنوعی با توابع پایه شعاعی (RBF-ANN) به طور گسترده در کالیبراسیون چندمتغیره غیرخطی استفاده می‌شوند. در زمینه کالیبراسیون چندمتغیره، RBF-ANN‌ها در برقراری روابط پیچیده بین متغیرهای مستقل و متغیرهای پاسخ بدون فرض مدل خاص موفق هستند. این انعطاف‌پذیری آن‌ها را برای کاربردهایی مناسب می‌کند که در آن‌ها روش‌های خطی کلاسیک نمی‌توانند الگوها و روابط پیچیده در داده‌ها را به طور مناسب مدل کنند. در RBF-ANN‌ها برای کالیبراسیون چندمتغیره، هر نورون ورودی، یک اسکور خاص برای یک نمونه دریافت می‌کند، که معمولاً از تحلیل مؤلفه اصلی^۱ بدست می‌آید، در واقع تعداد نورون‌های ورودی، به تعداد اسکورها‌های حاصل از اعمال PCA بر روی داده‌های مجموعه آموزش است. لایه پنهان شامل توابع گوسی است که توسط ورودی‌ها فعال می‌شوند. توابع گوسی که توسط یک مرکز و یک عرض مشخص می‌شوند، به طور کلی مبنای تبدیل غیرخطی داده‌های ورودی هستند. مراکز این توابع در یک ماتریس سازمان‌دهی می‌شوند، و در فرمول‌بندی معمولی، عرض توابع گوسی یکسان است. در طول آموزش، شبکه وزن‌های مرتبط با اتصالات میان نورون‌ها تنظیم می‌شوند، تا رابطه بین سیگنال‌های ورودی و خروجی هدف را یاد بگیرد، هدف در این یادگیری به حداقل رسیدن اختلاف بین مقادیر پیش‌بینی شده و واقعی پاسخ‌ها هستند. این رویکرد اجازه می‌دهد تا RBF-ANN‌ها به طور موثر در مدل کردن روابط غیرخطی در

1. PCA

مسائل کالیبراسیون چندمتغیره عمل کنند، و ابزاری قدرتمند برای تحلیل و پیش‌بینی در زمینه‌های مختلف باشد (۱۱، ۱۲).

۲-۲. کالیبراسیون حداقل مربعات جزئی^۱

روش حداقل مربعات جزئی یکی از روش‌های پرکاربرد برای کالیبراسیون خطی چندمتغیره در زمینه‌های مختلف، به ویژه در شیمی تجزیه و کمومتریکس است. PLS برای مدل‌سازی رابطه بین متغیرهای مستقل و متغیر پاسخ (هدف)، با تجزیه هر دو ماتریس (متغیرهای مستقل و وابسته) به متغیرها یا فاکتورهای پنهان، طراحی شده است. برخلاف رگرسیون کوچک‌ترین مربعات کلاسیک (CLS) که به طور مستقیم با متغیرهای مستقل کار می‌کند، PLS با در نظر گرفتن همبستگی بین متغیرهای مستقل و متغیر پاسخ عمل می‌کند. با ماکزیمم کردن همبستگی بین این دو مجموعه متغیر، PLS فاکتورهای پنهانی استخراج می‌کند که حداکثر واریانس را همزمان در هر دو ماتریس مستقل و پاسخ آن‌ها نشان می‌دهند. این فاکتورهای پنهان، الگوها یا ساختارهای نهفته در داده را نمایش می‌دهند که بیشترین ارتباط با پیش‌بینی متغیر پاسخ را دارند. در PLS، فاکتورهای پنهان به صورت تکراری استخراج می‌شوند، به طوری که هر فاکتور بخشی از واریانس متغیرهای مستقل و پاسخ را به خود اختصاص می‌دهد. تعداد فاکتورهایی که باید نگه داشته شود، براساس معیارهای مختلفی مانند اعتبارسنجی متقاطع تعیین می‌شود. پس از تعیین فاکتورهای پنهان، PLS یک مدل رگرسیون بین متغیرهای مستقل و پاسخ با استفاده از این فاکتورها ایجاد می‌شوند. سپس این مدل برای پیش‌بینی متغیر پاسخ (غلظت آنالیت) برای نمونه‌های جدید براساس اسکورهای متغیر مستقل آن‌ها استفاده می‌شود. PLS به ویژه در مواقعی مفید است که تعداد متغیرهای مستقل زیاد هستند، و از تعداد نمونه‌ها فراتر می‌روند. قابلیت مدیریت چنین پیچیدگی‌هایی، PLS را به ابزار مهمی برای کالیبراسیون خطی چندمتغیره در زمینه‌های مختلف تبدیل می‌کند (۷).

