



The Use of Gradient Boost Regression Model to Modeling of Gas Sensors in dried Raisins Diagnosis of Sun-dried, Sulphurous and Acidic solution

Mohammad Ghoushchian¹ | Seyed Saeid Mohtasebi^{2✉} | Shahin Rafiee³

1. Department of Agricultural Machinery Engineering, Faculty of Agriculture, College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran, E-mail: mghoushchian@ut.ac.ir
2. Corresponding Author, Department of Agricultural Machinery Engineering, Faculty of Agriculture, College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran. E-mail: mohtaseb@ut.ac.ir
3. Department of Agricultural Machinery Engineering, Faculty of Agriculture, College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran, E-mail: shahinrafiee@ut.ac.ir

Article Info

Article type: Research Article

Article history:

Received: Jan. 6, 2024

Revised: Feb. 11, 2024

Accepted: Feb. 13, 2024

Published online: Spring 2024

Keywords:

*Data modeling,
Gas sensors,
Gradient Boost Regression,
machine learning,
Raisin harmful substances*

ABSTRACT

Machine learning modeling can help overcome some of the limitations of gas sensors, such as high operational conditions, drift errors, limited selectivity, the need for a large amount of labeled data, and cost and fabrication challenges. In this research, an electronic nose system was developed for the detection of sulfur dioxide and acetic acid. three treatments, including sunny, acetic, and sulfuric, were prepared in three repetitions, and each was exposed to olfactory sensors for 60 minutes to record the sensor responses to each treatment. Then, the data obtained from the sensor responses were examined by machine learning models to determine the modeling accuracy of each method. The results showed that the utilized Gradient Boost Regression model with a determination coefficient of 0.9972, root mean square error of 0.0209, mean absolute error of 0.0026, and relative root mean square error of 0.0209 was able to model the gas sensor responses well for the introduced treatments. Furthermore, by analyzing the results, the type and degree of correlation between the sensor responses to each other and over time were determined to evaluate their behavior prediction. Then, based on the conducted modeling, it was revealed that MQ9, MQ3, MQ5, and TGS2620 sensors, with determination coefficients of 0.8668, 0.8786, 0.9458, and 0.9074, and root mean square errors of 0.0163, 0.0168, 0.0083, and 0.0227, respectively, provided more accurate and predictable responses compared to MQ135, TGS822, TGS810, and MQ4 sensors.

Cite this article: Ghoushchian, M., Mohtasebi, S. S., Rafiee, Sh. (2024) The Use of Gradient Boost Regression Model to Modeling of Gas Sensors in Diagnosis of Sun-dried, Sulphurous and Acidic solution dried Raisins, *Iranian Journal of Biosystem Engineering*, 55 (1), 1-18. <http://doi.org/10.22059/ijbse.2024.370678.665534>

© The Author(s).

Publisher: The University of Tehran Press.

DOI: <http://doi.org/10.22059/ijbse.2024.370678.665534>



استفاده از مدل رگرسیون گرادیان افزایشی برای مدلسازی حسگرهای گازی در تشخیص کشمش آفتابی، گوگردی و تیزابی

محمد قوشچیان^۱ | سید سعید محتسبی^۲ | شاهین رفیعی^۳

۱. گروه مهندسی ماشین‌های کشاورزی، دانشکده کشاورزی، دانشکده‌گان کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران، رایانامه:

mghoushchian@ut.ac.ir

۲. نویسنده مسئول، گروه مهندسی ماشین‌های کشاورزی، دانشکده کشاورزی، دانشکده‌گان کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران،

رایانامه: mohtaseb@ut.ac.ir

۳. گروه مهندسی ماشین‌های کشاورزی، دانشکده کشاورزی، دانشکده‌گان کشاورزی و منابع طبیعی دانشگاه تهران، کرج، ایران، رایانامه:

shahinrafee@ut.ac.ir

اطلاعات مقاله

چکیده

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۲/۱۰/۱۶

تاریخ بازنگری: ۱۴۰۲/۱۱/۲۲

تاریخ پذیرش: ۱۴۰۲/۱۱/۲۴

تاریخ انتشار: بهار ۱۴۰۳

واژه‌های کلیدی:

حسگرهای گازی،

رگرسیون گرادیان افزایشی،

مدلسازی،

مواد مضر کشمش،

یادگیری ماشین

مدلسازی یادگیری ماشین می‌تواند به غلبه بر برخی از محدودیت‌های حسگرهای گازی، مانند شرایط عملیاتی سخت، خطاهای رانش، انتخاب محدود، نیاز به مقدار زیادی از داده‌های برجسب‌گذاری شده و چالش‌های هزینه و ساخت کمک کند. در این پژوهش یک سامانه بینی الکترونیک جهت تشخیص کشمش آفتابی، گوگردی و تیزابی ساخته شد. تیمارها شامل سه تیمار آفتابی، تیزابی و گوگردی هر کدام در سه تکرار آماده شدند و هر کدام ۶۰ دقیقه در معرض حسگرهای بویایی قرار گرفتند تا پاسخ حسگرها به هر کدام از تیمارها ثبت شود. سپس داده‌های بدست آمده از پاسخ حسگرها توسط مدل‌های یادگیری ماشین مورد بررسی قرار گرفتند تا دقت مدلسازی هر روش مشخص شده و مورد بررسی قرار گیرد. نتایج نشان داد مدل رگرسیون گرادیان افزایشی استفاده شده با ضریب تبیین ۰/۹۹۷۲، ریشه میانگین مربعات خطای ۰/۰۲۰۹، میانگین مطلق خطای ۰/۰۰۲۶ و ریشه میانگین مربعات خطای نسبی ۰/۰۲۰۹. برای داده‌های آزمون توانسته است پاسخ حسگرهای گازی را به خوبی نسبت به تیمارهای معرفی شده مدلسازی کند. همچنین با بررسی و تحلیل نتایج بدست آمده، نوع و میزان همبستگی بین پاسخ حسگرها نسبت به هم و نسبت به زمان مشخص شد تا در پیش‌بینی رفتار آنها مورد ارزیابی قرار بگیرد. سپس با مدلسازی انجام شده مشخص شد حسگرهای MQ9، MQ3، MQ5، MQ9، TGS2620، به ترتیب با ضرایب تبیین ۰/۸۶۶۸، ۰/۸۷۸۶، ۰/۹۴۵۸، ۰/۹۰۷۴ و ریشه میانگین مربعات خطای ۰/۰۱۶۳، ۰/۰۱۶۸، ۰/۰۰۸۳ و ۰/۰۲۲۷ پاسخ‌های دقیق‌تر و پیش‌بینی پذیرتری نسبت به حسگرهای MQ135، MQ822، TGS810 و MQ4 نشان دادند.

استاد: قوشچیان؛ محمد، محتسبی؛ سید سعید، رفیعی؛ شاهین، (۱۴۰۳) استفاده از مدل رگرسیون گرادیان افزایشی برای مدلسازی حسگرهای گازی در تشخیص کشمش

آفتابی، گوگردی و تیزابی، مجله مهندسی بیوسیستم ایران، ۵۵ (۱)، ۱۸-۱. <http://doi.org/10.22059/ijbse.2024.370678.665534>



© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <http://doi.org/10.22059/ijbse.2024.370678.665534>

مقدمه

مدل رگرسیون گرادیان افزایشی، یک مدل یادگیری ماشین است که برای حل مسائل رگرسیون استفاده می‌شود. این مدل یک الگوریتم یادگیری ماشینی مبتنی بر مجموعه است که از چندین مدل درخت تصمیمی ساده تشکیل شده است. هر مدل درخت تصمیمی، خطاهای مدل قبلی را پیش‌بینی می‌کند و با استفاده از این پیش‌بینی، بهبود عملکرد کلی مدل دسته‌بندی را ممکن می‌سازد (James et al., 2013). نتایج رگرسیون گرادیان افزایشی می‌تواند به صورت یک مدل پیش‌بینی دقیق باشد که قابلیت پیش‌بینی و تفسیر خروجی را دارد. برای تفسیر نتایج، باید به دقت معیارهای ارزیابی مانند ضریب تبیین، میانگین مطلق خطا، خطای مجذور میانگین و ریشه میانگین مربعات خطا را بررسی کرد. همچنین، مهم است که ویژگی‌های مهم در مدل پیدا شوند تا بتوان بهترین نتیجه را از آن استخراج کرد. برای این کار، می‌توان از روش‌هایی مانند استخراج ویژگی استفاده کرد که اجازه می‌دهد تا ویژگی‌های مهم در مدل را پیدا و بر اساس آن‌ها تصمیم‌گیری کرد. یکی از مزایای این روش این است که با استفاده از داده‌های آموزشی، مدلی با دقت بالا و قابلیت تفسیر بالا ساخته می‌شود. به طور خلاصه برخی از مزایای این روش به بدین شرح است: رگرسیون گرادیان افزایشی معمولاً دقت بالایی در پیش‌بینی دارد و می‌تواند به خوبی با داده‌های پیچیده و با نویز سازگاری داشته باشد. همچنین این روش به دلیل استفاده از مدل درختی تصمیم‌گیری، به راحتی قابلیت تفسیر دارد و می‌توان بررسی کرد که هر ویژگی در مدل چه وزنی دارد. رگرسیون گرادیان افزایشی دارای پارامترهای قابل تنظیم است که می‌توان با تنظیم آن‌ها، دقت و عملکرد مدل را بهبود بخشید. این روش با استفاده از روش‌هایی مانند رگرسیون با خطای مجذور میانگین، مقاومت خوبی در برابر داده‌های دارد و در نهایت این روش در مقابل برخی روش‌های یادگیری ماشینی دیگر، سرعت بالایی دارد و می‌تواند برای داده‌های بزرگ مناسب باشد (Chen & Guestrin, 2016).

اندازه‌گیری باقی‌مانده عناصری مانند دی‌اکسید گوگرد و ترکیبات آن که به دلیل در دسترس و ارزان بودن به طور گسترده‌ای در خشکبار به عنوان ماده نگه‌دارنده به کار گرفته می‌شود، یکی از نکاتی است که کشورهای واردکننده در هنگام خرید کشمش از ایران مورد بررسی قرار می‌دهند. لذا کشمش صادراتی می‌بایست از لحاظ میزان وجود این عناصر در محصول نهایی بررسی شود. دی‌اکسید گوگرد خاصیت آنتی‌اکسیدانی و ضد میکروبی دارد و برای جلوگیری از رشد میکروارگانیسم‌ها و همچنین برای حفظ رنگ و طعم در غذاهای مختلف به ویژه در میوه‌های خشک (زردآلو، آلو، سیب، کشمش و غیره) استفاده می‌شود. از طرف دیگر، از گوگرد و سولفات مس مستقیماً به عنوان سموم دفع آفات به خصوص در باغ‌های انگور استفاده می‌شود که منبع دیگری از کاربرد گوگرد در مواد غذایی است (Mischek & Krapfenbaue, 2012).

دی‌اکسید گوگرد و سولفیت‌ها، اجزای آلرژی‌زایی هستند که می‌تواند منجر به واکنش‌های آلرژیک در بیماران آسم و افراد دارای حساسیت شود (Soubra et al., 2007). علاوه بر این، این مواد افزودنی غذایی می‌توانند باعث واکنش‌های پوستی و آسیب دی ان ای شوند (Meng et al., 2005؛ Vally et al., 2009). بنابراین، سازمان‌های بررسی کننده ایمنی مواد غذایی حد قابل قبول دی‌اکسید گوگرد و سولفیت را در مواد غذایی در نظر گرفته‌اند. در نتیجه، ذکر نام و غلظت ترکیبات سولفیت مورد استفاده در فرمولاسیون مواد غذایی یا نوشیدنی بر روی برچسب مواد غذایی در بسیاری از کشورها یک الزام قانونی است (Guido., 2016).

در اتحادیه اروپا (EU)، حد مجاز باقی مانده سولفیت در محصولات مختلف متفاوت است. برای زردآلو خشک، هلو، انگور، آلو خشک و انجیر؛ باقی مانده سولفیت باید کمتر از ۲۰۰۰ میلی گرم در کیلوگرم باشد در حالی که حداکثر میزان آن در سیب خشک و گلابی ۶۰۰ میلی گرم در کیلوگرم و در سایر میوه‌های خشک و آجیل کمتر از ۵۰۰ میلی گرم در کیلوگرم در نظر گرفته شد (EFSA, 2016). از این رو، Codex (۱۹۹۵) محدودیت‌های مختلفی برای باقی مانده دی‌اکسید گوگرد پیشنهاد کرد، از جمله ۱۵۰۰ میلی گرم در کیلوگرم در کشمش و ۲۰۰۰ میلی گرم در کیلوگرم در زردآلو خشک.

