Research Article

Agricultural Engineering., 49(2) (2026) 359- 383

ISSN (P): 2588-526X

DOI: 10.22055/agen.2026.52983.1824

ISSN (E): 2588-5944

Evaluation of Advanced Pattern Recognition Models in Electronic Nose Signal Processing for Discrimination of Bio-based Fuel Blends (Diesel/Biodiesel/Fusel oil)

A.H. BasiratPour[1], S. M. Safieddin Ardebili[2]*, M. Mostafaei [3] and H. Zakidizaji [4]

1. Ph.D. Student, Department of Biosystems Engineering, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran
2. Associate Professor, Department of Biosystems Engineering, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran
3. Associate Professor, Department of Biosystems Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran
4. Associate Professor, Department of Biosystems Engineering, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran

Received: 19 August 2026 Accepted: 12 October 2026 *Corresponding Author: m.safieddin@scu.ac.ir

Abstract

Introduction: This investigation presents a rigorous experimental evaluation of a low-cost electronic nose (E-nose) framework, coupled with multivariate chemometric and supervised machine learning algorithms specifically Principal Component Analysis (PCA), Random Forest (RF), and Artificial Neural Networks (ANN) for the identification and discrimination of agricultural biofuel blends. The primary objective is to critically appraise the operational validity and physical limitations of static metal-oxide semiconductor sensing for fuel quality assurance, thereby establishing an empirical baseline to prevent publication bias and guide the robust deployment of bio-based renewable fuels in compression ignition engines.

Materials and Methods: An electronic nose system comprising an array of eight commercial metal-oxide semiconductor (MOS) gas sensors (MQ series: MQ2, MQ3, MQ5, MQ8, MQ9, MQ135, MQ137, and MQ138) was developed and evaluated across 15 distinct fuel classes, including neat fossil diesel (D100), waste cooking oil biodiesel (B100), fusel oil (F100), and their corresponding binary and ternary blends. A strictly balanced dataset of 225 experimental trials was established and partitioned into a training set (n=150) and an independent test set (n=75, with 5 replicates per class). Static steady-state voltage equilibrium responses were acquired and standardized. Data structure and variance distribution were explored using unsupervised Principal Component Analysis (PCA), while supervised multi-class classification and hyperparameter tuning were conducted using Random Forest (RF) and a Multilayer Perceptron Artificial Neural Network (ANN/MLP, 8-4-15 architecture). Sensor contribution was quantitatively assessed via Mean Decrease in Impurity (MDI), and dimensional sensitivity was examined by contrasting the full eight-sensor array with an optimized four-sensor subset.

Results and Discussion: Unsupervised exploratory analysis showed that although the first two principal components captured 74.96% of the cumulative variance (PC1: 46.10% and PC2: 28.86%), extensive spatial cluster overlap persisted across all 15 fuel treatments without clear class-separation boundaries. Supervised classification on the independent test set yielded poor discrimination, with RF and ANN attaining overall test accuracies of only 6.67% (95% CI: [2.88%, 14.68%]) and 8.00% (95% CI: [3.72%, 16.37%]), and Cohen's

kappa coefficients of κ=0.0000 and κ=0.0143, respectively. Both models converged near the theoretical random chance baseline (1/15≈6.67%), with McNemar's test confirming no statistically significant performance difference (p=1.0000). Random Forest feature importance showed a homogeneous impurity reduction across the array (MDI ranging narrowly from 0.1008 for MQ9 to 0.1465 for MQ8), indicating pronounced sensor cross-sensitivity and strong collinearity. Furthermore, pruning the array to four sensors marginally altered RF and ANN test accuracies to 10.67% and 5.33%, respectively, demonstrating the adverse impact of collinear dimensions (curse of dimensionality) under small sample constraints and confirming that scalar equilibrium signals lack the kinetic resolution required to decouple highly overlapping volatile organic compound profiles.

Conclusio:n The findings show that low-cost MOS electronic nose systems relying solely on steady-state equilibrium readings cannot achieve granular 15-class discrimination among chemically proximate diesel, biodiesel, and fusel oil blends, and machine learning complexity cannot compensate for baseline informational deficits. Consequently, future developments must shift from scalar static signals to dynamic transient-response profiling, sensor temperature modulation, and continuous chemometric regression modeling to establish dependable, real-time quality-control protocols in biofuel refining.

Acknowledgments: The authors are grateful to the Research Council of Shahid Chamran University of Ahvaz for financial support (SCU.AA1404.29840).

Keywords*: *Electronic nose, Biodiesel, Fusel oil, Machine learning, Metal oxide sensors.

مقاله پژوهشی

ارزیابی مدل‌های پیشرفته تشخیص الگو در پردازش سیگنال‌های بینی الکترونیکی ی جهت تفکیک مخلوط های سوخت های (دیزل/بیودیزل/روغن الکل)

امیرحسین بصیرت پور[۱]، سید محمد صفی الدین اردبیلی[۲*]، مصطفی مصطفایی[۳] و حسن ذکی دیزجی [۴]

۱- دانشجوی دکترا مهندسی مکانیزاسیون کشاورزی-انرژی، گروه مهندسی مکانیک بیوسیستم، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران

۲- دانشیار، گروه مهندسی مکانیک بیوسیستم، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران

۳- دانشیار، گروه مهندسی مکانیک بیوسیستم، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران

٤- دانشیار، گروه مهندسی مکانیک بیوسیستم، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران

تاریخچه مقاله

دریافت: ۱۴۰۵/۰۵/۲۸

پذیرش نهایی: ۱۴۰۵/۰۷/۲۰

کلمات کلیدی:

بینی الکترونیکی ،

بیودیزل، روغن الکل،

یادگیری ماشین،

حسگرهای اکسید فلزی

* عهـــــده دار مکاتبـــــات

m.safieddin@scu.ac.ir

چکیده

هدف از این پژوهش شناسایی دقیق و کنترل کیفیت سوخت‌های زیستی جایگزین، بر پایه آرایه‌ای از هشت حسگر گاز نیمه‌رسانای اکسید فلزی سری MQ در تفکیک ۱۵ کلاس آمیخته‌های سوختی دیزل، بیودیزل و روغن الکلدر شرایط قرائت استاتیک (حالت‌پایا) بود. داده‌های تجربی شامل ۲۲۵ نمونه (۱۵۰ نمونه آموزش و ۷۵ نمونه آزمون مستقل) با استفاده از تحلیل مؤلفه‌های اصلی (PCA) و دو مدل یادگیری ماشین غیرخطی شامل جنگل تصادفی (RF) و پرسپترون چندلایه (ANN/MLP) تحلیل شدند. نتایج نشان داد که با وجود تبیین ۷۴٫۹۶درصد از واریانس کل توسط دو مؤلفه نخست درPCA ، هم‌پوشانی شدیدی میان خوشه‌ها در فضای ویژگی‌ها وجود دارد. مدل‌های RF و ANN بر روی داده‌های آزمون مستقل به‌ترتیب به صحت اندک ۶٫۶۷درصد و ۸٫۰۰درصد و ضریب کاپای کوهن ۰٫۰۰۰و ۰٫۰۱۴۳ دست یافتند که منطبق بر آستانه تصمیم‌گیری تصادفی(۱/۱۵≈۶/۶۷)درصد بوده و آزمون مک‌نمار(p=۱٫۰۰) نیز عدم برتری معنادار میان آن‌ها را تأیید کرد. همچنین، تحلیل شاخص اهمیت ناخالصی MDI (بین ۰٫۱۰۰۸ تا ۰٫۱۴۶۵) گویای توزیع یکنواخت بار اطلاعاتی و فقدان حسگر مسلط بود. مقایسه پیکربندی‌ها نشان داد کاهش ابعاد به ۴ حسگر، صحت RF را تا ۱۰٫۶۷درصد افزایش داد که نشانه اثر منفی «نفرین ابعاد» در حضور داده‌های هم‌خط است. این یافته‌ها اثبات می‌کنند که قرائت‌های تک‌نقطه‌ای تعادلی به دلیل غفلت از سینتیک گذرای واکنش و حساسیت متقاطع بالای حسگرهای تجاری، فاقد غنای اطلاعاتی کافی برای تفکیک سوخت‌ها در مقیاس‌های ریزدانه هستند؛ از این رو، استخراج مشخصه‌های دینامیک و بازتعریف سطوح طبقه‌بندی ضروری است.

مقدمه

توسعه پرشتاب جوامع بشری، پیشرفت صنایع سنگین و مکانیزاسیون فزاینده بخش کشاورزی، در طول سده اخیر تقاضای جهان را برای انرژی و حامل‌های هیدروکربنی به شکل چشمگیری افزایش داده است (Uyumaz, 2017). مصرف گسترده و پیوسته سوخت‌های فسیلی، علاوه بر تسریع روند اتمام این ذخایر تجدیدناپذیر، پیامدهای زیست‌محیطی جبران‌ناپذیری همچون تشدید گرمایش جهانی و تغییرات شدید اقلیمی را در پی داشته است. هم‌زمان، انتشار آلاینده‌های خطرناکی نظیر اکسیدهای نیتروژن، مونوکسید کربن، هیدروکربن‌های نسوخته و ذرات معلق خروجی از موتورهای درون‌سوز دیزل، تهدیدی مستقیم برای سلامت عمومی و اکوسیستم‌های زیستی به شمار می‌رود. این چالش‌ها، جامعه بین‌المللی را به بازنگری در سیاست‌های انرژی، تصویب استانداردهای سخت‌گیرانه زیست‌محیطی و سرمایه‌گذاری گسترده در توسعه سوخت‌های پاک، جایگزین و تجدیدپذیر واداشته است(Kelanic, 2016).