۳. بخش تجربی

۳-۱. داده‌های NIR تکاتور^۲

این داده‌ها شامل طیف‌های جذب NIR هستند که برای پیش‌بینی مقادیر رطوبت، چربی و پروتئین گوشت چرخ شده استفاده می‌شوند. این طیف‌ها با استفاده از یک تجزیه‌گر مواد غذایی و خوراکی اینفراتک تکاتور در محدوده طول موج ۸۵۰-۱۰۵۰ نانومتر ثبت شده‌اند. آزمایش با استفاده از یک

1. PLS

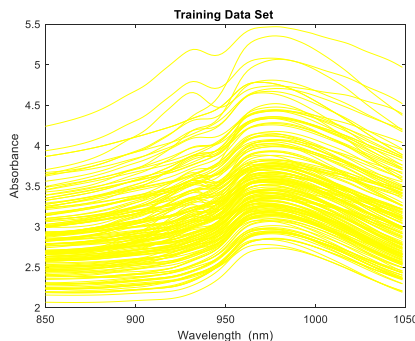
2. Tecator

طیف‌سنج NIR از شرکت LT Industries انجام شد. این طیف‌سنج با یک پروب واردکننده که می‌تواند در یک توده گوشت وارد شود، یا در یک مایع غوطه‌ور شود، تجهیز شده است. نور از طریق فیبر نوری منتشر و تشخیص داده می‌شود. طیف‌سنج طیفی با بازه طول موجی ۹۰۰ تا ۱۸۰۰ نانومتر را در ۱۲۰۱ طول موج ثبت می‌کند. هر نمونه شامل گوشت خالص چرخ شده با محتوای مختلف رطوبت، چربی و پروتئین است. این مجموعه داده معمولاً به دلیل وجود ارتباط غیرخطی بین سیگنال و ویژگی‌های اندازه‌گیری شده، برای ارزیابی عملکرد مدل‌های کالیبراسیون چندمتغیره مختلف استفاده می‌شود (۱۳، ۱۴). این مجموعه داده به صورت رایگان در بسته نرم‌افزاری MVC1 در دسترس است (۱۵، ۱۶). داده‌ها شامل ۲۴۰ طیف NIR بوده که هر کدام در ۱۰۰ طول موج ثبت شده است. مقادیر متناظر رطوبت، چربی و پروتئین برای هر نمونه با روش‌های استاندارد تعیین شده است. در این تحقیق نمونه‌ها به طور تصادفی به ۱۷۰ طیف برای کالیبراسیون و ۷۰ طیف برای آزمون تقسیم شدند. همه مدل‌سازی‌های PLS و RBF-ANN با نرم‌افزار MATLAB 2018b و با بسته MVC1 انجام شده است (۱۵، ۱۶).

۴. نتایج و بحث

برقراری ارتباط بین سیگنال‌های اندازه‌گیری شده با دستگاه‌های طیف‌سنجی و غلظت یا ویژگی آنالیت، از طریق کالیبراسیون انجام می‌گیرد. در شرایط ایده‌آل این ارتباط از یک تابع خطی تبعیت می‌کند؛ یعنی با تغییر در غلظت آنالیت سیگنال طیف‌سنجی آن در نواحی فعال طیفی نیز به همان نسبت تغییر می‌کند. مدل‌های کالیبراسیون متفاوتی براساس اینگونه ارتباط‌ها بنا شده‌اند. روش کالیبراسیون چندمتغیره PLS از یک تابع خطی برای مدل‌سازی استفاده می‌کند. فاکتورهای PLS که از ترکیب خطی سیگنال‌های طیف‌سنجی ساخته می‌شوند، از طریق یک مدل رگرسیون چندگانه خطی در مدل کالیبراسیون وارد می‌شوند. به دلایل متفاوتی در سیستم‌های پیچیده شیمیایی مخلوط‌ها ممکن است ارتباطی بین غلظت هر آنالیت با داده‌های طیفی ثبت شده خطی نباشد. در این حالت واضح است که سیگنال مرتبط به آنالیت به همان نسبتی که غلظت آن تغییر می‌کند، تحت تاثیر قرار نمی‌گیرد. کالیبراسیون داده‌های غیرخطی امکان آنالیز مخلوط‌های پیچیده و افزایش دامنه قابل کاربرد مدل‌های کالیبراسیون را ایجاد می‌کند. سوالی که پیش می‌آید این است که آیا روش‌های کالیبراسیون خطی مانند PLS قادر هستند با تغییر در پارامترهای قابل تنظیم مدل مانند تعداد فاکتورها در آنالیز داده‌های غیرخطی مورد استفاده قرار گیرند؟ شبکه‌های عصبی مصنوعی^۱ از