طبق استاندارد ملی ایران، حداکثر سطح قابل قبول دی‌اکسید گوگرد برای انواع مختلف انگور متفاوت است: ۷۰۰ میلی گرم در کیلوگرم در کشمش سلطانا (ISIRI, 2012)، ۱۵۰۰ میلی گرم در کیلوگرم در کشمش طلایی (angora; ISIRI, 2012)، ۲۰۰۰ میلی گرم در کیلوگرم در پوست میوه هلو (ISIRI, 2018) و ۲۰۰۰ میلی گرم در کیلوگرم در زردآلوهای خشک (ISIRI, 2014)، و همچنین زردآلوهای نیمه خشک (ISIRI, 2015).

روش‌های قدیمی برای بررسی میزان باقی‌مانده دی‌اکسید گوگرد در کشمش شامل استاندارد ملی شماره ۵۶۹ به روش یدومتری و روش‌های تجزیه‌ای، کروماتوگرافی یونی، جذب مولکولی، طیف‌سنجی جذب اتمی با منبع خطی و منبع پیوسته، آمپرومتری و ولتامتری و فلورسانس ماوراء بنفش است (Mischek & KrapfenbauerCermak., 2012).

بینی الکترونیکی یک فناوری است که برای تقلید از سیستم بویایی انسان طراحی شده است (Borowik et al., 2020). امروزه، بینی‌های الکترونیکی ابزاری هستند که مبتنی بر تعامل حسگرهای نیمه انتخابی با ترکیبات فرار هستند. انواع مختلفی از حسگرها در سامانه‌های بینی الکترونیکی به کار رفته‌اند و معروف‌ترین آنها نیمه هادی اکسید فلزی^۱، پلیمر رسانا^۲ و حسگرهای کریستالی پیزوالکتریک^۳ هستند (Borowik et al., 2020).

پیشینه پژوهش

رحمن پور و همکاران (۱۳۹۹) در پژوهشی، یک سامانه چند حسگری ماشین بویایی بر پایه حسگرهای نیمه هادی اکسید فلزی به منظور ایجاد روشی سریع، حساس و قابل اطمینان جهت تشخیص تقلب در سرکه انگور را ارزیابی کردند. سامانه ماشین بویایی به طور موفقیت آمیزی توانست تقلب سرکه سیب را در سرکه انگور با درصدهای مختلف تقلب توسط روشهای تحلیل مؤلفه‌های اصلی (PCA) و شبکه عصبی مصنوعی (ANN) تشخیص و طبقه بندی کند. نتایج بدست آمده از تحلیل PCA، حسگرهای MQ136 و MQ9 را به ترتیب به عنوان بهترین حسگر برای تشخیص تقلب سرکه سیب معرفی کرد. همچنین نتایج نشان داد که مؤلفه‌های اصلی PC1 و PC2، برای نمونه‌های مخلوط سرکه سیب در مجموع ۹۲٪ واریانس داده‌ها را مورد پوشش قرار داد. همچنین ماتریس اغتشاش بهترین ساختار ANN، برای تقلب سرکه سیب تفکیک ۳/۹۳٪ را نشان داد.

ملازاده و قهرمانی (۱۳۹۳) امکان استفاده از دستگاهی جهت دسته بندی و جداسازی کشمش‌های مطلوب با استفاده از سیستم پردازش تصویر را مورد بررسی قرار دادند. این سیستم از نوار نقاله، جعبه نور، کنترلر، سیستم پردازش تصویر و واحدهای جداسازی تشکیل شده است. در این آزمایش به منظور انجام پردازش تصویر و استخراج ویژگی‌های بسیار مفید از تصاویر گرفته شده توسط دوربین‌های نصب شده، از الگوریتم‌های پردازش تصویر استفاده شد. این الگوریتم شامل تقسیم بندی پس زمینه، انتخاب کشمش‌های مطلوب و استخراج آن‌ها می‌باشد. که در ابتدا پس زمینه تصاویر را حذف کرده و سپس انواع کشمش را مشخص کرده و در نهایت با محاسبه مرکز ثقل هر یک از کشمش‌ها برای جداسازی به صورت خودکار و رد کشمش‌های نامرغوب کار پردازش خود را به پایان می‌رساند.

در پژوهشی حسگر TGS^{۸۲۲} با استفاده از یک حسگر شبکه‌عصبی مصنوعی مدلسازی شد. این مدل با در نظر گرفتن غیرخطی بودن پاسخ آن، وابستگی به دما و رطوبت نسبی در نقطه اندازه‌گیری علاوه بر وابستگی به طبیعت گاز، رفتار حسگر گازی را در یک محیط دینامیکی به دقت بازتولید کرد. مدل ANN پیشنهادی به‌عنوان یک مؤلفه در کتابخانه شبیه ساز SPICE پیاده‌سازی شد و این مدل آزمایش و اعتبارسنجی شد. در نتایج این پژوهش تصریح شده است این روش را می‌توان با حسگرهای دیگر نیز استفاده کرد (Baha & Dibi, 2010).

در پژوهش دیگری عنوان شده است که یادگیری ماشین می‌تواند به طور موثر با محدودیت‌های زمان و هزینه آزمایش‌ها مقابله کند. در این پژوهش یادگیری ماشین برای مدل‌سازی کمیت و کیفیت محصولات (بیوچار، روغن زیستی و سینگاس) که در پیرولیز مایکروویو زیست توده تکامل می‌یابند، استفاده شده است. یک مجموعه داده فراگیر شامل انواع مختلف زیست توده، جذب مایکروویو و شرایط واکنش از پیشینه تحقیق انتخاب شده و تحت داده کاوی قرار گرفته و سه مدل یادگیری ماشین (رگرسیون بردار پشتیبان، رگرسیون جنگل تصادفی و رگرسیون گرادیان افزایشی) برای مدل‌سازی فرایند بر اساس ۱۴ توصیف کننده استفاده شد. در نتایج این پژوهش ذکر شده مدل رگرسیون گرادیان افزایشی عملکرد پیش‌بینی بهتری ($R_2 > 0.822$ و $RMSE < 12.38$ و $RRMSE < 0.765$) را نسبت به سایر مدل‌ها نشان داده است (Yang et al., 2022).

در پژوهش دیگری یک رویکرد مدل سازی ANN برای یک حسگر گازی صنعتی به نام MQ-9 پیشنهاد شده است. این مدل، غیرخطی بودن، حساسیت متقاطع، دما و رطوبت را در نظر می‌گیرد تا رفتار حسگر گازی را به درستی نشان دهد (Kouda et al., 2022). در پژوهشی، تشخیص غیرمخرب پیچ امین‌الدوله^۵ تحت تیمار دوددهی گوگرد بر اساس تصویربرداری فراطیفی را مورد بررسی قرار دادند. نتایج کلی نشان داد که همه مدل‌های امتحان شده در این پژوهش می‌تواند به شناسایی سریع و غیر مخرب فلوس لونیسرا دوددهی

1- Metal oxide semiconductor (MOS)
2- Conducting polymer (CP)
3- Piezoelectric crystal sensors
4- Taguchi Gas Sensors
5. Flos Lonicerae

شده با گوگرد دست یابند. مدل مبتنی بر CARS عملکرد بهینه را به دست آورد و مدل CARS - LS-SVM با ضریب همبستگی ۰/۹۱۱ و ریشه میانگین مربعات خطا معادل ۰/۳۳۵ بهترین اثر طبقه‌بندی را برای دسته پیش‌بینی داشت (Liu et al., 2018). در پژوهش دیگری یک مدل شبکه عصبی مصنوعی (ANN) برای یک حسگر گازی اکسید فلزی (MOX) مورد استفاده در تشخیص گاز، به ویژه برای گاز مایع ارائه شده است. در این پژوهش ثابت شد که روش ANN یک ابزار قدرتمند برای مدل‌سازی حسگر گازی LP است (Lamamra & Djamil, 2016). با توجه به اینکه حسگرهای گازی دارای محدودیت‌هایی هستند که با استفاده از یادگیری ماشین می‌توان بر آنها غلبه کرد و با بررسی‌های انجام شده در تحقیقات مشابه، مشخص شد تاکنون پژوهشی در خصوص مدل‌سازی پاسخ حسگرهای گازی اکسید فلزی به روش رگرسیون گرادیان افزایشی انجام نشده است. بنابر این در این پژوهش پاسخ حسگرهای انتخاب شده در آزمون تشخیص مواد مضر کشمش (تیزاب و دی‌اکسید گوگرد) با روش‌های رگرسیون و رگرسیون گرادیان افزایشی مدل‌سازی شدند و دقت مدل‌سازی آنها مورد تجزیه و تحلیل قرار گرفت.

روش‌شناسی پژوهش

برای این پژوهش، یک سامانه بینی الکترونیک شامل محفظه حسگری، آرایه حسگرها، سامانه داده‌برداری و دیگر ملزومات سامانه ساخته شد. تیمارها شامل سه تیمار آفتابی، تیزابی و گوگردی هر کدام در سه تکرار آماده شدند و هر کدام ۶۰ دقیقه در معرض حسگرهای بویایی قرار گرفتند تا پاسخ حسگرها به هر کدام از تیمارها ثبت شود. سپس داده‌های به دست آمده از پاسخ حسگرها توسط مدل‌های یادگیری ماشین مورد بررسی قرار گرفتند تا دقت مدل‌سازی هر روش مشخص شود و قابل مقایسه باشد. پارامترهای ارزیابی مدل در مدل‌سازی‌های انجام شده بررسی شد و تفسیر هر کدام مورد بحث و بررسی قرار گرفت که در ادامه به تفصیل به هر یک اشاره شده است.

جهت آماده‌سازی نمونه‌ها، از نمونه کشمش بدون دانه تاکستان قزوین، سه تیمار کشمش آفتابی، تیمار تیزابی و تیمار گوگردی و از هر تیمار سه تکرار تهیه شد و در آزمایش‌ها مورد ارزیابی قرار گرفت. مقدار هر کدام از نمونه‌ها مطابق کمترین مقدار استفاده شده در آزمون استاندارد، میزان ۵۰ گرم توزین شد و تمامی آزمایش‌ها در دما و رطوبت ثابت انجام شدند.

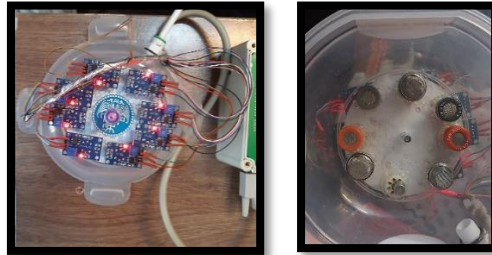
طراحی یک سامانه بینی الکترونیک جامع که قادر به تشخیص هر نوع بو باشد؛ غیرواقع‌بینانه است. بنابراین، اطلاعات، پردازش و ابزار دقیق باید برای هر کاربرد طراحی شود (ازگلی و همکاران، ۱۴۰۲). سامانه بینی الکترونیک شامل قسمت‌های: سامانه داده‌برداری، حسگرها، محفظه حسگرها، محفظه نمونه‌گیری، منبع تغذیه، اتصالات و لوازم جانبی، پمپ هوا و فیلتر هوا است. در حالت کلی می‌توان این سامانه را متشکل از ۳ بخش اصلی در نظر گرفت: ۱- سامانه نمونه‌گیری ۲- سامانه ردیابی ۳- سامانه پردازش داده (Borowik et al., 2020).

انتخاب مجموعه حسگرهای سامانه از اهمیت زیادی برخوردار است، به طوری که با انتخاب یک مجموعه حسگرهای مناسب می‌توان طیف وسیعی از بوها را شناسایی کرد (Ghasemi-Varnamkhasti et al., 2019). حسگرهای مورد استفاده بر روی یک صفحه الکترونیک مداربندی می‌شوند. مدار بندی این حسگرها بر اساس ساختار و پیکربندی منتشر شده توسط شرکت‌های سازنده انجام می‌شود. همچنین طبق برگه اطلاعاتی حسگرهای مورد استفاده، لازم است تا حسگرها پیش از شروع آزمایشات برای مدت زمان مشخصی گرم شوند. چون حسگرهای نوع TGS بیشترین مدت زمان یعنی هفت روز را به خود اختصاص می‌دهند، تمامی حسگرها به مدت یک هفته پیش از شروع هرگونه آزمایشی گرم می‌شوند. در جدول ۱ حسگرهای مورد استفاده در سامانه بینی الکترونیکی ساخته شده برای این پژوهش و کاربردهای آنها ذکر شده است.