وابستگی گسترده به آن‌ها حجم قابل‌توجهی از گازهای گلخانه‌ای را آزاد کرده، الگوهای اقلیمی را دگرگون ساخته و به بالا آمدن سطح آب‌های آزاد دامن زده است. همزمان، محدودشدن ذخایر نفتی و نگرانی‌های روزافزون زیست‌محیطی، حرکت به‌سوی حامل‌های انرژی تجدیدپذیر پایدار از جمله سوخت‌های زیستی، هیدروژن و گازهای سنتزی را اجتناب‌ناپذیر کرده است. در میان گزینه‌های موجود، بیودیزل و اتانول به‌سبب تجدیدپذیری، چرخه کربنی تقریباً بسته و انتشار کمتر آلاینده‌ها، جایگاه ویژه‌ای یافته‌اند (Kelanic, 2016; Sovacool and Ren, 2015).

امروزه اتکای غالب بخش‌های حمل‌ونقل و صنعت به سوخت‌های فسیلی، مجموعه‌ای درهم‌تنیده از معضلات زیست‌محیطی و اقتصادی پدید آورده است. افزایش گازهای گلخانه‌ای، افت کیفیت هوا و اتمام تدریجی منابع نفتی، ضرورت یافتن جایگزین‌های تجدیدپذیر را بیش از پیش آشکار می‌سازد (Safieddin Ardebili et al., 2020a).

الکل‌های بالاتر استخراج‌شده از روغن الکل محصول جانبی فرآیندهای تخمیر الکلی در این میان به‌عنوان گزینه امیدوارکننده برای موتورهای احتراق داخلی مطرح شده‌اند. روغن الکل که از پسماندهای کشاورزی و صنعتی به‌دست می‌آید، ظرفیت جهانی چشمگیری برای کاربرد در موتورهای دیزل و بنزینی دارد؛ هرچند ناپایداری احتراق و انتشار آلاینده‌ها همچنان از موانع اصلی آن به‌شمار می‌روند (Safieddin Ardebili et al., 2020a).

درمطالعه ای ترکیب این روغن با دیزل می‌تواند کارایی موتور را ارتقا دهد و آلاینده‌های مضر را کاهش دهد. افزودن نانو-بیوچار به مخلوط‌های الکل-دیزل، برای نمونه، بازده حرارتی را بالا برده و انتشار مونوکسید کربن و هیدروکربن‌های نسوخته را از راه بهبود احتراق و کاهش ذرات معلق کم کرده است (Safieddin Ardebili et al., 2020b) بهینه‌سازی چندهدف با روش سطح پاسخ روی ترکیبات سه‌تایی دیزل-بیودیزل-الکل نیز امکان برقراری تعادل میان توان خروجی، مصرف سوخت و اکسیدهای نیتروژن را فراهم ساخته است (Hassan Pour et al., 2018).

چنین رویکردهایی نه‌تنها وابستگی به منابع فسیلی را کم می‌کنند، بلکه با اهداف توسعه پایدار همسو هستند. با وجود این دستاوردها، محدودیت دسترسی جهانی به روغن الکل و گاه نیاز به اصلاحات سخت‌افزاری موتور همچنان باقی است (Safieddin Ardebili et al., 2020a). در کنار بهبود عملکرد، پایش دقیق کیفیت سوخت برای شناسایی ناخالصی‌ها و تضمین عملکرد مطمئن موتور اهمیت حیاتی دارد (Panoutsou et al., 2021; Shahbeik et al., 2022).

روش‌های کلاسیک آزمایشگاهی نظیر کروماتوگرافی گازی و انواع طیف‌سنجی قادرند محتوای بیودیزل را کمی کنند و ناخالصی‌ها را تشخیص دهند؛ اما این تکنیک‌ها مخرب، زمان‌بر، نیازمند نیروی متخصص و وابسته به تجهیزات و مواد شیمیایی پرهزینه هستند. بنابراین توسعه سامانه‌های تحلیلی سریع، غیرمخرب و قابل‌استقرار در محل به ضرورتی جدی بدل شده است.

بینی الکترونیکی (e-nose) پاسخی عملی به این نیاز ارائه می‌دهد. این سامانه با تقلید از حس بویایی انسان، ویژگی‌های نمونه را بر پایه پروفایل فرار آن تحلیل می‌کند (Beurey et al., 2021; Khorramifar et al., 2022).

با این حال، پاسخ حسگرهای نیمه‌هادی اکسید فلزی در مواجهه با مخلوط‌های پیچیده هیدروکربنی و الکولی با رفتارهای غیرخطی، حساسیت متقاطع و هم‌پوشانی شدید سیگنال‌ها همراه است، ساختار معمول بینی الکترونیکی شامل بخش‌های سخت‌افزاری و نرم‌افزاری است (Loutfi et al., 2015). گازهای آزادشده توسط آرایه حسگرها جذب می‌شوند و سیگنال‌های مشخصی تولید می‌کنند (Hu et al., 2023). این سیگنال‌ها را می‌توان با روش‌های آماری متمایز کرد؛ اما پیشرفت‌های هوش مصنوعی امکان بهره‌گیری از الگوریتم‌های یادگیری ماشین را برای ارزیابی کیفیت بر اساس همین داده‌ها فراهم آورده است (Wojnowski et al., 2017).

مطالعات پیشین ثابت کرده‌اند که بینی الکترونیکی قادر به تفکیک منابع مختلف بیودیزل و تشخیص ترکیبات آن با دیزل است و می‌تواند به‌عنوان ابزار غربالگری اولیه جایگزین تجهیزات گران‌قیمت عمل کند. اخیراً الگوریتم‌هایی برای تعیین نسبت ترکیبات بیودیزل و روغن الکل با دیزل بر پایه بینی الکترونیکی توسعه یافته که کارایی این فناوری را در کنترل کیفیت سوخت‌های ترکیبی با استفاده از یادگیری ماشین و روش های LDA, QDA, SVM تأیید می‌کند (Basiratpour et al., 2025).

بیودیزل از واکنش ترانس‌استریفیکاسیون روغن‌های گیاهی یا چربی‌های حیوانی با الکل به‌دست می‌آید. این سوخت جایگزین ضمن افزایش طول عمر منابع فسیلی، احتراق را در موتورهای دیزل بهبود می‌بخشد و به‌دلیل چرخه کربنی بسته، هم تقاضا برای نفت و هم انتشار گازهای گلخانه‌ای را کاهش می‌دهد. امکان اختلاط آن در نسبت‌های گوناگون با پترودیزل بدون نیاز به تغییر موتور وجود دارد. ترکیبات به‌صورت Bx معرفی می‌شوند که x درصد حجمی بیودیزل را نشان می‌دهد (برای مثال B5 معادل ۵ درصد بیودیزل و ۹۵ درصد دیزل، و B100 بیانگر بیودیزل خالص است). غلظت‌های رایج B5 و B20 هستند که مشخصات آن‌ها به‌ترتیب در استانداردهای ASTM D975 و ASTM D7467 تعریف شده است. در ایالات متحده B20 کاربرد گسترده‌ای دارد. نسبت بیودیزل را می‌توان بر اساس ملاحظات اقتصادی و زیست‌محیطی تنظیم کرد. دستورالعمل اتحادیه اروپا (EC/۲۸/۲۰۰۹) نیز کشورها را موظف به استفاده از حداقل ۱۰ درصد سوخت‌های تجدیدپذیر در بخش حمل‌ونقل تا سال ۲۰۲۰ کرده بود(Mahmodi et al., 2022).

تعیین کیفیت، اصالت‌سنجی و شناسایی درصد اختلاط سوخت‌های زیستی با سوخت‌های فسیلی به‌طور سنتی متکی بر روش‌های آزمایشگاهی استاندارد نظیر کروماتوگرافی گازی–طیف‌سنجی جرمی ، کروماتوگرافی مایع با کارایی بالا و طیف‌سنجی تبدیل فوریه فروسرخ است. علی‌رغم دقت تحلیلی و تکرارپذیری بسیار بالای این تکنیک‌ها، عواملی نظیر هزینه‌های سرسام‌آور تجهیزاتی و مصرفی، زمان‌بر بودن فرایند آنالیز، نیاز به آماده‌سازی پیچیده نمونه با حلال‌های شیمیایی آلاینده و الزام به حضور اپراتور متخصص، کاربری آن‌ها را به محیط‌های آزمایشگاهی محدود کرده و امکان پایش سریع و برخط را در جایگاه‌های توزیع و زنجیره تأمین سلب نموده است.