طریق وارد کردن توابع غیرخطی در پردازش داده‌های کالیبراسیون، حتی ارتباط‌های پیچیده غیرخطی را امکان‌پذیر می‌کنند. این ماشین‌های یادگیری در آنالیز مخلوط‌ها، در دامنه گسترده‌ای از غلظت آنالیت‌ها موفق عمل کرده و به ابزار قدرتمندی در دست پژوهشگران برای مدل‌سازی در سیستم‌های غیرخطی مبدل شده‌اند. بنابراین، روش RBF-ANN می‌تواند به عنوان مدل مرجع در کالیبراسیون غیرخطی در آنالیز این سیستم‌ها مورد استفاده قرار گیرد. بنابراین، می‌توان کارآمدی کالیبراسیون‌های خطی در مدل‌سازی داده‌های غیرخطی را با نتایج مدل‌های RBF-ANN مقایسه و ارزیابی کرد. در این پژوهش آنالیز مخلوط‌های پیچیده توسط کالیبراسیون خطی PLS و مدل‌سازی غیرخطی RBF-ANN تحت شرایط بهینه مورد مقایسه قرار می‌گیرند. شکل (۱) داده‌های طیف‌سنجی NIR برای ۱۷۰ نمونه گوشت چرخ شده را نشان می‌دهد. غلظت سه آنالیت (مقدار رطوبت، پروتئین و چربی) در این نمونه‌ها توسط روش‌های استاندارد، تعیین شده‌اند (۴) و امکان بنا کردن مدل‌های کالیبراسیون را میسر می‌سازند. ارتباط غیرخطی بین غلظت آنالیت‌ها و طیف‌های NIR در این داده‌ها قبلاً مورد توجه پژوهشگران قرار گرفته است (۴، ۵). بنابراین، در ادامه نتایج کالیبراسیون‌های خطی PLS و غیرخطی RBF-ANN در آنالیز این داده‌ها برای پیش‌بینی غلظت رطوبت، پروتئین و چربی در نمونه‌های گوشت، مورد ارزیابی و بررسی قرار می‌گیرد.



شکل ۱- طیف‌های NIR مربوط به ۱۴۰ نمونه گوشت که برای مرحله یادگیری در ساخت مدل‌های RBF-ANN و PLS به کار برده شده‌اند.

۵. مدل‌سازی یادگیری ماشین با شبکه‌های عصبی مصنوعی - تابع پایه شعاعی

طیف‌های ثبت شده از ۱۷۰ نمونه گوشت (شکل ۱) برای مرحله آموزش شبکه عصبی مورد استفاده قرار گرفت. هر طیف در ۱۰۰ طول موج ثبت شده است. برای پیشگیری از بیش‌برازش^۱ مدل،