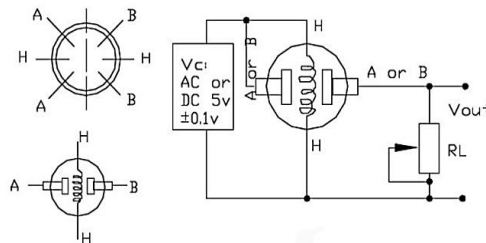
جدول ۱. حسگرهای مورد استفاده در سامانه بینی الکترونیکی و کاربردهای آنها

نوع حسگر	نشان تجاری	گازهای تشخیص دهنده
S 1	MQ-9	کربن مونواکسید-گازهای احتراق‌پذیر
S 2	MQ-3	الکل
S 3	MQ-5	ال بی جی و گازهای طبیعی
S 4	TGS-2620	الکل - بخار حلال‌های آلی
S 5	MQ-135	الکل، آمونیاک، دی‌اکسید کربن و بنزن
S 6	TGS-822	بخار حلال‌های آلی
S 7	TGS-813	بوتان، پروپان و متان
S 8	MQ-4	بوتان، پروپان، متان

باتوجه به اینکه حسگرها عملکرد اختصاصی برای گازهای هدف ندارند (Maho et al., 2022b)؛ در نهایت پس از بررسی ترکیبات نمونه مورد نظر نیز باید با آزمون و خطا و انتخاب آرایه‌های حسگری مختلف، آرایه حسگری که بهترین پاسخ را به تیمارهای انتخاب شده دارد مشخص کرد. بنابر این در این پژوهش نیز با آزمودن آرایه‌های حسگری مختلف، بهترین آرایه از حسگرها انتخاب شد. در شکل ۱ تصویر آرایه حسگری استفاده شده در سامانه بینی الکترونیکی ساخته شده برای این پژوهش و در شکل ۲ نحوه مداربندی این حسگرها آمده است.



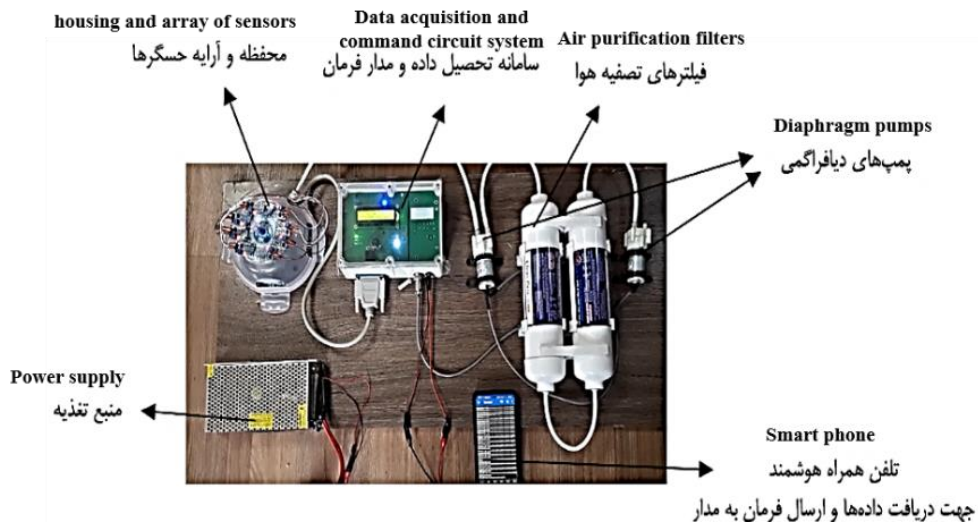
شکل ۱. آرایه حسگری مورد استفاده در سامانه بینی الکترونیکی



شکل ۲. مدار بندی یکی از حسگرهای مورد استفاده در سامانه بینی الکترونیکی و نحوه قرارگیری پایه های آن

فرآیند داده‌برداری

جهت انجام آزمایش‌های مربوط به تفکیک و طبقه‌بندی انواع نمونه، نمونه‌ها در تیمارهای گوگردی، تیزابی و آفتابی هر کدام با سه تکرار آماده شده و هر آزمایش نیز در سه تکرار مشخص انجام شد. سامانه ماشین بویایی شامل محفظه بویایی، آرایه حسگرها شامل ۸ حسگر گازی، سامانه داده‌برداری شامل برد اختصاصی طراحی شده جهت داده‌برداری، انتقال داده‌ها به رایانه از طریق کابل و انتقال به تلفن همراه از طریق بلوتوث و اپلیکیشن اندروید و پمپ و فیلترهای تصفیه هوا ساخته شد که در شکل ۳ به طور کامل تشریح شده است.



شکل ۳. تصویر سامانه ماشین بویایی ساخته شده جهت دریافت داده‌ها و ارسال فرمان به مدار

جهت دریافت و پردازش داده‌های مربوط به حسگرها و همچنین اتصال مدار فرمان به رایانه، از یک سامانه داده‌برداری با برد

اختصاصی استفاده شد. جهت دریافت داده‌های حسگرها از ورودی‌های آنالوگ و همچنین جهت کنترل مدار فرمان از خروجی‌های دیجیتال این سامانه استفاده شد. این سامانه از طریق کابل به رایانه و از طریق بلوتوث به تلفن هوشمند متصل می‌شود و ارتباط دوسویه جهت ارسال اطلاعات و دریافت تنظیمات و برنامه‌ها را دارد و جهت داده‌برداری نیز از هردو رابط می‌توان استفاده کرد ولی رابط تلفن همراه بیشتر برای کاربر نهایی و مصرف‌کننده این سامانه طراحی شده است تا بدون استفاده از لپ‌تاپ علاوه بر نمایشگر خود سامانه از تلفن همراه هوشمند نیز اطلاعات تکمیلی را بتواند مشاهده کند. به‌منظور تأمین توان الکتریکی مورد نیاز نیز از یک منبع تغذیه دوازده ولت استفاده می‌شود. منبع تغذیه مورد استفاده، ولتاژ ورودی متناوب را به ولتاژ مستقیم تبدیل می‌کند. از تبدیل منبع دوازده ولت به پنج ولت جهت راه‌اندازی مدار فرمان و تأمین توان مورد نیاز حسگرها استفاده می‌شود و منبع دوازده ولت جهت راه‌اندازی پمپ به کار می‌رود.

همچنین جهت رساندن پاسخ حسگرها به سطح مبناء و همچنین تمیز کردن محفظه حسگرها از بوی نمونه، از هوای تصفیه شده استفاده می‌شود. برای این منظور از دو کپسول کربن فعال جهت تصفیه هوا و دو پمپ دیافراگرامی ۱۲ ولتی جهت انتقال هوا استفاده شد.

پیش‌پردازش داده‌های سامانه بینی الکترونیکی

برای استخراج ویژگی‌ها از پاسخ حسگرهای سامانه بینی الکترونیکی توسعه‌یافته، حالت مبنای پاسخ حسگرها (پایان نقطه تزریق گاز تصفیه شده) در نظر گرفته می‌شود. برای استخراج ویژگی‌ها از داده‌های به دست آمده از حسگرهای سامانه بینی الکترونیکی مورد استفاده، زمان‌بندی در نظر گرفته می‌شود. سپس از یکی از روش‌های معمول پیش‌پردازش داده‌های بینی الکترونیکی استفاده می‌شود (چهرمی و همکاران، ۱۳۹۸).

روش‌های متداول پیش‌پردازش داده‌های سامانه بینی الکترونیکی شامل روش کسری، تفاضلی و نسبی است که در این تحقیق روش کسری جهت پیش‌پردازش داده‌های آرایه حسگری بینی الکترونیکی و نرمال‌سازی داده‌های هر حسگر استفاده شد. در این روش، خط مبنا از پاسخ حسگر کم می‌شود و سپس تقسیم صورت می‌گیرد. اندازه‌گیری‌های کسری نه تنها بدون بعد بوده بلکه نرمال شده هم هستند. زیرا پاسخ حاصل، یک تغییر پیش واحد نسبت به خط مبنا است که برای حسگرهایی که ذاتا سطح پاسخ بالا و پایینی دارند مفید است. پیش‌پردازش سیگنال‌های به دست آمده از آرایه حسگری به‌منظور افزایش کارایی اطلاعات موجود و بهینه‌سازی خروجی آرایه، پیش از انجام تحلیل آن‌ها با روش‌های تشخیص الگو است (Tohidi et al., 2018b).

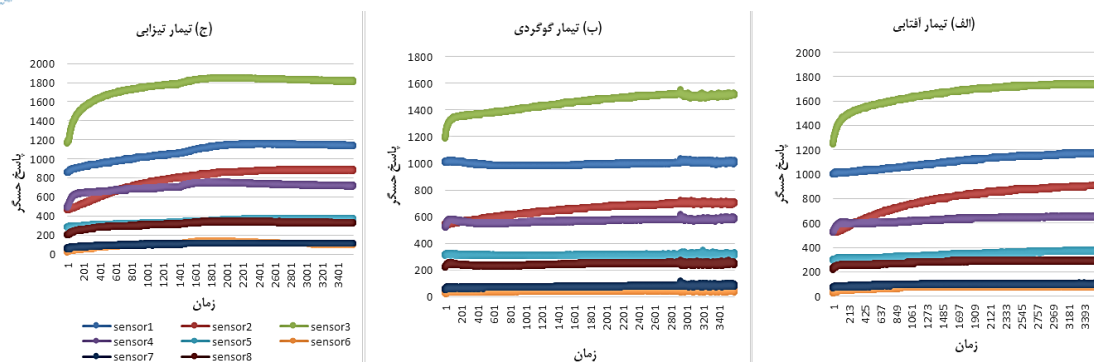
مدل‌سازی داده‌ها

برای ارزیابی کیفیت پیش‌بینی مدل، می‌توان از معیارهایی مانند ضریب تبیین، ریشه میانگین مربعات خطا، میانگین مطلق خطا و ریشه میانگین مربعات خطای نسبی استفاده کرد که در مدل رگرسیون گرادیان افزایشی نیز قابل‌استفاده هستند. همچنین، برای بررسی تفاوت میان مقادیر واقعی و پیش‌بینی‌شده، می‌توان از آزمون فرضیه و آزمون تی استفاده کرد.

برای داده‌های به دست آمده از پاسخ حسگرها پس از نرمال‌سازی به روش کسری، ماتریس همبستگی پاسخ حسگرها نسبت به هم و نسبت به زمان برای تیمارها و تکرارهای مختلف استخراج شد. سپس نمودار مقایسه مقادیر پیش‌بینی شده در مقابل مقادیر واقعی برای مدل‌سازی به روش رگرسیون و رگرسیون گرادیان افزایشی ترسیم شد و دقت پیش‌بینی هریک با مؤلفه‌های کیفیت پیش‌بینی مورد تجزیه و تحلیل قرار گرفت و پارامترهای مدل‌سازی برای مدل بهینه هرکدام گزارش شد. سپس آنالیز واریانس مدل رگرسیون گرادیان افزایشی برای هریک از مؤلفه‌های کیفیت پیش‌بینی مدل به تفکیک تیمارها، حسگرها و ترکیب حسگرها با تیمارها هم برای داده‌های آزمایش و هم برای داده‌های آزمون انجام شد که از تفسیر هرکدام نکات مختلف در خصوص مدل‌سازی انجام شده استخراج شد و در بخش بحث و نتایج مورد تجزیه و تحلیل قرار گرفت.

یافته‌های پژوهش و بحث

در شکل ۴، میانگین داده‌ها شامل اندازه‌گیری‌های ۸ حسگر گازی نیمه‌هادی اکسید فلزی در سه تکرار مختلف برای سه تیمار تیزابی، گوگردی و آفتابی برای یک نمونه کشمش رسم شده است. این نوع حسگر به‌عنوان یک حسگر گازی، برای تشخیص تغییرات در محیط با استفاده از تغییرات مقاومت الکتریکی در مقابل گازها عمل می‌کند. باتوجه‌به پاسخ هر یک از این حسگرها نسبت به هر نمونه مورد آزمایش، یک خط اثر اختصاصی به دست می‌آید که رفتار آن حسگر نسبت به نمونه آزمایش شده را نشان می‌دهد و می‌توان با مدل‌سازی آن و بررسی جمیع حسگرها، به اطلاعات ارزشمندی از نمونه دست یافت.



شکل ۴. نمودار پاسخ حسگرهای گازی نسبت به زمان برای تیمارهای مختلف

جداول ۲، ۳ و ۴ ماتریس همبستگی بین هشت حسگر مختلف برای سه تیمار تیزابی، گوگردی و افتابی را نشان می‌دهند که با استفاده از روش تحلیل همبستگی برای داده‌هایی که در طی زمان جمع‌آوری شده‌اند، تولید شده است. هر خانه در این جدول، مقدار ضریب همبستگی بین دو حسگر را نشان می‌دهد. همبستگی بیانگر رابطه مستقیم یا معکوس بین دو متغیر است. در اینجا، همبستگی صفر بین دو حسگر بیانگر عدم وجود هرگونه رابطه بین آن دو حسگر است. همبستگی ۱ بین دو حسگر نیز بیانگر رابطه مستقیم و معکوس کامل بین دو حسگر است. همبستگی مثبت بین دو حسگر نیز بیانگر رابطه مستقیم بین آن دو حسگر است، به‌عنوان مثال همبستگی بین حسگر ۱ و حسگر ۲ برابر با ۰/۹۷۶ است که نشان می‌دهد دو حسگر با هم مرتبط هستند و تغییرات مشابهی در آن‌ها رخ می‌دهد. باتوجه به مقادیر ماتریس همبستگی، می‌توان نتیجه گرفت که بین بعضی از حسگرها رابطه مثبت و قوی وجود دارد. به‌عنوان مثال، در تیمار تیزابی بین حسگر ۱ و حسگر ۲ رابطه مثبت و قوی وجود دارد. همچنین، بین حسگر ۴ و حسگر ۵ نیز رابطه مثبت و قوی با مقدار ضریب همبستگی ۰/۹۶۶ وجود دارد. از طرف دیگر، بین بعضی حسگرها رابطه منفی وجود دارد. باتوجه به این نتایج، می‌توان در طراحی سامانه‌های کنترل و پیش‌بینی فرایندها از این اطلاعات استفاده کرد. به‌عنوان مثال، باتوجه به رابطه مثبت بین حسگر ۱ و حسگر ۲، می‌توان پیش‌بینی کرد که در صورت افزایش مقدار حسگر ۱، مقدار حسگر ۲ نیز به تبع آن افزایش خواهد یافت.