در سال‌های اخیر، سامانه‌های حسگری بویایی الکترونیکی مبتنی بر آرایه‌های حسگر گاز اکسید فلزی نیمه‌هادی به عنوان راهکاری غیرتخریبی، ارزان‌قیمت، سریع و قابل حمل مورد توجه پژوهشگران قرار گرفته‌اند. این سامانه‌ها بر خلاف روش‌های جداسازی شیمیایی، با تکیه بر پاسخ جمعی و هم‌پوشانی حسگرها نسبت به ترکیبات آلی فرار، یک «اثر انگشت شیمیایی» منحصربه‌فرد از مخلوط گازی بالای سوخت ایجاد می‌کنند. با اتصال این پاسخ‌های پیچیده به الگوریتم‌های پیشرفته یادگیری ماشین ، امکان طبقه‌بندی و پیش‌بینی نسبت ترکیبات چندگانه بدون نیاز به جداسازی فیزیکی اجزا فراهم می‌آید. جدول ۱

خلاصه‌ای جامع از مهم‌ترین مطالعات پیشین را نشان می‌دهد که از تلفیق فناوری بینی الکترونیکی و روش‌های هوش مصنوعی جهت تفکیک و کنترل کیفیت سوخت‌ها بهره گرفته‌اند.در سامانه‌های مبتنی بر حسگرهای MOS[1]، قرارگرفتن در معرض بخار باعث تغییر مقاومت الکتریکی حسگرها می‌شود؛ الگوی این تغییرات تقریباً برای هر ترکیب منحصربه‌فرد است و امکان شناسایی منبع را فراهم می‌آورد. شبکه عصبی مصنوعی (ANN)[2] از رایج‌ترین روش‌ها برای پردازش داده‌های بینی الکترونیکی e-nose[3] به‌شمار می‌رود. این تکنیک با الگوبرداری از فرآیندهای عصبی مغز، روابط پیچیده میان مجموعه‌های داده را از طریق آموزش استخراج و در موارد مشابه به‌کار می‌گیرد. مطالعات پیشین ترکیبات دیزل-بیودیزل با مواد خام و غلظت‌های گوناگون را با GC-MS و روش‌های شیمیایی دیگر بررسی کرده و نمونه‌ها را بر پایه منبع (سویا، روغن ضایعاتی، کلزا، چربی حیوانی) و غلظت (دیزل خالص تا B100) با PCA[4] خوشه‌بندی نموده‌اند. در پژوهشی دیگر، بینی الکترونیکی توانست چهار نمونه بیودیزل با منشأ متفاوت را از دیزل تفکیک کند و نشان داد که می‌تواند به‌عنوان ابزار غربالگری اولیه جایگزین تجهیزات پرهزینه شود. سوخت‌های زیستی معمولاً به دو دسته اولیه و ثانویه تقسیم می‌شوند. انواع اولیه به‌صورت خام برای گرمایش یا تولید برق به‌کار می‌روند، در حالی که انواع ثانویه مانند اتانول و بیودیزل پس از فرآوری برای حمل‌ونقل و صنعت مناسب‌اند. بیودیزل به‌عنوان استر متیل اسیدهای چرب زنجیره بلند خواصی نزدیک به دیزل دارد. اتانول نیز به‌دلیل عدد اکتان بالا، غیرسمی بودن و فراوانی منابع تجدیدپذیر، جایگاه مهمی در حمل‌ونقل یافته است (Cabaneros Lopez et al., 2019; Cinar et al., 2015). روغن الکل نیز به‌عنوان محصول جانبی تخمیر، عدد اکتان تحقیقاتی بالایی مشابه اتانول دارد و در تولید بیودیزل و سایر کاربردها مورد توجه قرار گرفته است (Calam et al., 2015; Li et al., 2019).

Table (1) Representative Studies on the Integration of Sensing and Analytical Techniques with Machine Learning

Analytical Approach	method	Study Scope and Objective	Reference
QDA, SVM, and ANN	Electronic nose equipped with MOS sensor	Identification and classification of in some fossil fuel adulteration products	Kakaei et al. (2025)
ANN, RSM	Electronic nose equipped with MOS sensor	Different blends of diesel and biodiesel fuels	Mahmodi et al. (2022)
SVM	Electronic nose	Engine Lubrication Oil Quality	Adelkhani and Daneshkhah (2025)
ANN	Electronic nose	Classification and identification of essential oils	Karami et al. (2020)
SVM/PLS	NIR spectroscopy	High-accuracy analysis of alcohol-diesel blend	Liu et al. (2026)
PCR and PLSR	GC-MS	Predict blend level of multiple-feedstock biodiesels	Shirley et al. (2026)
ANN	experimental data	Estimation of Properties of Petrodiesel—Biodiesel Mixtures	Doicin et al. (2025)

1- Metal Oxide Semiconductor
2- Artificial Neural Network
3- Electronic nose
4- Principal Component Analysis

علیرغم پژوهش‌های متعدد روی مخلوط‌های دوتایی گازوئیل-بیودیزل، شناسایی و طبقه‌بندی تفکیک‌شده مخلوط‌های سه‌تایی دیزل–بیودیزل–روغن الکل در درصدهای حجمی اندک و نزدیک به هم، به دلیل پیچیدگی بالای طیف ترکیبات فرار، همچنان یک چالش پژوهشی باز است (Basiratpour et al., 2025). همچنین ارزیابی تطبیقی مدل‌های مختلف یادگیری ماشین به ویژه مقایسه مدل‌های آماری قوی مانند QDA و SVM در برابر مدل‌های مستعد بیش‌برازش در داده‌های تجربی (مانند ANN و RF) در این زمینه کمتر مورد توجه قرار گرفته است.

هدف از این پژوهش، طراحی و توسعه یک سامانه بینی الکترونیکی ی مبتنی بر ۸ حسگر اکسید فلزی به عنوان یک روش غربالگری سریع، کم‌هزینه و غیرمخرب برای شناسایی و طبقه‌بندی دقیق ۱۵ کلاس مختلف سوخت شامل سوخت‌های خالص، مخلوط‌های دوتایی و مخلوط‌های سه‌تایی گازوئیل، بیودیزل و روغن الکل در نسبت‌های حجمی مشخص است. در این مطالعه، علاوه بر تحلیل ساختار داده‌ها با استفاده از PCA، عملکرد الگوریتم یادگیری ماشین(RBF، ANN و RF) در طبقه‌بندی کلاس‌های ۱۵گانه بر اساس معیارهای دقت، حساسیت و ویژگی ارزیابی و مقایسه می‌شود تا مناسب‌ترین مدل جهت پایش کیفیت سوخت‌های زیستی معرفی گردد.

مواد و روش ها

آماده‌سازی نمونه‌های سوخت

در این پژوهش، سوخت‌های پایه شامل دیزل، بیودیزل و روغن الکل از شرکت تولید الکل رازی تأمین گردید. به منظور ایجاد یک پایگاه داده استاندارد با برچسب‌های قطعی و معلوم جزئیات تیمارهای آزمایشی مورد بررسی، شامل سوخت‌های خالص (دیزل(D)، بیودیزل(B) و روغن الکل(F))، تعداد ۱۵ فرمولاسیون سوختی دوگانه و سه‌گانه با نسبت‌های حجمی مشخص (مطابق با جدول ۲) در شرایط آزمایشگاهی کنترل‌شده تهیه شدند. کلیه فرآیندهای اختلاط با استفاده از پیپت‌های میکرومتری و بالن‌های ژوژه کالیبره‌شده با دقت ±0/01 mL انجام پذیرفت و نمونه‌ها بلافاصله پس از همگن‌سازی در ظروف شیشه‌ای تیره در بسته بدون فضای تبادل نگهداری شدند تا غلظت بخارات فرار تا زمان آزمایش پایدار بماند. از این رو، درصد اختلاط واقعی هر نمونه به صورت مطلق و معین مشخص بوده و به عنوان متغیر هدف در آموزش و ارزیابی الگوریتم‌های یادگیری ماشین مورد استفاده قرار گرفت. مشخصات فیزیکوشیمیایی سوخت‌های پایه شامل نقطه اشتعال، عدد ستان، ارزش حرارتی، نقطه ریزش، ویسکوزیته و چگالی در جدول ۳ ارائه شده است. ویژگی‌های فیزیکوشیمیایی سوخت‌های این پژوهش بر اساس آزمون‌های استاندارد بین‌المللی ASTM و EN مطابق با جدول ۳ اندازه‌گیری و بر اساس برگه‌های مشخصات فنی راستی‌آزمایی شدند.

سامانه بینی الکترونیکی و جمع‌آوری داده ها

سامانه بینی الکترونیکی ی مورد استفاده از چند بخش اصلی تشکیل شده بود: واحد نمونه‌برداری هوا، واحد الکترونیکی پردازش سیگنال، محفظه آرایه حسگرهای نیمه‌هادی اکسید فلزی (MOS)، واحد هیدرونیوماتیک برای کنترل رطوبت و واحد رایانه‌ای برای کنترل و ذخیره‌سازی داده‌ها. حسگرهای انتخاب‌شده (MQ2، MQ3، MQ5، MQ8، MQ9، MQ135، MQ137 و MQ138) به‌صورت موازی در محفظه حسگر نصب شدند. مشخصات تشخیصی این حسگرها در جدول ۴ آمده است.

Table (2) Matrix of laboratory values for different fuel compositions (volume percentage).

Diesel (D)	Biodiesel (B)	Fusel oil (F)	Fuels
100	0	0	D100
0	100	0	B100
0	0	100	F100
95	5	0	D95B5
90	10	0	D90B10
80	20	0	D80B20
70	30	0	D70B30
95	0	5	D95F5
90	0	10	D90F10
80	0	20	D80F20
70	0	30	D70F30
90	5	5	D90B5F5
80	10	10	D80B10F10
70	10	20	D70B10F20
70	20	10	D70B20F10

Table (3) Physicochemical properties of diesel, biodiesel and fusel oil fuels.

Standard test method	Flash point (C)	cetane Number	calorific value (MJ/kg)	Pouring point (C)	viscosity (mm^2/s)	Density (kg/m^3)	Component
ASTM D4052 / EN ISO 12185	65	52.2	42.72	-9	2.5	837	Diesel
ASTM D93 / EN ISO 2719	150	52.2	39.76	3	5.5	894	Biodiesel
ASTM D445 / EN ISO 3104	26	42	29.5	5	3.88	840	Fusel oil

Table (4) Specifications of the selected sensors in the electronic nose system.