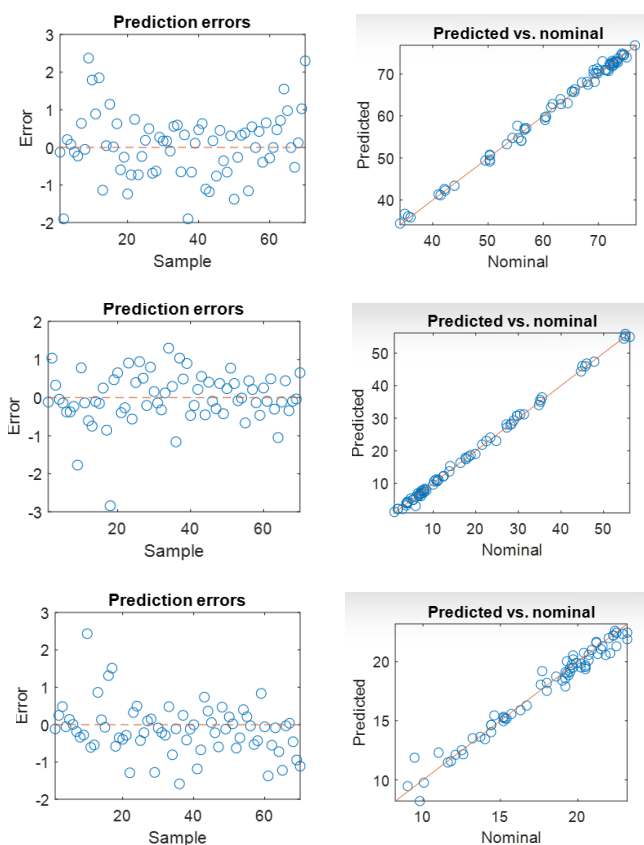
1. Over fitting

نیاز به کاهش تعداد متغیرها هست. از آنالیز مولفه‌های اصلی^۱ برای کاهش ابعاد داده و متعاقب آن کاهش تعداد نورون‌های ورودی استفاده شد. برنامه MVC1 از مدل کالیبراسیون رگراسیون مولفه‌های اصلی^۲ و الگوریتم اعتبارسنجی متقاطع^۳ برای تخمین تعداد مولفه‌های اصلی مناسب استفاده می‌کند. همچنین MVC1 تعداد نورون‌های لایه پنهان را نیز به صورت خودکار بهینه می‌نماید. تعداد نورون لایه خروجی نیز همیشه (۱) بوده که متناظر با غلظت آنالیتی است که مدل کالیبراسیون برای آن بنا شده است. به این ترتیب معماری شبکه عصبی مصنوعی در مرحله بهینه‌سازی برای هر آنالیت تعیین می‌شود. سه ماشین یادگیری RBF-ANN متناظر با سه آنالیت رطوبت، چربی و پروتئین آموزش داده شدند که معماری آنها به ترتیب ۱۳-۴۴-۱، ۱۴-۶۲-۱ و ۱۲-۴۱-۱ هستند. این اعداد به ترتیب متناظر با تعداد نورون‌ها در لایه خروجی-لایه پنهان-لایه ورودی هستند. معیار کارآمدی ماشین طراحی شده در یادگیری ارتباط بین طیف‌ها و غلظت آنالیت در نمونه‌های آموزش را می‌توان بر مبنای میانگین ریشه دوم خطای کالیبراسیون^۴ ارزیابی کرد. مقادیر این خطا برای سه مدل آموزش دیده به ترتیب ۰/۵۶۱۵۷، ۰/۳۷۲۴۴ و ۰/۴۶۲۵۸ برای رطوبت، پروتئین و چربی به دست آمدند که نشان‌دهنده موفق بودن مرحله یادگیری در سه ماشین طراحی شده است.

۷۰ نمونه مستقل گوشت برای ارزیابی قدرت پیش‌بینی مدل‌های RBF-ANN آموزش دیده، مورد استفاده قرار گرفتند. شکل (۲) نتایج بررسی ارتباط بین غلظت‌های پیش‌بینی شده و غلظت‌های معلوم برای ۷۰ نمونه پیش‌بینی را برای هر سه آنالیت نمایش می‌دهد. در ستون اول، غلظت‌های پیش‌بینی شده با ماشین‌های آموزش دیده، نسبت به غلظت‌های معلوم آنالیت‌ها ترسیم شده‌اند. قرار گرفتن مطلوب نقاط روی خط، موفقیت ماشین‌های یادگیری را در پیش‌بینی غلظت آنالیت‌ها در ۷۰ نمونه مستقل نمایش می‌دهد. همچنین در ستون دوم شکل (۲) توزیع خطای مطلق برای غلظت هر آنالیت در ۷۰ نمونه پیش‌بینی نشان داده شده است. پخش تصادفی نقاط در اطراف خط متناظر با خطای مطلق صفر، حاکی از قدرت ماشین‌های آموزش دیده در پیش‌بینی غلظت این آنالیت‌ها است. معیار دیگری که می‌تواند در ارزیابی قدرت مدل‌های RBF-ANN در پیش‌بینی غلظت آنالیت‌ها در نمونه‌های مستقل مورد توجه قرار گیرند، میانگین ریشه دوم خطای