جدول ۲. ماتریس همبستگی حسگرهای گازی در تیمار تیزابی

	زمان	حسگر ۱	حسگر ۲	حسگر ۳	حسگر ۴	حسگر ۵	حسگر ۶	حسگر ۷	حسگر ۸
زمان	۱	۰/۹۱۵۸۷۶	۰/۸۹۱۵۶۱	۰/۷۲۷۳۷۲	۰/۷۰۹۸۸۳	۰/۹۵۲۵۲۵	۰/۶۰۵۰۰۷	۰/۷۷۷۸۷۶	۰/۷۹۰۱۴
حسگر ۱	۰/۹۱۵۸۷۶	۱	۰/۹۷۶۰۵۶	۰/۸۸۴۰۷۹	۰/۹۰۷۰۷۹	۰/۹۷۲۱۷۷	۰/۸۴۳۳۴۵	۰/۹۰۲۶۲۱	۰/۹۴۱۰۱۴
حسگر ۲	۰/۸۹۱۵۶۱	۰/۹۷۶۰۵۶	۱	۰/۹۳۲۰۷۱	۰/۹۰۷۳۹۹	۰/۹۴۱۶	۰/۸۷۸۷۰۶	۰/۹۳۵۰۷۴	۰/۹۵۶۸۲۶
حسگر ۳	۰/۷۲۷۳۷۲	۰/۸۸۴۰۷۹	۰/۹۳۲۰۷۱	۱	۰/۹۵۰۳۳۶	۰/۸۱۶۴۱۸	۰/۹۲۹۴۱۲	۰/۹۳۳۲۹۸	۰/۹۶۴۳۲۴
حسگر ۴	۰/۷۰۹۸۸۳	۰/۹۰۷۰۷۹	۰/۹۰۷۳۹۹	۰/۹۵۰۳۳۶	۱	۰/۸۳۲۴۸	۰/۹۴۵۷۹۳	۰/۹۱۹۸۶۴	۰/۹۶۵۷۰۵
حسگر ۵	۰/۹۵۲۵۲۵	۰/۹۷۲۱۷۷	۰/۹۴۱۶	۰/۸۱۶۴۱۸	۰/۸۳۲۴۸	۱	۰/۷۵۱۵۶۹	۰/۸۷۲۹۷۷	۰/۸۹۱۵۴۴
حسگر ۶	۰/۶۰۵۰۰۷	۰/۸۴۳۳۴۵	۰/۸۷۸۷۰۶	۰/۹۲۹۴۱۲	۰/۹۴۵۷۹۳	۰/۷۵۱۵۶۹	۱	۰/۹۱۳۶۷۶	۰/۹۴۱۱۹۱
حسگر ۷	۰/۷۷۷۸۷۶	۰/۹۰۲۶۲۱	۰/۹۳۵۰۷۴	۰/۹۳۳۲۹۸	۰/۹۱۹۸۶۴	۰/۸۷۲۹۷۷	۰/۹۱۳۶۷۶	۱	۰/۹۵۷۳۵۵
حسگر ۸	۰/۷۹۰۱۴	۰/۹۴۱۰۱۴	۰/۹۵۶۸۲۶	۰/۹۶۴۳۲۴	۰/۹۶۵۷۰۵	۰/۸۹۱۵۴۴	۰/۹۴۱۱۹۱	۰/۹۵۷۳۵۵	۱

جدول ۳. ماتریس همبستگی حسگرهای گازی در تیمار گوگردی

	زمان	حسگر ۱	حسگر ۲	حسگر ۳	حسگر ۴	حسگر ۵	حسگر ۶	حسگر ۷	حسگر ۸
زمان	۱	۰/۴۲۷۰۰۱	۰/۹۵۲۹۲۱	۰/۹۳۶۵۳۹	۰/۷۶۷۱۲۳	۰/۳۹۶۳۲	۰/۷۵۲۰۸۲	۰/۶۹۰۴۵۱	۰/۶۲۸۹۷۳
حسگر ۱	۰/۴۲۷۰۰۱	۱	۰/۲۶۱۱۵۸	۰/۲۵۶۱۸	۰/۶۱۱۴۱۲	۰/۶۰۳۳۵۹	۰/۲۳۸۱۷۷۵	۰/۳۰۸۰۵۶	۰/۴۷۹۵۸۷
حسگر ۲	۰/۹۵۲۹۲۱	۰/۲۶۱۱۵۸	۱	۰/۹۸۰۵۷۲	۰/۷۲۴۹۰۴	۰/۲۹۵۴۳۴	۰/۷۸۵۹۰۹	۰/۶۹۶۳۸۴	۰/۶۳۹۱۵۷
حسگر ۳	۰/۹۳۶۵۳۹	۰/۲۵۶۱۸	۰/۹۸۰۵۷۲	۱	۰/۷۵۶۷۵۸	۰/۲۷۷۴۴۵	۰/۷۷۲۷۰۳	۰/۶۸۹۲۵۶	۰/۶۶۰۱۱۳
حسگر ۴	۰/۷۶۷۱۲۳	۰/۶۱۱۴۱۲	۰/۷۲۴۹۰۴	۰/۷۵۶۷۵۸	۱	۰/۵۲۰۰۱۸	۰/۶۴۹۴۴۶	۰/۶۲۲۲۳۳	۰/۷۷۳۵۸
حسگر ۵	۰/۳۹۶۳۲	۰/۶۰۳۳۵۹	۰/۲۹۵۴۳۴	۰/۲۷۷۴۴۵	۰/۵۲۰۰۱۸	۱	۰/۴۰۶۷	۰/۵۱۱۵۵۸	۰/۵۰۰۸۶۹
حسگر ۶	۰/۷۵۲۰۸۲	۰/۲۳۸۱۷۷۵	۰/۷۸۵۹۰۹	۰/۷۷۲۷۰۳	۰/۶۴۹۴۴۶	۰/۴۰۶۷	۱	۰/۷۰۲۵۴۱	۰/۶۱۷۸۶۸
حسگر ۷	۰/۶۹۰۴۵۱	۰/۳۰۸۰۵۶	۰/۶۹۶۳۸۴	۰/۶۸۹۲۵۶	۰/۶۲۲۲۳۳	۰/۵۱۱۵۵۸	۰/۷۰۲۵۴۱	۱	۰/۶۵۶۹۴۳
حسگر ۸	۰/۶۲۸۹۷۳	۰/۴۷۹۵۸۷	۰/۶۳۹۱۵۷	۰/۶۶۰۱۱۳	۰/۷۷۳۵۸	۰/۵۰۰۸۶۹	۰/۶۱۷۸۶۸	۰/۶۵۶۹۴۳	۱

جدول ۴. ماتریس همبستگی حسگرهای گازی در تیمار آفتابی

حسگر ۸	حسگر ۷	حسگر ۶	حسگر ۵	حسگر ۴	حسگر ۳	حسگر ۲	حسگر ۱	زمان
۰/۸۹۲۲۷۵	۰/۷۵۸۶۹۳	۰/۷۹۳۶۱۳	۰/۹۷۲۸۹	۰/۰۷۹۲	۰/۸۹۱۹۱۲	۰/۹۳۶۵۰۱	۰/۹۸۴۶۸۹	۱
۰/۹۳۳۶۰۱	۰/۷۹۷۳۳۹	۰/۸۴۳۴۳۸	۰/۹۸۱۹۷۴	۰/۹۴۹۲۸۲	۰/۹۲۴۸۱۶	۰/۹۶۵۹۸۰	۰/۹۸۴۶۸۹	حسگر ۱
۰/۹۴۱۹۵۵	۰/۸۲۶۶۱۹	۰/۹۱۲۵۹۳	۰/۹۴۱۴۷۸	۰/۹۴۵۰۶۳۶	۰/۹۷۸۱۷۷۸	۱	۰/۹۳۶۵۰۱	حسگر ۲
۰/۹۳۴۶۹۳	۰/۸۱۸۱۴۱	۰/۹۲۳۰۸۱	۰/۹۰۱۸۹۰	۰/۹۴۶۳۵۲	۱	۰/۹۷۸۱۷۷۸	۰/۹۲۴۸۱۶	حسگر ۳
۰/۹۶۰۵۷۶	۰/۸۳۷۰۰۹	۰/۸۱۷۷۶۴	۰/۹۴۰۵۳۹	۱	۰/۹۴۶۳۵۲	۰/۹۴۵۰۶۳۶	۰/۹۴۹۲۸۲	حسگر ۴
۰/۹۳۵۰۹۹	۰/۸۲۲۲۰۵	۰/۸۴۶۹۹۲	۱	۰/۹۴۰۵۳۹	۰/۹۰۱۸۹۰	۰/۹۴۵۰۶۳۶	۰/۹۸۱۹۷۴	حسگر ۵
۰/۹۱۰۱۱۵	۰/۹۰۹۱۵۶	۱	۰/۸۴۶۹۹۲	۰/۸۸۷۷۶۴	۰/۹۲۳۰۸۱	۰/۹۱۲۵۹۳	۰/۸۴۳۴۳۸	حسگر ۶
۰/۸۷۶۴۷۰	۱	۰/۹۰۹۱۵۶	۰/۸۲۲۲۰۵	۰/۸۳۷۰۰۹	۰/۸۱۸۱۴۱	۰/۸۲۶۶۱۹	۰/۷۹۷۳۳۹	حسگر ۷
۱	۰/۸۷۶۴۷۰	۰/۹۱۰۱۱۵	۰/۹۳۵۰۹۹	۰/۹۶۰۵۷۶	۰/۹۳۴۶۹۳	۰/۹۴۱۹۵۵	۰/۹۳۳۶۰۱	حسگر ۸

الگوریتم‌های یادگیری ماشینی

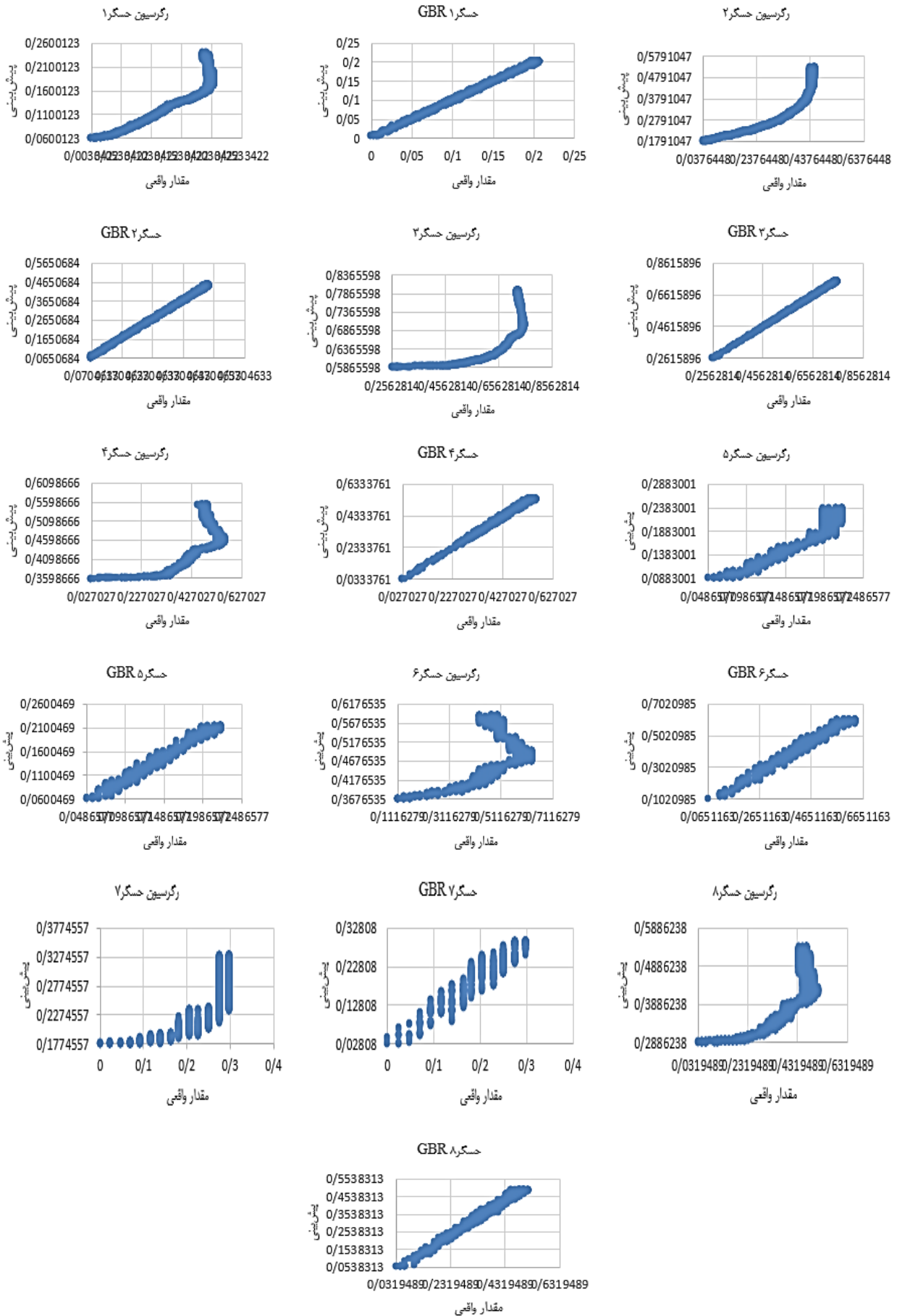
در این پژوهش، الگوریتم‌های یادگیری ماشینی برای تحلیل داده‌های حسگر استفاده شدند. با استفاده از این روش می‌توان الگوهای مختلف در داده‌های حسگر را شناسایی و تحلیل کرد. بدین منظور الگوریتم رگرسیون گرادیان افزایشی با مشخصات ذکر شده در جدول ۵ در پایتون پیاده‌سازی شد.