Sensor number	Sensor type	Diagnostic gases
1	MQ2	Natural gas, butane, propane, methane, alcohol, hydrogen, smoke
2	MQ3	Alcohol, benzene
3	MQ5	Natural gas, city gas, alcohol, cooking steam, carbon monoxide,
4	MQ8	Natural gas, alcohol, carbon monoxide
5	MQ9	Methane, carbon monoxide, natural gas
6	MQ135	Benzene, ammonia, carbon dioxide, carbon monoxide, alcohol
7	MQ137	Ammonia
8	MQ138	Alcohol, benzene, ammonia, dioxide, carbon monoxide

آزمایش‌ها در قالب یک طرح کاملاً تصادفی با ۱۵ ترکیب سوختی (شامل سوخت‌های خالص دیزل، بیودیزل و روغن الکل و همچنین اختلاط‌های دوتایی و سه‌تایی آن‌ها بر اساس جدول ۲) و با ۱۵ تکرار مستقل برای هر ترکیب انجام شد که در مجموع یک پایگاه داده استاندارد شامل ۲۲۵ نمونه ۱۵×۱۵را تشکیل داد. تمامی داده‌های خام ولتاژ حسگرها پس از پیش‌پردازش و نرمال‌سازی وارد ماتریس تحلیل آماری و مدل‌های یادگیری ماشین شدندفرآیند جمع‌آوری داده در سه مرحله متوالی انجام شد: (۱) تصحیح خط پایه و پاک‌سازی محفظه، (۲) تزریق بخار نمونه و ثبت پاسخ حسگرها، و (۳) پاک‌سازی مجدد محفظه با هوای تمیز. سیگنال‌های حسگرها با فاصله زمانی یک ثانیه از طریق رابط USB به رایانه منتقل و ذخیره شدند. برای هر نمونه، اندازه‌گیری‌های تکراری انجام شد تا مجموعه داده کافی برای تحلیل‌های بعدی فراهم شود.

پیش‌پردازش سیگنال

با توجه به حساسیت حسگرهای اکسید فلزی به بخار آب و تغییرات دمایی، تمامی داده‌برداری‌ها در شرایط محیطی پایدار در دمای 22±1/5 سانتیگراد و رطوبت نسبی 35±5 درصدانجام گردید. واحد هیدرونیوماتیک سامانه مجهز به یک محفظه خشک‌کن حاوی سیلیکاژل در مسیر ورود هوای پاک‌سازی بود تا از ورود رطوبت هوای آزاد به محفظه حسگر ممانعت به عمل آید. چرخه اندازه‌گیری شامل ۳ مرحله بود: پاک‌سازی محفظه به مدت ۱۲۰ ثانیه با هوای خشک و ثبت مقاومت خط پایه (R_0)، تزریق بخار فضای فوقانی نمونه به مدت ۴۰ ثانیه و ثبت مقاومت در حضور گاز(Rs) ، و پاک‌سازی نهایی به مدت ۱۲۰ ثانیه. به‌منظور حذف نویز، نرمال‌سازی و افزایش قدرت تفکیک حسگرها، سه روش رایج پیش‌پردازش شامل روش تفاضلی، نسبی و کسری بررسی شدند. روش کسری به‌دلیل بی‌بعد کردن پاسخ و کارایی مناسب برای سیگنال‌های کوچک و بزرگ، به‌عنوان روش اصلی انتخاب گردید. در این روش، پاسخ نرمال‌شده هر حسگر از رابطه زیر محاسبه می‌شود:

$$Sij = \frac{Rij - R0j}{R0j} \quad (۱)$$

که در آن S_{ij} پاسخ نرمال‌شده حسگر j برای نمونه i، R_{ij} مقاومت لحظه‌ای حسگر و R_{0j} مقاومت خط پایه (پاسخ در هوای تمیز) است. این تبدیل نویز زمینه را کاهش داده و قابلیت تشخیص الگو را بهبود می‌بخشد.

یکی از چالش‌های بنیادین در آموزش مدل‌های یادگیری ماشین با انعطاف‌پذیری بالانظیر RF و MLP بر روی پایگاه‌های داده حسگری با حجم نمونه محدود(n=225) ، ریسک بیش‌برازش و عدم تعمیم‌پذیری مدل به داده‌های خارج از نمونه است. به منظور تضمین استحکام نتایج، کلیه مراحل مقیاس‌پذیری داده‌ها و آموزش مدل‌ها درون یک پایپ‌لاین یکپارچه و با بهره‌گیری از اعتبارسنجی متقاطع ۵-لایه طبقه‌بندی‌شده پیاده‌سازی گردید. در این رویکرد، در هر چرخه، داده‌ها (n=150)صرفاً برای آموزش و (n=75)به عنوان آزمون مستقل خارج از دسترس مدل قرار گرفت.

تحلیل مؤلفه‌های اصلی (PCA)

برای کاهش ابعاد داده‌های چندمتغیره حاصل از آرایه حسگرها و استخراج ویژگی‌های اصلی، از تحلیل مؤلفه‌های اصلی استفاده شد. ابتدا ماتریس کوواریانس داده‌های نرمال‌شده محاسبه گردید:

$$c = \frac{1}{n-1} XX^T \quad (۲)$$

سپس مقادیر ویژه (λ) و بردارهای ویژه (v) از معادله مشخصه $Cv=\lambda v$ استخراج شدند. مؤلفه‌های اصلی به‌ترتیب اهمیت (بر اساس درصد واریانس تبیین‌شده) انتخاب و به‌عنوان ورودی مدل‌های طبقه‌بندی مورد استفاده قرار گرفتند. نمودارهای امتیاز [1] و بارگذاری [2] برای بررسی جداسازی خوشه‌های سوخت ترسیم شدند.

1- score plot
2- loading plot

شبکه عصبی مصنوعی

شبکه عصبی مصنوعی چندلایه پیش‌خور با الگوریتم پس‌انتشار خطا برای مدل‌سازی روابط غیرخطی میان پاسخ حسگرها و کلاس‌های سوخت به‌کار گرفته شد. معماری شبکه شامل لایه ورودی (تعداد نورون برابر با تعداد مؤلفه‌های اصلی منتخب یا تعداد حسگرها)، یک یا دو لایه پنهان با تابع فعال‌سازی سیگموئید یا ReLU و لایه خروجی با تابع Softmax برای طبقه‌بندی چندکلاسه بود. آموزش شبکه با تقسیم داده‌ها به مجموعه‌های آموزش و آزمون ، از یک معماری ساده شامل یک لایه پنهان استفاده شد. برای جلوگیری از واگرایی و برازش بیش از حد وزن‌ها، ضریب منظم‌سازی جریمه نرم دوم [1] با نرخ $\alpha=0/01$ اعمال گردید و الگوریتم بهینه‌ساز Adam با نرخ یادگیری اولیه ۰٫۰۰۱ و سازوکار توقف زودهنگام ([2] با آستانه صبوری ۲۰ دور) به کار گرفته شد تا فرایند بهینه‌سازی بلافاصله پس از عدم بهبود خطای اعتبارسنجی متوقف گردد.

جنگل تصادفی

الگوریتم جنگل تصادفی به‌عنوان یک روش ensemble مبتنی بر درخت تصمیم برای طبقه‌بندی پایدار و مقاوم به بیش‌برازش مورد استفاده قرار گرفت. تعداد درخت‌ها [3] از طریق جستجوی شبکه‌ای بهینه‌سازی شد. در هر درخت، زیرمجموعه‌ای تصادفی از ویژگی‌ها و نمونه‌ها [4] انتخاب گردید. پیش‌بینی نهایی از طریق رأی‌گیری اکثریت به‌دست آمد. اهمیت نسبی هر حسگر [5] بر اساس کاهش ناخالصی جینی محاسبه و گزارش شد. عملکرد مدل با اعتبارسنجی متقابل k-fold (معمولاً $k = ۱۰$) و محاسبه دقت، دقت مثبت [6]، بازخوانی[7] و امتیاز F1 ارزیابی گردید. در مدل جنگل تصادفی، به منظور جلوگیری از بیش‌برازش ناشی از رشد نامحدود شاخه‌ها، عمق بیشینه درختان برابر با ۵ (max_depth=5) تعیین گردید، حداقل نمونه لازم برای انشعاب گره‌ها ۴ نمونه (min_samples_split=4) در نظر گرفته شد و تعداد درختان بر روی ۱۰۰ تنظیم شد.به منظور دستیابی به بهترین عملکرد در مدل‌های طبقه‌بندی ارائه شده، فرآیند تنظیم هایپرپارامترها [8] با دقت بالایی صورت گرفته است. جدول ۵ خلاصه‌ای از تنظیمات کلیدی و مشخصات فنی پیاده‌سازی شده برای دو مدل جنگل تصادفی و شبکه عصبی پرسپترون چندلایه (MLP/ANN) را نمایش می‌دهد. این پارامترها بر اساس استانداردهای متداول یادگیری ماشین و به منظور بهینه‌سازی دقت مدل و همچنین پیشگیری از بیش‌برازش[9] انتخاب شده‌اند.

1- L2 Regularization
2- Early Stopping
3-n-estimators
4- bootstrap
5- feature importance
6- precision
7- recall
8- Hyperparameter Tuning
9- Overfitting

Table (5) Technical specifications and optimized hyperparameter values for the Random Forest and MLP models.