1. PCA
2. PCR
3. CV
4. RMSEC

پیش‌بینی^۱ است. مقادیر RMSEP برای رطوبت، پروتئین و چربی در ۷۰ نمونه گوشت به ترتیب ۰/۸۵۰۰، ۰/۶۵۳۸ و ۰/۶۷۶۱ محاسبه شدند. معیار درصد خطای نسبی پیش‌بینی (%REP) که نسبت RMSEP به مقدار میانگین غلظت آنالت در نمونه‌های پیش‌بینی را نمایندگی می‌کند نیز می‌تواند مبنای مناسبی برای مقایسه باشد. مقادیر %REP برای رطوبت، پروتئین و چربی در ۷۰ نمونه پیش‌بینی به ترتیب ۱/۳۴۹۴، ۳/۵۶۸۰ و ۳/۸۲۵۴ درصد محاسبه شدند. قدرت ماشین‌های یادگیری غیرخطی طراحی شده در دستیابی به درصد‌های خطای نسبی در پیش‌بینی نمونه‌های مستقل، به خوبی در نتایج قابل مشاهده است.



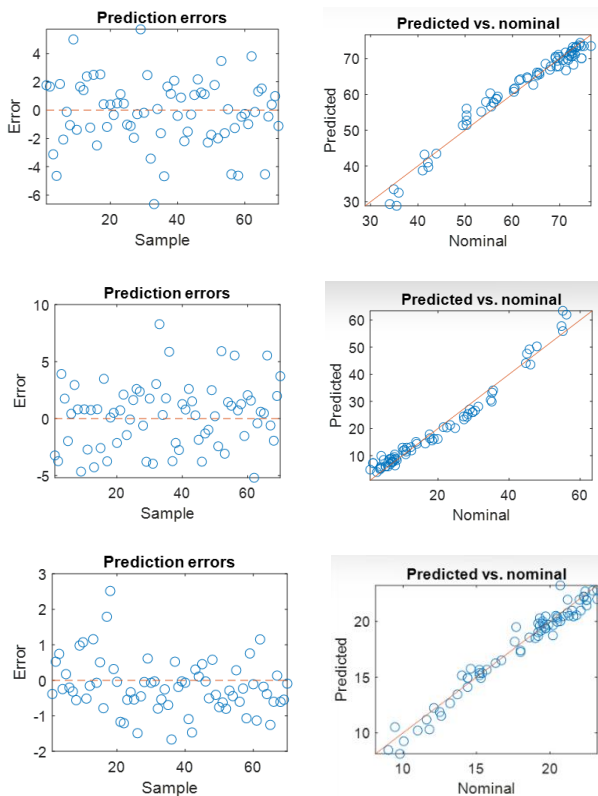
شکل ۲- نتایج پیش‌بینی غلظت رطوبت (ردیف اول)، پروتئین (ردیف دوم) و چربی (ردیف سوم) در ۷۰ نمونه گوشت با مدل‌های بهینه شده RBF-ANN. پلات‌های غلظت‌های پیش‌بینی شده نسبت به غلظت‌های واقعی (ستون اول) و میزان خطای مطلق پیش‌بینی غلظت نمونه‌ها (ستون دوم).