جدول ۵. مشخصات الگوریتم یادگیری ماشینی رگرسیون گرادیان افزایشی

حسگر ۸	حسگر ۷	حسگر ۶	حسگر ۵	حسگر ۴	حسگر ۳	حسگر ۲	حسگر ۱	
۰/۰۱	۰/۰۱	۰/۰۱	۰/۰۱	۰/۰۱	۰/۰۱	۰/۰۱	۰/۰۱	نرخ آموزش
۴	۴	۴	۴	۴	۴	۴	۴	عمق بیشینه
۵۰۰	۵۰۰	۵۰۰	۵۰۰	۵۰۰	۵۰۰	۵۰۰	۵۰۰	برآوردگر η
۰/۵	۰/۵	۰/۵	۰/۵	۰/۵	۰/۵	۰/۵	۰/۵	زبربخش

شکل ۵، میزان دقت و بازخوانی دو الگوریتم مختلف رگرسیون و رگرسیون گرادیان افزایشی برای پاسخ ۸ حسگر گازی مختلف آرایه حسگری به تیمار تیزابی را نشان می‌دهد. البته این بررسی برای هر دو بخش داده‌های آموزش و آزمون انجام شده است که در نهایت فقط نتایج داده‌های آزمون در شکل‌ها نشان داده شده است. ستون «مقدار واقعی» شامل مقادیر واقعی متغیر هدف که حسگرها نشان می‌دهند است و ستون «پیش‌بینی» شامل مقادیر پیش‌بینی شده توسط مدل برای پاسخ حسگرها از متغیر هدف است.

همان‌طور که در نمودارها مشخص است، در تمام حسگرها، مدل رگرسیون گرادیان افزایشی توانسته است پیش‌بینی دقیق‌تر و بهتری را ارائه کند، بنابراین مدل‌سازی توسط این مدل با مؤلفه‌های تعیین کیفیت مدل تجزیه و تحلیل و نتایج مدل‌سازی بررسی شد. در جدول ۶، داده‌ها مربوط به کیفیت پیش‌بینی بهترین مدل رگرسیون گرادیان افزایشی برای داده‌های حسگر است که شامل دو بخش آموزشی و آزمون است. ستون "آموزش" شامل مقادیر معیارهای کیفیت پیش‌بینی مدل برای داده‌های آموزشی و ستون "آزمون" شامل مقادیر معیارهای کیفیت پیش‌بینی مدل برای داده‌های آزمون است. مقدار ضریب تبیین برای داده‌های آموزشی و آزمون به ترتیب برابر با ۰/۹۹۷۶ و ۰/۹۹۷۳ است که نشان می‌دهد که مدل به خوبی با داده‌ها سازگار است. میانگین ریشه میانگین مربعات خطا برای داده‌های آموزشی و آزمون به ترتیب برابر با ۰/۰۰۲۸ و ۰/۰۰۳۱ است. این معیار نشان می‌دهد که میانگین خطای پیش‌بینی مدل برای داده‌های آزمون کمی بیشتر از داده‌های آموزش است. میانگین مطلق خطا برای داده‌های آموزش و آزمون به ترتیب برابر با ۰/۰۰۲۴ و ۰/۰۰۲۶ است. ریشه میانگین مربعات خطای نسبی برای داده‌های آموزشی و آزمون به ترتیب برابر با ۰/۰۱۸۷ و ۰/۰۲۰۹ است. باتوجه به مقادیر معیارها، به نظر می‌رسد که مدل به خوبی با داده‌ها سازگار است و دارای کیفیت پیش‌بینی بسیار خوبی در پیش‌بینی متغیر هدف برای داده‌های آموزشی و آزمون است. به‌طور کلی، باتوجه به مقادیر بالای معیارهای کیفیت پیش‌بینی، می‌توان نتیجه گرفت که مدل رگرسیون گرادیان طراحی شده، به خوبی با مجموعه داده سازگار است و می‌تواند برای پیش‌بینی مقادیر متغیر هدف به خوبی عمل کند.

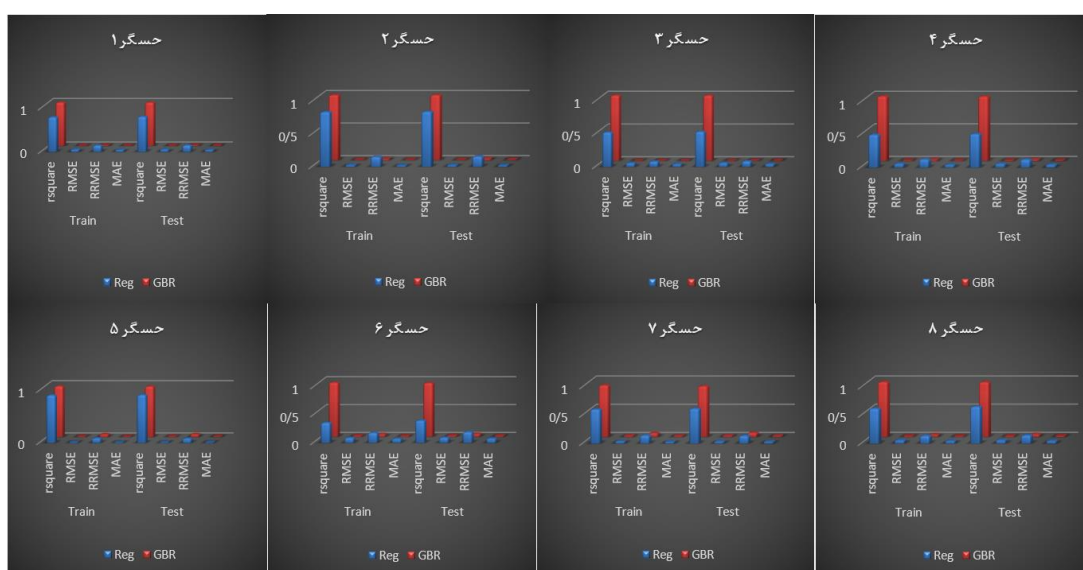


شکل ۵. مقایسه مقادیر واقعی و پیش‌بینی شده حسگرها در مدل‌های رگرسیون و رگرسیون گرادبان افزایشی

جدول ۶. مؤلفه‌های کیفیت پیش‌بینی بهترین مدل رگرسیون گرادیان افزایشی

	آموزش	تست
rsquare	۰/۹۹۷۵۹	۰/۹۹۷۲۴۴
RMSE	۰/۰۰۲۸۱۷	۰/۰۰۳۰۹۸
RRMSE	۰/۰۱۸۶۶۹	۰/۰۲۰۹۸۵
MAE	۰/۰۰۲۴۰۱	۰/۰۰۲۶۴۷

در شکل ۶، مؤلفه‌های پیش‌بینی کیفیت مدل در مدل‌های رگرسیون و گرادیان بردار افزایشی برای ۸ حسگر آرایه حسگری مقایسه شده‌اند تا مدل دقیق‌تر شناسایی شود.



شکل ۶. مقایسه کیفیت پیش‌بینی مدل در مدل‌های رگرسیون و گرادیان بردار افزایشی برای ۸ حسگر آرایه حسگری

باتوجه به مجموع نتایج و بررسی و مقایسه مقادیر مولفه‌های کیفیت مدل‌سازی برای مدل رگرسیون و مدل رگرسیون گرادیان افزایشی، کاملاً مشهود است که در مدل‌سازی تمام حسگرها، رگرسیون گرادیان افزایشی عملکرد بهتری داشته است و با دقتی بسیار خوب و قابل قبول توانسته است پاسخ حسگرهای مختلف را مدل‌سازی کند. در کل، باتوجه به نتایج این نمودارها می‌توان نتیجه گرفت که روش رگرسیون گرادیان افزایشی نسبت به روش رگرسیون برای هر چهار معیار ارزیابی بهتر عمل کرده است. بنابراین در جدول به آنالیز واریانس بررسی عملکرد مدل رگرسیون بردار افزایشی پرداخته شد.

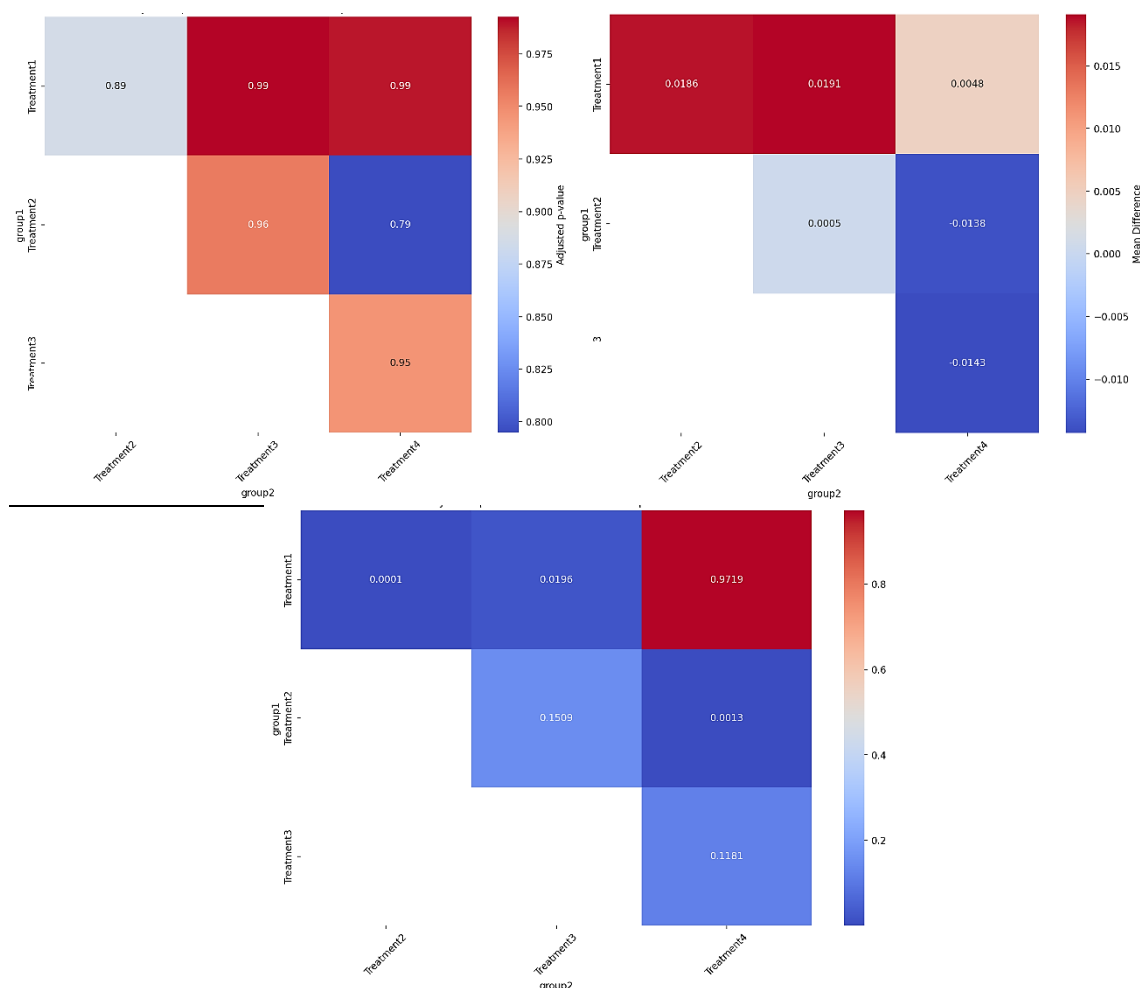
جدول ۷. آنالیز واریانس بررسی عملکرد مدل رگرسیون بردار افزایشی

	sum_sq	df	F	PR(>F)
Intercept	۷۰/۶۰۳۴۴	۱	۳۹۴۹/۴۶۶	۵/۳۵E-۸۱
C(Treatment, Sum)	۰/۴۷۶۳۲۴	۳	۸/۸۸۱۶۵۳	۲/۹E-۰۵
C(Sensore, Sum)	۰/۸۹۵۷۳۶	۷	۷/۱۵۸۰۴۹	۶/۶۱E-۰۷
C(Replication, Sum)	۰/۰۳۵۹۵۳	۳	۰/۶۷۰۲۸۳	۰/۵۷۲۲۲۲
Residual	۱/۷۵۱۹۱۷	۹۸		