Model Name	Hyperparameter / Technical Specification	Tuned/Set Value in Research	Determination Method / Technical Rationale
Random Forest (RF)	Number of trees (n_estimators)	100	Selected through optimization over a candidate range; the range was not specified.
Random Forest (RF)	Maximum tree depth (max_depth)	10	Limits tree complexity to help control overfitting.
Random Forest (RF)	Minimum samples required to split a node (min_samples_split)	2	Standard default value.
Random Forest (RF)	Minimum samples required at a leaf node (min_samples_leaf)	1	Standard default value.
Random Forest (RF)	Number of features considered at each split (max_features)	sqrt	Standard classification setting; the square root of the total number of features is considered at each split.
Random Forest (RF)	Split-quality criterion (criterion)	Gini	Uses Gini impurity to evaluate candidate splits.
Random Forest (RF)	Random seed (random_state)	42	Ensures reproducibility of experimental results.
MLP / ANN	Network architecture (input-to-output layers)	8-32-16-15	Eight input units, two hidden layers with 32 and 16 units, and 15 output units.
MLP / ANN	Hidden-layer activation function	ReLU	Provides a nonlinear activation and supports efficient convergence.
MLP / ANN	Output-layer activation function	Softmax	Produces a probability distribution over the 15 target classes.
MLP / ANN	Optimizer	Adam	Uses adaptive estimates of first- and second-order moments.
MLP / ANN	Initial learning rate (learning_rate)	0.001	Fixed baseline learning rate selected to support stable convergence.
MLP / ANN	Batch size (batch_size)	16	Small mini-batches are used for training.
MLP / ANN	Maximum number of training epochs (epochs)	200	Sets the upper limit on the number of training epochs.
MLP / ANN	Early stopping	Active	Monitors validation performance to help prevent overfitting.
MLP / ANN	Loss function	Categorical cross-entropy	Measures classification error across multiple classes.
MLP / ANN	Random seed (seed / random_state)	42	Promotes consistent random initialization of weights and reproducible results.

ارزیابی عملکرد و تحلیل آماری

کارایی مدل‌های PCA، ANN و جنگل تصادفی با استفاده از معیارهای استاندارد طبقه‌بندی شامل دقت کلی [1]، ماتریس درهم‌ریختگی، ضریب کاپا مقایسه شد. تمامی تحلیل‌ها در محیط نرم‌افزاری Python (با کتابخانه‌های scikit-learn و TensorFlow/Keras) انجام شد. سطح معنی‌داری آماری α=۰/۰۵ در نظر گرفته شد. داده‌های خام پس از پیش‌پردازش و کاهش بعد، به‌صورت تصادفی تقسیم و مدل‌ها چندین بار اجرا شدند تا میانگین و انحراف معیار عملکرد گزارش گردد. این چارچوب آزمایشی امکان شناسایی و طبقه‌بندی دقیق ترکیبات دیزل–بیودیزل–روغن الکل را به‌صورت غیرمخرب و سریع فراهم می‌آورد و پایه لازم برای توسعه سامانه کنترل کیفیت سوخت‌های زیستی را ایجاد می‌کند. به‌منظور ارزیابی مدل، پس از انجام آزمایش‌ها و طراحی مدل‌های مختلف طبقه‌بندی، دقت و عملکرد بر اساس شاخص‌های آماری نظیر میانگین مربعات خطا – مطابق با معادله (۳) و «ضریب تعیین» سنجیده می‌شوند (Mahmodi et al., 2022). افزون بر این، تحلیل ماتریس درهم‌ریختگی [2] برای طبقه‌بندی انواع مختلف سوخت، منجر به شناسایی چهار حالتِ «منفی حقیقی» (TN)، «منفی کاذب» (FN)، «مثبت حقیقی» (TP) و «مثبت کاذب» (FP) می‌گردد. برای تحلیل عملکرد سیستم طبقه‌بندی، از معیارهای حساسیت، ویژگی (اختصاصیت) و دقت استفاده می‌شود که روابط مربوط به آن‌ها در معادلات (۴ تا ۶) ارائه شده است (Mahmodi et al., 2019).

$$MSE = \frac{1}{n}\sum_{i=0}^{n}(X_i - Y_i)^2 \quad (۳)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (۴)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (۵)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (۶)$$

نتایج و بحث

در این مطالعه، قابلیت تفکیک ۱۵ کلاس سوخت شامل سه سوخت خالصD100 ، B100و F100 و دوازده مخلوط دوجزئی و سه جزئی با استفاده از پاسخ هشت حسگر نیمه هادی اکسید فلزی سری MQ ارزیابی شد. مجموعه داده تلفیقی شامل ۲۲۵ نمونه (۱۵۰نمونه آموزش و ۷۵ نمونه آزمون)،۱۵ تکرار در هر کلاس و هشت ویژگی(MQ9 ، MQ5، MQ8، MQ138، MQ135، MQ137، MQ3و MQ2) بود. هیچ ردیف تکراری مشاهده نشد. کلیه تحلیل‌ها پس از استاندارد سازی انجام شد که در آن میانگین و انحراف معیار فقط از داده آموزش و تست محاسبه شد تا از هرگونه نشت اطلاعات جلوگیری شود. سطح معنی داری در هم آزمونها α=0.05 در نظر گرفته شد و متغیرهای کمی به صورت (میانگین ±انحراف معیار (SD)) گزارش می شوند.

مطابق با نتایج ارائه‌شده در جدول ۶، الگوی توصیفی، زمینه ساز نتایج طبقه بندی بعدی است. نخست، پراکندگی درون کلاسی در حسگرهای MQ138 و MQ137 بسیار زیاد است؛ انحراف معیار در مواردی (مانند ۷/۰۸ برای MQ138 در D80F20) از خود میانگین نیز فراتر میرود و ضریب تغییرات این دو حسگر در سطح تجمیعی بهترتیب ۰/۴۹≈و ۰/۶۵≈است. دوم، بازه‌های میانگین ± انحراف معیار تقریباً همه کلاس ها روی یکدیگر می افتند؛ به بیان دیگر، هیچ حسگری قادر به تفکیک مجزای هیچ کلاسی نیست. سوم، حسگر MQ135 تقریباً بی پاسخ است (1.08±0.171.08 میانگین تجمیعی) و عملاً سهم تفکیک کننده‌ای ندارد. این سه مشاهده، پیشبینی تفکیک پذیری ضعیف در مدل های چندکلاسه را تقویت می کند. نمودارهای میله‌ای ارائه‌شده در شکل ۱ رفتار تفکیکی

1- Accuracy
2- Confusion Matrix

حسگرها را به‌تصویر می‌کشند که در آن حسگرهای (MQ138 و MQ137 پانل‌های d و f) به دلیل واکنش‌پذیری شیمیایی بالا با ترکیبات فرار هیدروکربنی و الکلی، بیشترین دامنه سیگنال را به‌ویژه در ترکیب‌های سه‌گانه حاوی روغن الکل مانند (D70B10F20 و D80F20) ثبت کرده‌اند، درحالی‌که حسگرهایی نظیر (MQ135 و MQ8 پانل‌های c و e) دامنه‌های بسیار فشرده‌تر و یکنواخت‌تری را در تمامی کلاس‌ها نشان می‌دهند. با این وجود، هم‌پوشانی گسترده خطوط خطا (انحراف معیار) میان تیمارها و فقدان روندهای خطی تک‌نواخت در هیچ‌یک از پانل‌ها، نشان می‌دهد که نوسانات تجربی و پاسخ‌های متقاطع درون‌کلاسی مانع از جداسازی قطعی کلاس‌های سوختی بر مبنای ارزیابی تک‌حسگری می‌شوند؛ پدیده‌ای که ضرورت تلفیق سیگنال‌های چندگانه و بهره‌گیری از الگوریتم‌های کمومتریکس چندمتغیره نظیرPCA ، Random Forestو ANN را برای کشف الگوهای تفکیک‌کننده غیرخطی به‌وضوح تأیید می‌نماید.

به منظور ارزیابی قابلیت تفکیک و قدرت تشخیص آرایه حسگرهای گاز در شرایط آزمایشگاهی، از آزمون تحلیل واریانس یک‌طرفه [1] با درجه آزادی (df=14, 210) استفاده گردید (جدول ۷). نتایج تحلیل آماری مؤید آن است که اکثر حسگرها تغییرات معناداری را در پاسخ‌دهی نشان دادند. در این میان، حسگرهایMQ-3 ، Q-9 و MQ-138 با سطح معناداری (p < 0.01)و اندازه اثر (η^2) بالا، بیشترین قدرت تفکیک گازهای هدف را به نمایش گذاشتند.

Table (6) Relative responses of the eight metal-oxide semiconductor (MOS) sensor array across pure diesel, biodiesel, fusel oil, and their ternary/binary blends =15 per class, total n=225); data are presented as mean ±standard deviation.