1. RMSEP

۶. مدل‌سازی یادگیری ماشین با رگرسیون حداقل مربعات جزئی

میزان موفقیت یک ماشین یادگیری خطی مثل PLS می‌تواند از مقایسه معیارهای کارآمدی آن با نتایج یک ماشین یادگیری غیرخطی مثل RBF-ANN مورد ارزیابی قرار گیرد. به این منظور مراحل آموزش و پیش‌بینی ماشین PLS هم دقیقاً با همان نمونه‌هایی که در بخش قبلی با شبکه عصبی مصنوعی استفاده شدند، تکرار شدند. ۱۷۰ نمونه در ساخت مدل کالیبراسیون مشارکت داده شدند. تعداد فاکتورهای بهینه هر مدل PLS بر مبنای حداقل خطای پیش‌بینی در اجرای الگوریتم اعتبارسنجی متقاطع (CV) برای هر آنالیت تعیین شد. تعداد فاکتورهای بهینه برای آموزش مدل خطی PLS برای رطوبت، پروتئین و چربی به ترتیب ۳، ۳ و ۱۰ تعیین شدند. سه مدل کالیبراسیون بهینه شده برای پیش‌بینی غلظت‌های رطوبت، پروتئین و چربی در ۷۰ نمونه گوشت مستقل به کار برده شدند. نتایج غلظت‌های پیش‌بینی شده به دو صورت در شکل (۳) نمایش داده شده‌اند. ستون اول در این شکل ارتباط بین غلظت‌های پیش‌بینی شده با غلظت‌های هر آنالیت در این ۷۰ نمونه گوشت را نشان می‌دهد. به منظور نمایش بهتر قدرت پیش‌بینی ماشین‌های طراحی شده، خطای مطلق در پیش‌بینی غلظت آنالیت‌ها در ستون دوم شکل (۳) نمایش داده شده است. در یک نگاه سریع کارآمدی کم‌تر مدل‌های PLS در مقایسه با مدل‌های ANN (شکل ۲) قابل مشاهده است. بررسی‌ها نشان داد که افزایش تعداد فاکتورهای PLS به خصوص در مورد مدل‌های مربوط به آنالیت‌های رطوبت و پروتئین که ۳ فاکتور در ساخت مدل‌های آنها شرکت کرده‌اند، بهبودی در نتایج آنالیز نمونه‌های پیش‌بینی ایجاد نمی‌کند.

برای مقایسه کمی دقیق‌تر مقادیر RMSEP برای رطوبت، پروتئین و چربی به ترتیب ۲/۳۱۳۲، ۲/۸۰۲۳ و ۰/۷۶۱۵ محاسبه شدند. هر سه مقدار در مقایسه با مقادیر متناظرشان در مدل‌های ANN به طور قابل توجهی بزرگ‌تر هستند که حاکی از ضعف مدل خطی در رسیدن به شرایط یکسان با مدل‌های غیرخطی است. برای واضح‌تر شدن مقایسه، مقادیر REP% هم برای سه مدل PLS محاسبه شدند. مقادیر درصد خطای پیش‌بینی برای رطوبت، پروتئین و چربی به ترتیب ۳/۶۷۲۵، ۱۵/۲۹۲۲ و ۴/۳۰۸۳ درصد محاسبه شدند. هرچند قدرت مدل خطی PLS در پیش‌بینی غلظت نمونه‌های مستقل به اندازه مدل‌های غیرخطی نیست، ولی نکته قابل توجه این است که ماشین یادگیری خطی PLS در شرایط اپتیم، توانسته به خطای قابل قبول در پیش‌بینی غلظت دست یابد. خطای نسبی پیش‌بینی در مورد آنالیت چربی برای مدل خطی PLS تنها حدود یک درصد از مدل غیرخطی ANN بیشتر است، که با این وجود هنوز خطای قابل قبول زیر ۵ درصد را دارد.



شکل ۳- نتایج پیش‌بینی غلظت رطوبت (ردیف اول)، پروتئین (ردیف دوم) و چربی (ردیف سوم) در ۷۰ نمونه گوشت با مدل‌های بهینه شده PLS. پلات‌های غلظت‌های پیش‌بینی شده نسبت به غلظت‌های واقعی (ستون اول) و میزان خطای مطلق پیش‌بینی غلظت نمونه‌ها (ستون دوم).

۷. نتیجه‌گیری

روش‌های یادگیری ماشین در مدل‌سازی ارتباط بین طیف‌های NIR ثبت شده از نمونه‌های گوشت و غلظت رطوبت، پروتئین و چربی بسیار توانا هستند. در مرحله یادگیری متغیرهای قابل تنظیم ماشین به کار برده شده، در موقعیت بهینه خود قرار داده می‌شوند. بافت پیچیده نمونه‌های گوشت باعث می‌شود که داده‌های طیفی اندازه‌گیری شده حامل رفتار غیرخطی باشند. این ویژگی نیاز به استفاده از ماشین‌های یادگیری غیرخطی را ایجاد می‌کند. RBF-ANN به عنوان یک هوش مصنوعی، در کالیبراسیون سه ویژگی مورد جستجو بسیار مطلوب عمل می‌کند و مدل‌های ساخته شده برای هر آنالیت با میزان دقت و صحت مناسب، قادر به پیش‌بینی در نمونه‌های مستقل هستند. مقایسه ارقام شایستگی مدل‌های خطی و غیرخطی مورد استفاده نشان می‌دهد که، ماشین یادگیری خطی PLS قادر است در مواردی به نتایج قابل قبول در آنالیز داده‌های غیرخطی برسد.