اعداد جدول ۷ نشان‌دهنده آماره‌های مختلف آنالیز واریانس هستند که کمک می‌کنند تا عملکرد مدل بررسی شود. در تحلیل واریانس ارائه شده، مجموع مربعات خطا برای هر یک از متغیرهای مستقل برابر با ۷۰/۶ است که نشان می‌دهد میزان تغییر در متغیر وابسته که قابل توضیح توسط متغیر رهگیری نیست، برابر با ۷۰/۶ است. آماره اف برای هر یک از متغیرهای مستقل، نسبت میان مربعات تبدیل شده به فاصله میان میانگین‌های گروه را می‌سنجد. آماره اف

برای هر یک از متغیرهای مستقل باید بیشتر از ۱ باشد تا نشان دهنده تفاوت معنادار باشد. این پارامترها به صورت آزمایشی تعیین شده‌اند و ممکن است با توجه به نوع داده‌ها و مسئله مورد نظر، بهتر باشد که پارامترهای متفاوتی انتخاب شوند. به عنوان مثال، افزایش مقدار عمق حداکثر ممکن است باعث بالا رفتن دقت الگوریتم شود، اما همچنین ممکن است باعث بروز بیش‌برازش شود. به همین ترتیب، تعیین مقدار مناسب برای هر پارامتر باید با توجه به خصوصیات داده‌ها و مسئله مورد نظر انجام شود.

با توجه به شکل ۷ که جهت بررسی نتایج آزمون تفاوت معناداری بین میانگین تیمارها ارائه شده است؛ ضریب تبیین مدل سازی در میانگین تیمار تیزابی تفاوت معنی داری با تیمار گوگردی داشته و بهتر بوده است که نشان دهنده این موضوع است که تیمار تیزابی بیشترین قابلیت مدل سازی و در نتیجه پیش بینی پذیری پاسخ حسگرها را داشته که با توجه به تفاوت بوی محسوس تر ایجاد شده توسط تیزاب این موضوع توجیه پذیر است. همچنین تیمار کشمش آفتابی کمترین قابلیت مدل سازی را نسبت به بقیه تیمارها از خود نشان داده است که می تواند دلیل مشابهی داشته باشد. در هر حال در نتایج ریشه میانگین مربعات خطا، میانگین مطلق خطا و ریشه میانگین مربعات خطای نسبی بین میانگین تیمارها تفاوت معنی داری مشاهده نشد.



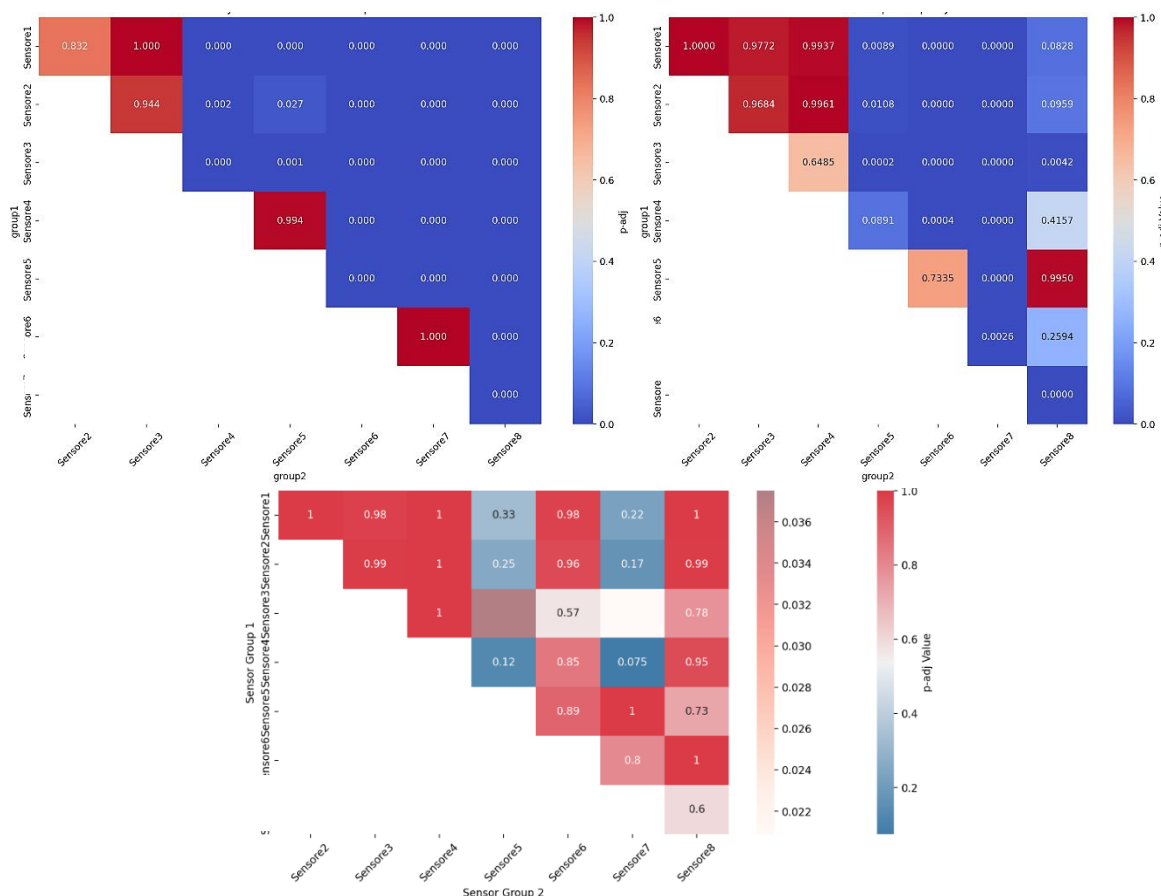
شکل ۷. نتایج آزمون تفاوت معناداری بین میانگین تیمارها الف) RMSE (ب) RRMSE (ج) rsquare

شکل ۸، مقایسه میان میانگین پاسخ هر یک از حسگرها را نشان می‌دهد و بررسی می‌کند آیا تفاوت معنی داری بین آنها وجود دارد یا خیر. به عنوان مثال، مقایسه اول، میانگین تفاوت بین حسگر ۱ و ۲ است که برابر با ۰/۰۰۱۲ است. مقدار p-adj نیز برابر با ۰/۰۹۱۱ است که نشان دهنده عدم وجود تفاوت معناداری در این مقایسه است. همچنین، محدوده CI نیز برای این مقایسه از ۰/۰۰۱۸- تا ۰/۰۰۴۳ است که نشان دهنده این است که تفاوت میانگین دو تیمار در کنار این حسگر، با اطمینان ۹۵٪، در این بازه قرار می‌گیرد. مقدار reject

نیز نشان دهنده تأیید یا رد شدن فرض صفر (عدم تفاوت معنادار) است که با رنگ آبی یا قرمز مشخص شده است. اگر مقدار reject برابر با True باشد، فرض صفر رد شده و به این معنی است که تفاوت معناداری بین دو حسگر وجود دارد. اما اگر مقدار reject برابر با False باشد، فرض صفر تأیید شده و به این معنی است که تفاوت معناداری بین دو حسگر وجود ندارد. در این شکل، برای مقایسه حسگرهای ۱، ۲، ۳، ۴، ۵، ۶ و ۷ فرض صفر تأیید شده است، به این معنی که تفاوت معناداری بین این حسگرها وجود ندارد و در بقیه مقایسه‌ها، فرض صفر رد شده و به این معنی است که تفاوت معناداری بین آنها وجود دارد.

با بررسی جداول مشابه برای بررسی نتایج آزمون تفاوت معناداری بین میانگین حسگرها در داده‌های آزمون برای ضریب تبیین، فقط میانگین پاسخ حسگرهای ۳ و ۵ و ۳ و ۷ تفاوت معنی‌داری داشتند که حسگر ۳ ضریب تبیین بهتری از دو حسگر دیگر به دست آورده بود. همچنین در خصوص ریشه میانگین مربعات خطا، حسگر ۱، ۲ و ۳ از حسگرهای ۴، ۵، ۶ و ۷ با تفاوت معنی‌داری خطای کمتری را ثبت کرده بودند و همچنین حسگر ۴ و ۵ نسبت به حسگرهای ۶، ۷ و ۸ بهتر بودند. و هر دو حسگر ۶ و ۷ نیز از حسگر ۸ با خطای کمتری مدل‌سازی شده بودند.

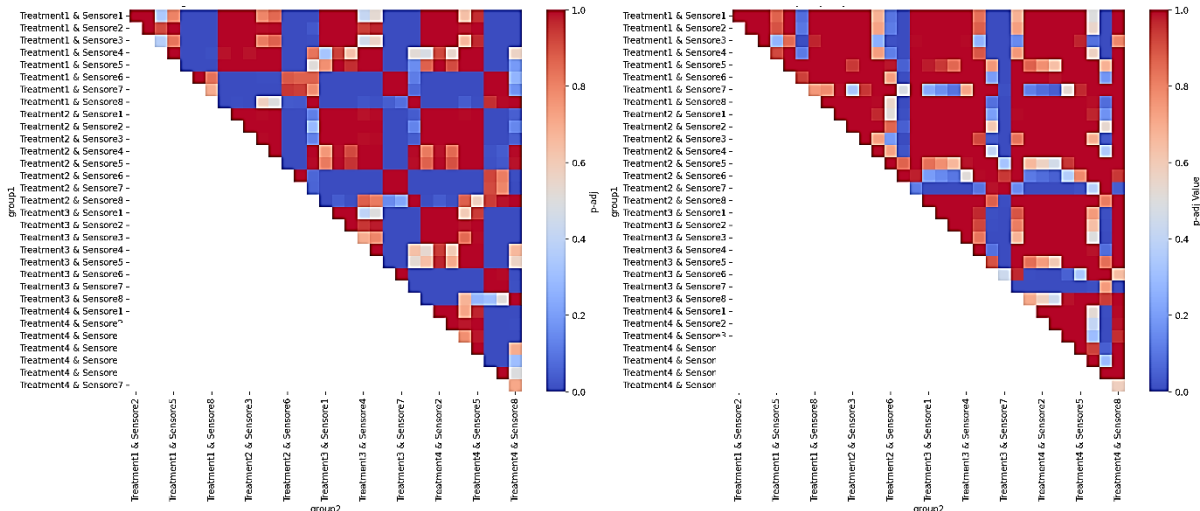
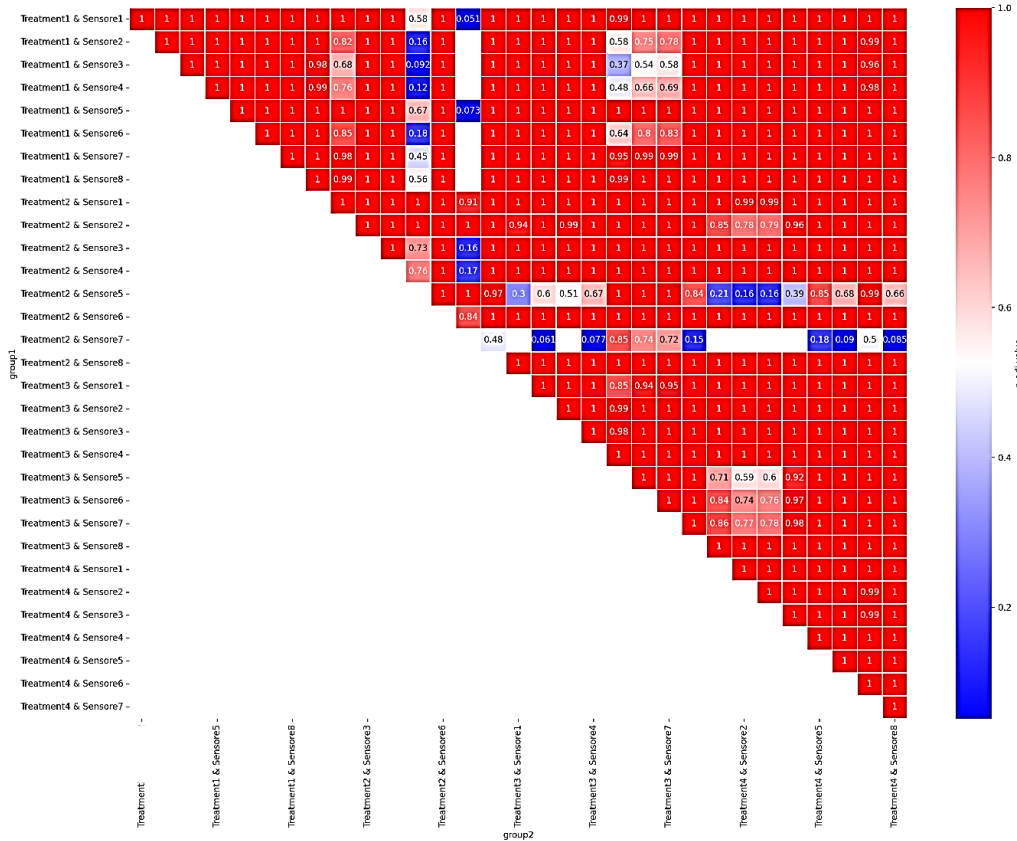
در بررسی نتایج آزمون تفاوت معناداری بین میانگین حسگرها در داده‌های آزمون برای ریشه میانگین مربعات خطای نسبی نیز مشخص شد؛ حسگر ۱ و ۲ از حسگرهای ۵، ۶ و ۷ با تفاوت معنی‌داری کم خطا تر بوده‌اند و حسگر ۳ از حسگرهای ۵، ۶ و ۷ بهتر بوده است. همچنین حسگر ۴ از حسگرهای ۶ و ۷ و حسگر ۵ و ۶ از ۷ و حسگر ۷ از حسگر ۸ با تفاوت معنی‌داری کم خطا تر مدل‌سازی شده‌اند. در نهایت در خصوص میانگین مطلق خطا نیز می‌توان گفت، حسگر ۱ از حسگرهای ۴، ۵، ۶ و ۷ و ۸ خطای کمتری را به شکل معنی‌داری ثبت کرده است و حسگر ۲ و ۳ نسبت به حسگرهای ۴، ۵، ۶ و ۷ و ۸ چنین شرایطی داشته است. همچنین حسگرهای ۴ و ۵ نسبت به حسگرهای ۴، ۵، ۶ و ۷ و ۸ و حسگر ۶ و ۷ نسبت به حسگر ۸ به شکل معنی‌داری کم خطا تر مدل‌سازی شده بودند.