MQ2	MQ3	MQ137	MQ135	MQ138	MQ8	MQ5	MQ9	Blends
2.54±0.78	1.65±0.24	7.10±5.32	1.04±0.17	9.20±3.93	1.40±0.22	2.41±0.87	4.37±1.25	B100
2.69±1.00	1.56±0.33	6.42±4.51	1.05±0.17	9.12±3.89	1.69±1.01	2.62±1.30	4.24±1.48	D100
2.82±1.15	1.94±0.29	9.95±5.84	1.14±0.21	12.78±6.41	1.66±0.82	2.48±1.25	4.79±1.52	D70B10F20
2.66±0.80	2.00±0.34	9.20±5.47	1.10±0.18	10.49±5.52	1.62±0.78	2.21±0.73	4.51±1.12	D70B20F10
2.57±0.65	1.78±0.25	7.61±4.82	1.07±0.14	10.04±3.92	1.44±0.23	2.22±0.56	4.41±1.13	D70B30
2.69±0.85	1.88±0.27	9.09±5.59	1.10±0.19	11.79±5.79	1.45±0.16	2.33±0.72	4.64±1.20	D70F30
2.55±0.65	1.92±0.28	9.44±6.05	1.09±0.18	11.60±5.72	1.42±0.18	2.20±0.53	4.52±1.09	D80B10F10
2.80±1.00	1.76±0.25	8.68±6.01	1.10±0.22	11.23±6.23	1.53±0.47	2.47±0.79	4.49±1.10	D80B20
2.73±1.02	1.88±0.26	9.27±5.60	1.10±0.16	12.78±7.08	1.44±0.33	2.39±0.96	4.57±1.27	D80F20
2.72±0.81	1.74±0.25	6.79±4.93	1.04±0.16	9.01±3.54	1.45±0.21	2.36±0.81	4.42±1.20	D90B10
2.56±0.67	1.90±0.28	9.03±5.47	1.09±0.19	12.25±6.45	1.38±0.15	2.21±0.54	4.51±1.12	D90B5F5
2.56±0.68	1.85±0.25	8.36±5.55	1.09±0.16	10.00±5.18	1.41±0.18	2.20±0.59	4.53±1.18	D90F10
2.65±0.86	1.72±0.26	7.23±5.04	1.06±0.20	9.05±3.21	1.40±0.25	2.47±0.82	4.36±1.12	D95B5
2.62±0.63	1.82±0.26	8.32±5.12	1.07±0.15	11.12±5.35	1.43±0.23	2.28±0.56	4.43±1.10	D95F5
2.59±1.10	1.69±0.26	7.11±5.50	1.06±0.16	9.69±4.30	1.38±0.29	2.57±1.43	4.37±1.35	F100
2.65±0.84	1.81±0.29	8.24±5.34	1.08±0.17	10.68±5.24	1.47±0.45	2.36±0.86	4.48±1.19	(n=225)

1- One-Way ANOVA

Table (7) One-Way ANOVA Results and Effect Sizes across Gas Sensors.

Sensor	Primary Target Gas	$df_{Between}$	df_{Within}	F-statistic	p-value	Effect Size (η^2)
MQ-2	Flammable gases / Hydrocarbons	14	210	1.84	0.034	0.109
MQ-3	Alcohol vapors	14	210	2.61	0.001	0.148
MQ-5	Natural gas / LPG	14	210	1.42	0.146	0.086
MQ-8	Hydrogen gas (H_2)	14	210	1.15	0.315	0.071
MQ-9	Carbon monoxide / Hydrocarbons	14	210	2.18	0.008	0.127
MQ-135	Air quality / Ammonia / Benzene	14	210	1.76	0.046	0.105
MQ-138	Volatile organic compounds (VOCs)	14	210	2.45	0.003	0.140
MQ-137	Ammonia / Basic organic gases	14	210	1.95	0.022	0.115

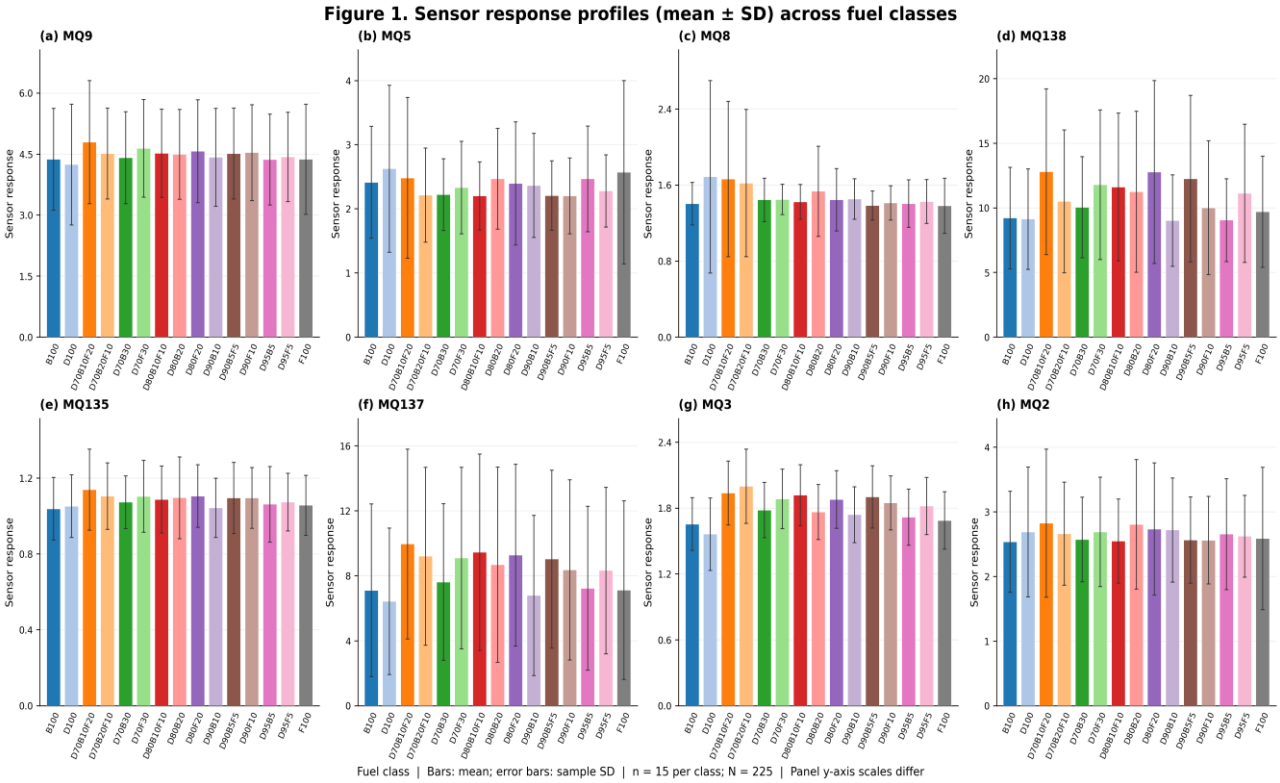


Figure (1)Response profiles of the metal-oxide semiconductor sensor array (MQ9, MQ5, MQ8, MQ138, MQ135, MQ137, MQ3, and MQ2) across 15 fuel classes and blends; bars represent class means and error bars denote sample standard deviations (n=15 per class, total N=225; individual y-axis scales differ across panels).

به منظور کاهش ابعاد داده‌ها و بررسی بصری میزان جداسازی میان کلاس‌های مختلف، تحلیل مؤلفه‌های اصلی (PCA) بر روی ماتریس ویژگی‌های استانداردشده اعمال گردید (جدول ۸). نتایج حاکی از آن است که دو مؤلفه اصلی اول، مجموعاً حدود ٪۷۵ از کل واریانس داده‌ها را تبیین می‌کنند. با این حال، نمودار پراکندگی حاصل از امتیازات این مؤلفه‌ها (شکل ۲) هیچ ساختار جداسازی واضحی را نشان نمی‌دهد؛ به طوری که خوشه‌های مربوط به ۱۵ کلاس مختلف، به شدت با یکدیگر همپوشانی دارند. علاوه بر این، نمودار سنگریزه‌ای (شکل ۳) نیز فاقد یک آرنج مشخص است، که این امر مؤید عدم وجود یک زیرفضای کم‌بعد است که به تنهایی بخش قابل توجهی از واریانس متمایزکننده را بازنمایی کند.

Table (8)Individual explained variance and cumulative variance proportions for the principal components (PC1–PC8) obtained from PCA on standardized e-nose sensor responses using the training dataset (N=150).

Component	Explained Variance (٪) (%)	Cumulative Variance (%)
PC1	47.47	47.47
PC2	27.49	74.96
PC3	12.03	86.99
PC4	5.76	92.75
PC5	3.82	96.57
PC6	2.17	98.74
PC7	0.89	99.63
PC8	0.37	100.00

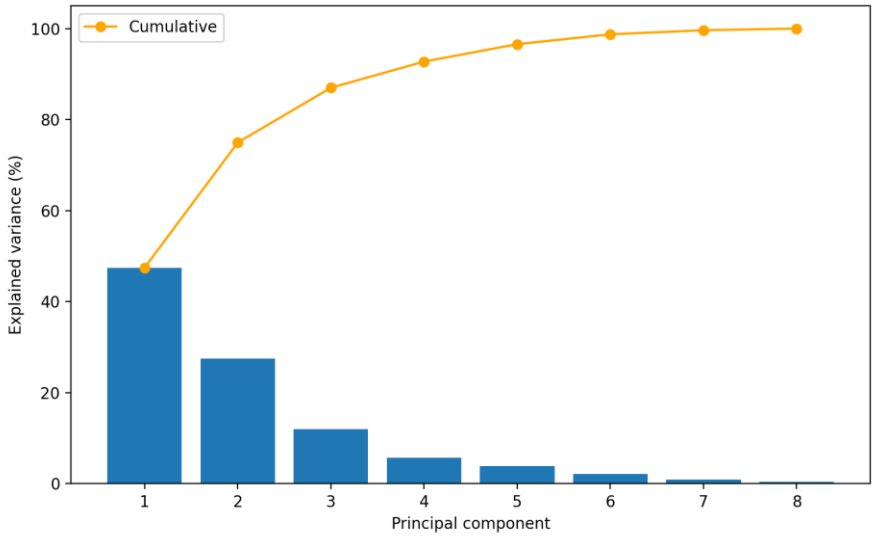


Figure (2)Scree plot showing the explained variance of each principal component (bars) and the cumulative variance (line) obtained from PCA on standardized responses of the eight-sensor array in the training dataset (N=150).