References

1. Dias CM, Nunes H, Melo T, Rosa H, Silva C & Borba A. Application of Near Infrared Reflectance (NIR) spectroscopy to predict the moisture, protein, and fat content of beef for gourmet hamburger preparation. *Livestock Science*. 2021; 254: 104772. <https://doi.org/10.1016/j.livsci.2021.104772>
2. Kamruzzaman M, ElMasry G, Sun D-W & Allen P. Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression. *Innovative Food Science & Emerging Technologies*. 2012; 16: 218-26. <https://doi.org/10.1016/j.ifset.2012.06.003>
3. Tøgersen G, Isaksson T, Nilsen B, Bakker E & Hildrum K. On-line NIR analysis of fat, water and protein in industrial scale ground meat batches. *Meat science*. 1999; 51(1): 97-102. [https://doi.org/10.1016/S0309-1740\(98\)00106-5](https://doi.org/10.1016/S0309-1740(98)00106-5)
4. Tøgersen G, Arnesen J, Nilsen B & Hildrum K. On-line prediction of chemical composition of semi-frozen ground beef by non-invasive NIR spectroscopy. *Meat science*. 2003; 63(4): 515-23. [https://doi.org/10.1016/S0309-1740\(02\)00113-4](https://doi.org/10.1016/S0309-1740(02)00113-4)
5. Massart DL, Vandeginste BG, Buydens L, Lewi P, Smeyers-Verbeke J & Jong SD. *Handbook of chemometrics and qualimetrics*. Elsevier Science Inc., 1998
6. Martens H & Naes T. *Multivariate calibration*. John Wiley & Sons; 1992.
7. Haaland DM & Thomas EV. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Analytical chemistry*. 1988; 60(11): 1193-202. <https://doi.org/10.1021/ac00162a020>
8. Russell SJ & Norvig P. *Artificial intelligence a modern approach*: London: 2010.
9. Hippe Z. Problems in the application of artificial intelligence in analytical chemistry. *Analytica Chimica Acta*. 1983;150: 11-21. [https://doi.org/10.1016/S0003-2670\(00\)85455-0](https://doi.org/10.1016/S0003-2670(00)85455-0)
10. Nayak J, Vakula K, Dinesh P, Naik B & Pelusi D. Intelligent food processing: Journey from artificial neural network to deep learning. *Computer Science Review*. 2020; 38: 100297. <https://doi.org/10.1016/j.cosrev.2020.100297>
11. Wentzell PD & Lohnes MT. Maximum likelihood principal component analysis with correlated measurement errors: theoretical and practical considerations. *Chemometrics and Intelligent Laboratory Systems*. 1999; 45(1-2): 65-85. [https://doi.org/10.1016/S0169-7439\(98\)00090-2](https://doi.org/10.1016/S0169-7439(98)00090-2)
12. Orr MJ. Regularization in the selection of radial basis function centers. *Neural computation*. 1995;7(3): 606-23
13. Borggaard C & Thodberg HH. Optimal minimal neural interpretation of spectra. *Analytical chemistry*. 1992; 64(5): 545-51. <https://doi.org/10.1021/ac00029a018>
14. Allegrini F & Olivieri AC. Sensitivity, prediction uncertainty, and detection limit for artificial neural network calibrations. *Analytical chemistry*. 2016; 88(15): 7807-12. <https://doi.org/10.1021/acs.analchem.6b01857>
15. Chiappini FA, Goicoechea HC & Olivieri AC. MVC1_GUI: A MATLAB graphical user

-
- interface for first-order multivariate calibration. An upgrade including artificial neural networks modelling. *Chemometrics and Intelligent Laboratory Systems*. 2020; 206.
<https://doi.org/10.1016/j.chemolab.2020.104162>
16. Vali Zade S. Practical approaches in detection of food fraud with multivariate calibration. *Human, Health and Halal Metrics*. 2023; 4(2): 25-37.
10.30502/JHHHM.2023.424076.1077