شکل ۸. نتایج آزمون تفاوت معناداری بین میانگین حسگرها الف) RMSE ب) RRMSE ج) rsquare

پس از بررسی روندهای مشابه و شکل ۹ در نتایج آزمون تفاوت معناداری بین میانگین حسگرها در تیمارهای مختلف در داده‌های آزمون نیز مشخص شد؛ حسگر ۷ در تیمار تیزابی با تفاوت معنی‌داری ضریب تبیین بالاتری نسبت به تیمار آفتابی مدل‌سازی شده است که

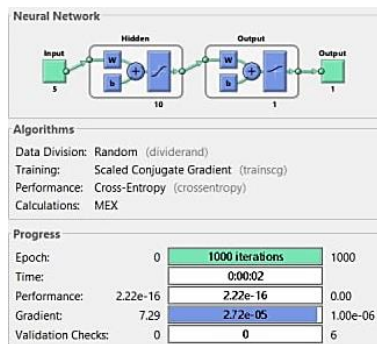
این مسئله، نتایج مقایسه تیمارها با هم را تأیید می‌کند. همچنین در خصوص ریشه میانگین مربعات خطا، حسگرهای ۶، ۷ و ۸ در تیمار تیزابی با تفاوت معنی‌داری خطای بیشتری نسبت به حسگر ۱ در این تیمار داشته‌اند. همچنین حسگرهای ۶، ۷ و ۸ نیز با تفاوت معنی‌داری از حسگرهای ۱، ۳، ۴ و ۵ در تیمار تیزابی عملکرد بدتری داشته‌اند. در مورد تیمار گوگردی نیز حسگرهای ۱، ۲ و ۳ از حسگرهای ۶ و ۷ با تفاوت معنی‌داری خطای کمتری ثبت کرده‌اند و حسگر ۴ و ۵ نیز همین شرایط را نسبت به حسگر ۶ و ۷ داشتند. البته در تیمار آفتابی نیز حسگرهای ۱، ۲ و ۳ نسبت به حسگرهای ۶ و ۷ و ۸ همین شرایط را داشتند و حسگرهای ۴ و ۵ نیز از حسگرهای ۶ و ۷ با تفاوت معنی‌داری عملکرد کم‌خطاتری را ثبت کرده بودند که تمام نتایج، مقایسه میانگین نتایج حسگرها را تأیید می‌کنند.



شکل ۹. نتایج آزمون تفاوت معناداری بین میانگین حسگرها در تیمارهای مختلف الف) RMSE (ب) RRMSE (ج) R-square

همچنین با بررسی شبکه عصبی چندلایه MLP که برای پاسخ حسگرهای گازی با ده لایه مخفی و تابع فعالسازی تانژانت سیگموئید در نرم افزار نتورال سلوشن ۵ پیاده‌سازی شد، دقت این شبکه عصبی برای تفکیک نمونه‌ها برابر با ضریب همبستگی برای کل داده‌ها برابر

۰/۹۸۶، برای داده‌های آموزش برابر ۰/۹۹۷، برای داده‌های آزمون برابر ۰/۹۹۷، برای داده‌های اعتبارسنجی برابر ۰/۹۹۱ و همچنین مقدار RMSE برای کل داده‌ها برابر ۰/۱۱۵، برای داده‌های آموزش برابر ۰/۱۱۰، برای داده‌های آزمون برابر ۰/۲۵۸ و برای داده‌های اعتبارسنجی برابر ۰/۲۶۳ بدست آمد.



شکل ۱۰. شبکه عصبی مصنوعی چندلایه ایجاد شده جهت تفکیک نمونه‌ها

نتیجه‌گیری

حسگرهای گازی دارای محدودیت‌هایی هستند که با استفاده از یادگیری ماشین می‌توان بر آنها غلبه کرد. از جمله این محدودیت‌ها می‌توان به انتخاب‌پذیری محدود اشاره کرد، به این معنی که این حسگرها می‌توانند به چندین گاز پاسخ دهند که می‌تواند در برنامه‌های کاربردی که در آن انتخاب‌پذیری بالا مورد نیاز است، زیان‌آور باشد. همچنین حسگرهای گازی اغلب به شرایط عملیاتی کنترل شده نیاز دارند که می‌تواند در شرایط واقعی چالش‌برانگیز باشد. حسگرهای گازی ممکن است خطاهای رانش داشته باشند که این مسئله می‌تواند بر دقت و قابلیت اطمینان آنها در طول زمان تأثیر بگذارد و در نهایت توسعه و ساخت حسگرهای گازی می‌تواند پرهزینه و پیچیده باشد که یادگیری ماشین می‌تواند با بهبود عملکرد حسگرهای موجود و کاهش نیاز به سخت‌افزار اضافی، راه حل کارآمد و مقرون به صرفه‌تری ارائه دهد (Leon-Medina et al., 2020; Ye et al., 2021).

در این پژوهش با توجه به جمیع بررسی‌ها و تحلیل‌های انجام شده جهت مدل‌سازی پاسخ حسگرهای گازی استفاده شده و با بررسی مؤلفه‌های کیفیت مدل، اولاً مشخص شد؛ مدل رگرسیون گردایان افزایشی با ضریب تبیین ۰/۹۹۷۲، ریشه میانگین مربعات خطای ۰/۰۲۰۹، میانگین مطلق خطای ۰/۰۲۶ و ریشه میانگین مربعات خطای نسبی ۰/۰۲۰۹ برای داده‌های آزمون به خوبی توانسته است پاسخ حسگرهای گازی نسبت به تیمارهای معرفی شده را مدل‌سازی کند که این کار می‌تواند به غلبه بر محدودیت‌های حسگرهای گازی، مانند شرایط عملیاتی سخت، خطاهای رانش، انتخاب محدود، نیاز به مقدار زیادی از داده‌های برچسب‌گذاری شده و چالش‌های هزینه و ساخت کمک کند و از این نتایج استفاده کرد.

همچنین با بررسی و تحلیل نتایج به دست آمده، نوع و میزان همبستگی بین پاسخ حسگرها نسبت به هم و نسبت به زمان مشخص شد تا در پیش‌بینی رفتار آنها مورد ارزیابی قرار بگیرد. سپس با مدل‌سازی انجام شده مشخص شد حسگرهای ۱، ۲، ۳ و ۴ به ترتیب با ضرایب تبیین ۰/۸۶۶۸، ۰/۸۷۸۶، ۰/۹۴۵۸ و ۰/۹۰۷۴ و ریشه میانگین مربعات خطای ۰/۰۱۶۳، ۰/۰۱۶۸، ۰/۰۰۸۳ و ۰/۰۲۲۷ پاسخ‌های دقیق‌تر و پیش‌بینی‌پذیرتری نسبت به حسگرهای ۵، ۶، ۷ و ۸ نشان دادند که می‌توان نتیجه گرفت برای استفاده در آرایه حسگری سامانه ماشین بویایی مورد نظر گزینه‌های مناسب‌تری هستند. همچنین با تجزیه و تحلیل مدل‌سازی پاسخ حسگرها در تیمارهای مختلف مشخص شد، تیمار تیزابی قابل پیش‌بینی‌ترین تیمار و تیمار آفتابی کم‌دقت‌ترین تیمار در مدل‌سازی‌های انجام شده است که نشان می‌دهد سامانه ماشین بویایی نسبت به کدام تیمارها با دقت بیشتری می‌تواند مدل مناسبی جهت ارزیابی ارائه نماید. همچنین این نتایج با تحلیل نتایج مدل‌سازی حسگرهای مختلف در تیمارهای مختلف بررسی شد که با مشابَهت نتایج، تصدیق شدند.

همچنین مدل شبکه عصبی ایجاد شده نیز توانست با ضریب همبستگی برابر ۰/۹۸۶ برای کل داده‌ها، برای داده‌های آموزش برابر ۰/۹۹۷، برای داده‌های آزمون برابر ۰/۹۹۷، برای داده‌های اعتبارسنجی برابر ۰/۹۹۱ و همچنین مقدار ریشه میانگین مربعات خطا برای کل داده‌ها برابر ۰/۱۱۵، برای داده‌های آموزش برابر ۰/۱۱۰، برای داده‌های آزمون برابر ۰/۲۵۸ و برای داده‌های اعتبارسنجی برابر ۰/۲۶۳ تفکیک مناسبی را انجام دهد.



منابع

- کرامت چهرمی، مهدی، محتسبی، سید سعید، موسی زاده، حسین، قاسمی ورنامخواستی، مهدی، رفیعی، شاهین و سوند رومی، احسان. (۱۳۹۸). ارزیابی قابلیت ماشین بویایی در تفکیک کیفیت خرماي خشک شده با روش های الکتروهیترودینامیک، هوای گرم و ترکیبی. *مهندسی بیوسیستم ایران*. ۲۴۱-۲۵۱، (۱)۵۰.
- ازگلی، حافظ، محتسبی، سید سعید، حسین پور، سلیمان و حسین پور زرنق، محمد. (۱۴۰۲). بررسی کیفیت گوشت و روغن مصرف شده در ناگت مرغ با استفاده از بینی الکترونیکی و پردازش تصویر. *مهندسی بیوسیستم ایران*. ۱۴-۱۴۵۴، (۲)۵۴.