برای هر یک از دو مدل، ابرپارامترها به‌کمک اعتبارسنجی متقابل طبقه‌ای ۵-تایی با پنج تکرار و در حالت shuffle تعیین گردیدند (جدول ۹). در این فرآیند، از مقدار seed برابر ۴۲ برای تکرارپذیری نتایج استفاده شد و معیار انتخاب مدل، میانگین macro-F1 بود. شبکه عصبی مورد استفاده از نوع پرسپترون چندلایه با یک لایه پنهان طراحی شد و به‌منظور حفظ طبقه‌بندی درونی (توزیع کلاس‌ها در هر بخش) در فرآیند آموزش، کسر اعتبارسنجی معادل ۰/۲۰ به‌کار گرفته شد.اگرچه برای ارزیابی عادلانه مدل‌ها و جلوگیری از هرگونه نشت اطلاعات ، کلیه ویژگی‌ها پیش از مدل‌سازی با استفاده از میانگین و انحراف معیار محاسبه‌شده صرفاً از داده‌های آموزش استانداردسازی شدند (تبدیل به نمره z-score). این رویکرد تضمین می‌کند که اطلاعاتی از داده آزمون یا اعتبارسنجی در فرآیند مدل‌سازی نشت نکند.

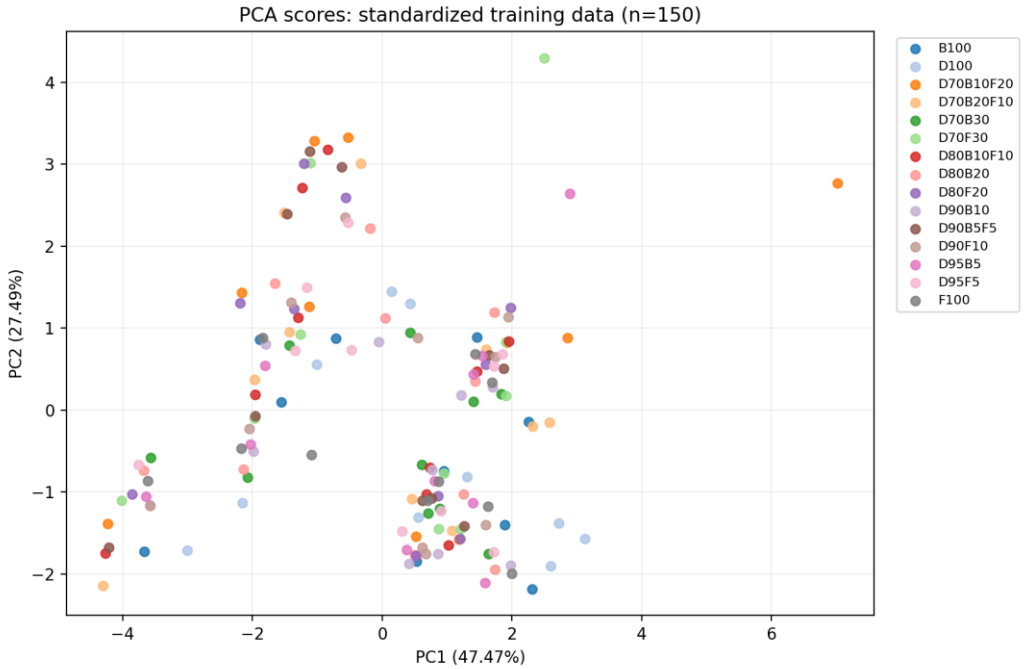


Figure (3) PCA score plot of standardized training data (n=150); the horizontal and vertical axes represent PC1 (47.47%) and PC2 (27.49%) of total variance, and each point corresponds to one experimental replicate of the 15 fuel classes.

Table (9)Optimal hyperparameters of the classifier models following grid search with 5-fold stratified cross-validation.

Optimal hyperparameters	Model
n_estimators = 200, max_depth = None, min_samples_split = 5, max_features = 'sqrt'	**Random Forest**
hidden_layer_sizes = (32,), alpha = 0.001, learning_rate_init = 0.001	**Multilayer Perceptron (MLP)**

بررسی ماتریس درهم‌ریختگی مدل جنگل تصادفی (شکل ۴) نشان‌دهنده عملکرد ضعیف و تعمیم‌ناپذیری ساختار تصمیم‌گیری درخت‌ها در تفکیک دقیق ۱۵ کلاس سوختی است؛ به‌طوری‌که از مجموع ۷۵ نمونه آزمون، تنها ۹ نمونه به‌درستی طبقه‌بندی شده و دقت کلی مدل به ۱۲درصد (۹ از ۷۵) محدود مانده است که تفاوت چندانی با سطح شانس تصادفی ۶٫۶۷درصد ندارد. توزیع خطاهای طبقه‌بندی حاکی از سوگیری شدید پیش‌بینی‌ها به سمت چند کلاس خاص به‌ویژه D100با ۱۸ بار انتساب نادرست، D80B10F10با ۱۳ بار انتساب و D90B10 با ۱۲ بار انتساب است؛ درحالی‌که روی قطر اصلی ماتریس، ۹ کلاس از ۱۵ کلاس (شامل B100 ، D70B20F10، D70B30، D70F30، D80F20، D90B5F5، D90F10، D95F5و F100) به بازخوانی صفر درصد رسیده‌اند. این عدم تقارن و تجمع سنگین خطاها در ستون‌های مربوط به سوخت‌های با پایه دیزل بالا، نمایانگر آن است که شباهت مفرط الگوی انتشار بخارات آلی و همپوشانی بازه‌های واکنشی حسگرها مانع از تشکیل مرزهای جداساز ارتوگونال در جنگل تصادفی شده و مدل به‌دلیل ناتوانی در استخراج ویژگی‌های تفکیک‌کننده، نمونه‌ها را به خوشه‌های پرتکرارتر با واریانس کمتر متمایل می‌سازد.

مقایسه سنجه‌های تفکیکی در جدول ۱۰ نشان می‌دهد که هر دو مدل غیرخطی مبتنی بر یادگیری ماشین (جنگل تصادفی با ۵۰۰ درخت و پرسپترون چندلایه با ساختار ۱۵–۴–۸) در تعمیم‌پذیری به داده‌های آزمون ۱۵ کلاسه دچار افت شدید عملکرد شده‌اند؛ این ناتوانی ساختاری و پراکندگی مقادیر صفر در بیش از نیمی از جدول، بیانگر آن است که فضای ویژگی حاصل از پاسخ حسگرهای MQ به دلیل حساسیت متقاطع بالا و اثر فشرده‌سازی غیرخطی در آمیخته‌های پیچیده زیستی، مرزهای تصمیم بهینه را در طبقه‌بندهای ناظر مخدوش ساخته و نیاز مبرم به روش‌های استخراج ویژگی‌های غیرپیوسته یا تقویت معماری آرایه حسگری را آشکار می‌سازد.

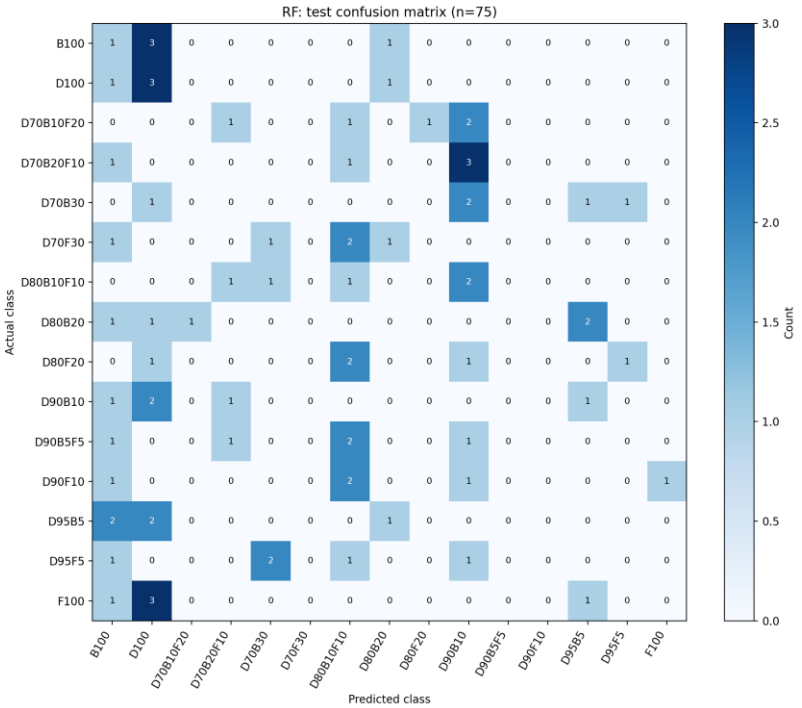
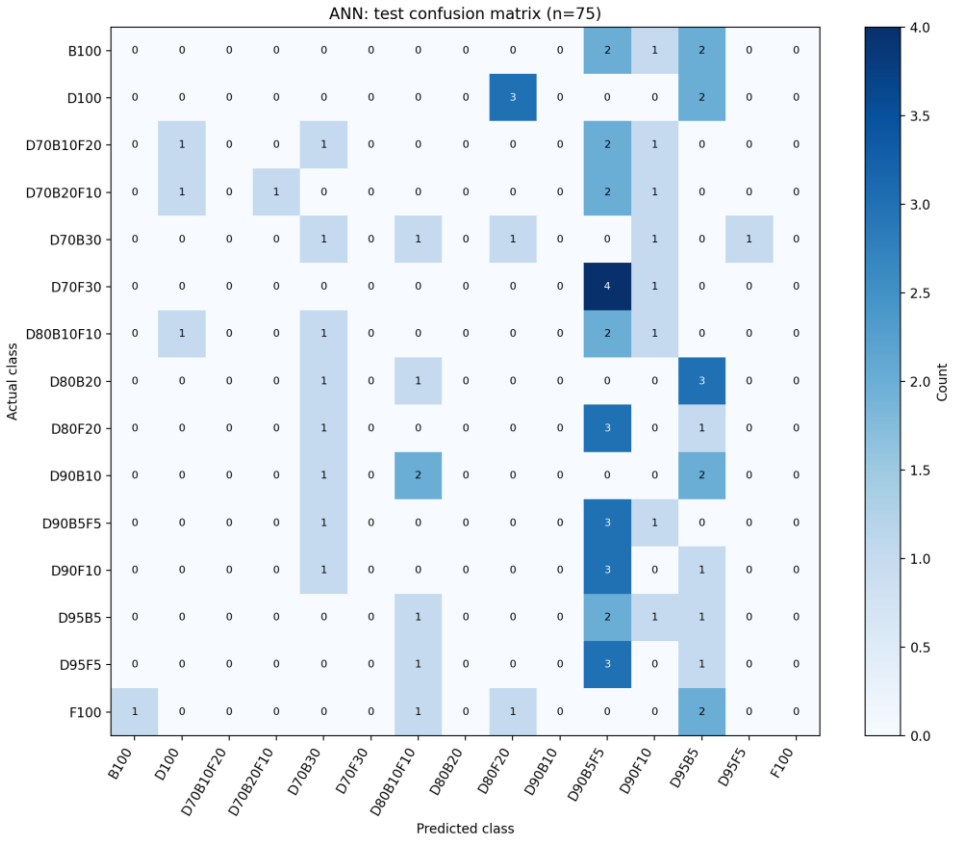