REFERENCES

- Baha, H., & Dibi, Z. (2010). ANN modeling of a gas sensor. *Journal of Electrical Engineering & Technology*, 5(3), 493–496. <https://doi.org/10.5370/jeet.2010.5.3.493>.
- Borowik, P., Adamowicz, L., Tarakowski, R., Siwek, K., & Grzywacz, T. (2020a). Odor detection using an E-Nose with a reduced sensor array. *Sensors*, 20(12), 3542. <https://doi.org/10.3390/s20123542>
- European Food Safety Authority (EFSA). (2016). Panel on food additives and nutrient sources added to food (ANS). Scientific opinion on the re-evaluation of sulfur dioxide (E 220), sodium sulfite (E 221), sodium bisulfite (E 222), sodium metabisulfite (E 223), potassium metabisulfite (E 224), calcium sulfite (E 226), calcium bisulfite (E 227), and potassium bisulfite (E 228) as food additives. *EFSA Journal*. 14: 4438. [DOI: 10.2903/j.efsa.2016.4438]
- Ghasemi-Varnamkhasti, M., Mohammad-Razdari, A., Yoosefian, S. H., Izadi, Z., & Rabiei, G. (2019). Selection of an optimized metal oxide semiconductor sensor (MOS) array for freshness characterization of strawberry in polymer packages using response surface method (RSM). *Postharvest Biology and Technology*, 151, 53–60. <https://doi.org/10.1016/j.postharvbio.2019.01.016>
- Guido L.F. (2016). Sulfites in beer: reviewing regulation, analysis and role. *Scientia Agricola*. 73: 189-197. [DOI: 10.1590/0103-9016-2015-0290]. <https://doi.org/10.1016/j.jspr.2021.101805>
- Institute of Standards and Industrial Research of Iran (ISIRI). (2014). Gheisi (Whole dried apricot) specifications and test methods. National Standard No. 13. 4th revision. URL: <http://standard.isiri.gov.ir/StandardView.aspx?Id=41629>. Accessed 7 June 2014.
- Institute of Standards and Industrial Research of Iran (ISIRI). (2015). Dried apricots - specification and test methods. National Standard No. 11. 5th revision. URL: <http://standard.isiri.gov.ir/StandardView.aspx?Id=40024>. Accessed 22 November 2015.
- Institute of Standards and Industrial Research of Iran (ISIRI). (2017). Dried fruits- determination of sulfur dioxide. National Standard No. 569. URL: <http://standard.isiri.gov.ir/StandardView.aspx?Id=47184>. Accessed 25 December 2017.
- Institute of Standards and Industrial Research of Iran (ISIRI). (2018). Specification and methods of test for fruit snack (fruit paste). National Standard No. 3308. 2nd revision. URL: <http://standard.isiri.gov.ir/StandardView.aspx?Id=50065>. Accessed 25 August 2018.
- Jamalzadeh, F., Ghasemi-Varnamkhasti, M., Ghasemi Nafchi, M., Tohidi, M., & Dowlati, M. (2020). Implementation of an olfactory machine system for the classification of different types of black pepper based on geographical origin and detection of cheating in Indian black pepper. *Iranian Food Science and Technology Research Journal*, 16(4), 479-491.
- Keramat-Jahromi, M., Mohtasebi, S. S., Mousazadeh, H., Ghasemi-Varnamkhasti, M., rafiee, S., & Savand-Roumi, E. (2019). Evaluation of a Machine Olfaction to Classify the Quality of Dried Date Fruit by Electrohydrodynamic, Hot Air, and the Hybrid Drying Techniques. *Iranian Journal of Biosystems Engineering*, 50(1), 241-251. doi: 10.22059/ijbse.2018.248873.665023. (In Persian).
- Kheireddine, Lamamra., Djamil, Rechem. (2016). Artificial neural network modelling of a gas sensor for liquefied petroleum gas detection. doi: 10.1109/ICMIC.2016.7804292
- Khodamoradi, F., Mirzaee-Ghaleh, E., Dalvand, M. J., & Sharifi, R. (2021). Classification of basil plant based on the level of consumed nitrogen fertilizer using an olfactory machine. *Food Analytical Methods*, 14(12), 2617–2629. <https://doi.org/10.1007/s12161-021-02089-y>
- Kiani, S., Minaei, S., & Ghasemi-Varnamkhasti, M. (2017). Integration of computer vision and electronic nose as non-destructive systems for saffron adulteration detection. *Computers and Electronics in Agriculture*, 141, 46–53. <https://doi.org/10.1016/j.compag.2017.06.018>
- Leon-Medina, J. X., Pineda-Muñoz, W. A., & Burgos, D. a. T. (2020). Joint distribution adaptation for drift

- correction in electronic NOSE type sensor arrays. *IEEE Access*, 8, 134413–134421. <https://doi.org/10.1109/access.2020.3010711> Ye, Z., Liu, Y., & Li, Q. (2021). Recent Progress in Smart Electronic Nose Technologies Enabled with Machine Learning Methods. *Sensors*, 21(22), 7620. <https://doi.org/10.3390/s21227620>
- Liu, Y., Wang, Q., Xu, Q., Feng, J., Yu, H., & Yin, Y. (2018). Non-destructive detection of Flos *Lonicerae* treated by sulfur fumigation based on hyperspectral imaging. *Journal of Food Measurement and Characterization*, 12(4), 2809–2818.
- Liu, Y., Wang, Q., Xu, Q., Feng, J., Yu, H., & Yin, Y. (2018b). Non-destructive detection of Flos *Lonicerae* treated by sulfur fumigation based on hyperspectral imaging. *Journal of Food Measurement and Characterization*, 12(4), 2809–2818. <https://doi.org/10.1007/s11694-018-9896-z>
- Maho, P., Herrier, C., Livache, T., Comon, P., & Barthelmé, S. (2022). A calibrant-free drift compensation method for gas sensor arrays. *Chemometrics and Intelligent Laboratory Systems*, 225, 104549. <https://doi.org/10.1016/j.chemolab.2022.104549>
- Makarichian, A., Chayjan, R. A., Ahmadi, E., & Mohtasebi, S. S. (2021). Assessment the influence of different drying methods and pre-storage periods on garlic (*Allium Sativum* L.) aroma using electronic nose. *Food and Bioproducts Processing*, 127, 198–211. <https://doi.org/10.1016/j.fbp.2021.02.016>
- Men, H., Shi, Y., Jiao, Y., Gong, F., & Liu, J. (2018). Electronic nose sensors data feature mining: a synergetic strategy for the classification of beer. *Analytical Methods*, 10(17), 2016–2025. <https://doi.org/10.1039/c8ay00280k>
- Meng Z., Qin G., Zhang B. (2005). DNA damage in mice treated with sulfur dioxide by inhalation. *Environmental and Molecular Mutagenesis*. 46: 150-155. [DOI: 10.1002/em.20142].
- Mischek D., Krapfenbauer-Cermak C. (2012). Exposure assessment of food preservatives (sulphites, benzoic and sorbic acid) in Austria. *Food Additives and Contaminants: Part A*. 29: 371-382. [DOI: 10.1080/19440049.2011.643415]
- Mohammad-Razdari, A., Ghasemi-Varnamkhasti, M., Yoosefian, S. H., Izadi, Z., & Siadat, M. (2019). Potential application of electronic nose coupled with chemometric tools for authentication assessment in tomato paste. *Journal of Food Process Engineering*, 42(5). <https://doi.org/10.1111/jfpe.13119>
- Ordukaya, E., & Karlık, B. (2017). Quality control of olive oils using machine learning and electronic nose. *Journal of Food Quality*, 2017, 1–7. <https://doi.org/10.1155/2017/9272404>
- Ozgoi, H., Mohtasebi, S. S., Hosseinpour, S., & Hosseinpour-Zarnaq, M. (2023). Investigating meat and oil quality in chicken nuggets using electronic nose and image processing techniques. *Iranian Journal of Biosystems Engineering*, 54(2), 1-14. doi: 10.22059/ijbse.2023.363744.665517. (In Persian).
- Rahimzadeh, H., Sadeghi, M., Mireei, S. A., & Ghasemi-Varnamkhasti, M. (2022). Unsupervised modelling of rice aroma change during ageing based on electronic nose coupled with bio-inspired algorithms. *Biosystems Engineering*, 216, 132–146. <https://doi.org/10.1016/j.biosystemseng.2022.02.010>
- Soubra L., Sarkis D., Hilan C., Verger P. (2007). Dietary exposure of children and teenagers to benzoates, sulphites, butylhydroxyanisol (BHA) and butylhydroxytoluen (BHT) in Beirut (Lebanon). *Regulatory Toxicology and Pharmacology*. 47: 68-77. [DOI: 10.1016/j.yrtph.2006.07.005].
- Souhil, Kouda., T., Bendib., Samir, Barra., Abdelghani, Dendouga. (2018). ANN modeling of an industrial gas sensor behavior. doi: 10.1109/CCEE.2018.8634510
- Tahri, K., Bari, N. E., & Bouchikhi, B. (2016). Geographical classification and adulteration detection of cumin by using electronic sensing coupled to. . . ResearchGate. https://www.researchgate.net/publication/346355867_Geographical_classification_and_adulteration_detection_of_cumin_by_using_electronic_sensing_coupled_to_multivariate_analysis
- Tohidi, M., Ghasemi-Varnamkhasti, M., Ghafarinia, V., Mohtasebi, S. S., & Bonyadian, M. (2018). Identification of trace amounts of detergent powder in raw milk using a customized low-cost artificial olfactory system: A novel method. *Measurement*, 124, 120–129. <https://doi.org/10.1016/j.measurement.2018.04.006>
- Vally H., Misso N.L.A., Madan V. (2009). Clinical effects of sulphite additives. *Clinical and Experimental Allergy*. 39: 1643-1651. [DOI: 10.1111/j.1365-2222.2009.03362.x].
- Yang, Y., Shahbeik, H., Shafizadeh, A., Masoudnia, N., Rafiee, S., Zhang, Y., Pan, J., Tabatabaei, M., & Aghbashlo, M. (2022). Biomass microwave pyrolysis characterization by machine learning for sustainable rural biorefineries. *Renewable Energy*, 201, 70–86. <https://doi.org/10.1016/j.renene.2022.11.028>
- Zhou, M., Khir, R., Pan, Z., Campbell, J. F., Mutters, R., & Hu, Z. (2021). Feasibility of detection of infested rice using an electronic nose. *Journal of Stored Products Research*, 92, 101805.



The Use of Gradient Boost Regression Model to Modeling of Gas Sensors in Diagnosis of Sun-dried, Sulphurous and Acidic solution dried Raisins

EXTENDED ABSTRACT

Introduction

Remaining elements such as sulfur dioxide and its compounds, which are widely used in dried fruits as preservatives due to their availability and affordability, are one of the factors that importing countries consider when purchasing raisins from Iran. Therefore, exported raisins should be examined for the presence of these elements in the final product. Machine learning modeling can help overcome some of the limitations of gas sensors, such as high operating conditions, drift errors, limited selectivity, the need for a large amount of labeled data, and cost and manufacturing challenges. The gradient boost regression model is a machine learning model used to solve regression problems.

Materials and Methods

In this study, three treatments, including sun-dried, Acidic solution dried Raisins and sulfur-treated each with three replicates, were prepared and exposed to the gas sensors for 60 minutes to record the sensor responses to each treatment. The obtained data were then analyzed using machine learning models to determine the accuracy of each modeling method and make them comparable. The model evaluation parameters were examined, and the interpretation of each was discussed in detail. Finally, the analysis of variance of the gradient boost regression model was performed for each quality prediction component separately for treatments, sensors, and combinations of sensors with treatments, and various points were extracted from the interpretation of each in the discussion and results section.

Results and Discussion

Based on the charts and results, the gradient boost regression model has been able to provide more accurate and better predictions in all sensors. Therefore, the modeling by this model with the quality determining components of the model was analyzed and the modeling results were examined. Overall, considering the high values of quality prediction metrics, it can be concluded that the designed gradient boost regression model is well compatible with the dataset and can effectively predict the target variable. The significant difference test results also showed significant differences between the mean treatments. Treatments 1 to 3, corresponding to acidic solution dried raisins, sulfuric-treated raisins, and sun-dried raisins, were found to have significant differences. According to the results, the coefficient of determination in the Acidic solution treatment had a significant difference compared to the Sun-dried and sulfuric treatments, and performed better than both. Additionally, the Sun-dried treatment had a significant difference compared to the solar treatment, and the results of the solar treatment were better, indicating that the Acidic solution treatment had the highest modeling capability and predictability of sensor responses, which can be justified by the more noticeable odor created by acid. Furthermore, the Sun-dried treatment showed the lowest modeling capability compared to other treatments, which can be justified by the lack of clear processing performed on it compared to other treatments. However, no significant differences were observed in the root mean squared error, mean absolute error, and root mean squared error of relative error between the mean treatments. Comparing the means of responses of each sensor also showed that for the comparisons of sensors 1 and 2, 1 and 3, 2 and 3, 2 and 5, 4 and 5, and 6 and 7, the reject value was False, indicating that the null hypothesis was accepted, meaning that there is no significant difference between these sensors. In the other comparisons, the reject value was True, indicating that the null hypothesis was rejected and there is a significant difference between them.

Conclusion

The results showed that the gradient boost regression model with the coefficient of explanation of 0.9972 and the root mean square error of 0.0209 for the test data was able to model the response of the gas sensors compared to treatments. Also, by examining and analyzing the obtained results, the type and degree of correlation between the response of the sensors in relation to each other and in relation to time was determined to be evaluated in predicting their behavior. Then, with the modeling done, it was determined that the MQ9, MQ3, MQ5, TGS2620 sensors have coefficients of explanation of 0.8668, 0.8786, 0.9458, and 0.9074, respectively, and the root mean square error of 0.0163, 0.0168, and 0.0083. 0 and 0.0227 showed more accurate and predictable responses than MQ135, TGS822, TGS810 and MQ4 sensors.