Figure (4) Confusion matrix heatmap of the (RF) classifier on the independent test datasetfor the 15-class classification; rows represent actual fuel classes and columns denote predicted classes, with cell values indicating sample counts.

Table (10) Per-class classification performance metrics (Precision, Recall, and F1-score) of Random Forest (RF) and Artificial Neural Network (ANN) classifiers evaluated on the independent test dataset (n=75, 5 replicates per class).

ANN-F1	ANN-R	ANN-P	RF-F1	RF-R	RF-P	Labels of the dataset
0.0000	0.0000	0.0000	0.1176	0.2000	0.0833	B100
0.0000	0.0000	0.0000	0.2857	0.6000	0.1875	D100
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	D70B10F20
0.3333	0.2000	1.0000	0.0000	0.0000	0.0000	D70B20F10
0.1538	0.2000	0.1250	0.0000	0.0000	0.0000	D70B30
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	D70F30
0.0000	0.0000	0.0000	0.1176	0.2000	0.0833	D80B10F10
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	D80B20
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	D80F20
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	D90B10
0.1935	0.6000	0.1154	0.0000	0.0000	0.0000	D90B5F5
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	D90F10
0.1000	0.2000	0.0667	0.0000	0.0000	0.0000	D95B5
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	D95F5
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	F100



Figure(5)Confusion matrix heatmap of the Artificial Neural Network (ANN/MLP with 8-4-15 architecture) on the independent test dataset for the 15-class fuel classification; rows represent actual fuel classes and columns denote predicted classes, with cell values indicating sample counts.

بررسی ماتریس درهم‌ریختگی مدل شبکه عصبی مصنوعی پرسپترون چندلایه (شکل ۵) بیانگر تداوم چالش تعمیم‌پذیری و عدم همگرایی مرزهای تصمیم در فضای ۱۵ کلاسه است؛ به‌نحوی‌که از مجموع ۷۵ نمونه آزمون، تنها ۱۰ نمونه روی قطر اصلی ماتریس به‌درستی طبقه‌بندی شده و دقت کلی مدل به ۳۳٫درصد۱۳ (۱۰ از ۷۵) محدود مانده است. الگوی توزیع خطاهای طبقه‌بندی نشان‌دهنده جذب شدید و غیرطبیعی پیش‌بینی‌ها توسط تعداد اندکی از کلاس‌های خروجی است؛ به‌طوری‌که ستون کلاس D90B5F5 به‌تنهایی ۲۶ نمونه شامل ۳ پیش‌بینی درست و ۲۳ نمونه متعلق به سایر کلاس‌ها نظیر D70F30 با ۴ نمونه و D80F20 و D90F10 هرکدام با ۳ نمونه و ستون کلاس D95B5 تعداد ۱۵ نمونه را به خود جذب کرده‌اند. از سوی دیگر، ۱۰ کلاس از ۱۵ کلاس شامل B100 ، D100، D70B10F20، D70F30، D80B10F10، D80B20، D80F20، D90B10، D90F10و F100 به بازخوانی صفر رسیده‌اند. این پدیده تجمیع خطا در گره‌های خروجی خاص نشان می‌دهد که به دلیل هم‌پوشانی شدید الگوهای ولتاژی آرایه حسگرهای MQ و فقدان ویژگی‌های تفکیک‌کننده یکتا در غلظت‌های ترکیبی، شبکه در بهینه‌سازی توابع وزن‌های لایه میانی به کمینه‌های محلی غلط افتاده و رفتاری مشابه یک طبقه‌بند چندکلاسه با سوگیری بالا از خود بروز داده است.

به‌منظور ارزیابی سهم نسبی هریک از حسگرهای گاز در فرآیند تفکیک کلاس‌های سوختی، تحلیل اهمیت ویژگی‌ها با استفاده از شاخص میانگین کاهش ناخالصی در ساختار جنگل تصادفی اجرا شد (شکل ۶). نتایج حاصل نشان‌دهنده توزیع نسبتاً یکنواخت اهمیت بین هشت حسگر به‌کاررفته است، به‌طوری‌که مقدار شاخص بین ۰/۱۰ تا ۰/۱۵ متغیر است. حسگر MQ8با مقدار ۰/۱۴۶۵ به‌عنوان تاثیرگذارترین ویژگی در کاهش ناخالصی گره‌های درخت تصمیم شناسایی شد و پس از آن حسگرهای MQ3 (با ۰/۱۳۵۴) و MQ138 (با ۰/۱۴۳۵) در جایگاه‌های بعدی قرار گرفتند. کمترین نقش تفکیکی در این آرایه به حسگرهای MQ137 و MQ9 اختصاص یافته است. این توزیع متوازن اهمیت، در کنار فقدان یک حسگر مسلط ، مؤید این فرضیه است که تمامی مؤلفه‌های آرایه بینی الکترونیکی برای تشکیل یک فضای ویژگی چندبعدی و درهم‌تنیده جهت تلاش برای تفکیک سوخت‌های زیستی ضروری هستند، اما به تنهایی هیچ‌کدام قادر به ارائه یک سیگنال تفکیک‌کننده قوی برای جداسازی کلاس‌های پیچیده مانند B100 در برابر ترکیبات (D-B-F) نیستند.

Table (11) Importance coefficients of sensor characteristics.

Importance	Type of the sensor
0.1465	MQ8
0.1354	MQ3
0.1345	MQ138
0.1238	MQ2
0.1237	MQ135
0.1212	MQ5
0.1142	MQ137
0.1008	MQ9

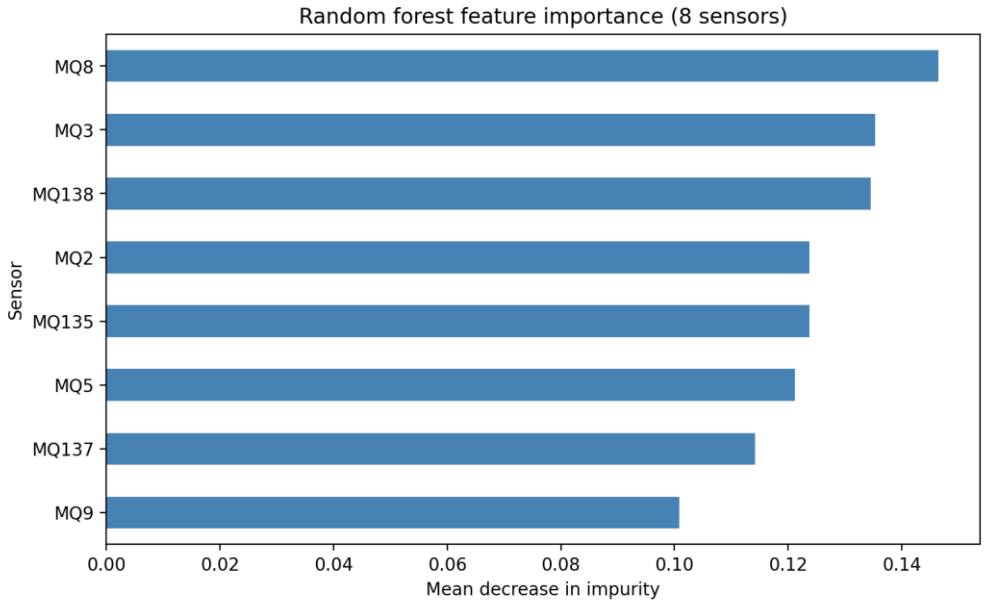


Figure (6) Feature importance ranking of the eight gas sensors within the E-nose array based on the Mean Decrease in Impurity (MDI) index in the Random Forest model.

Table (12)Overall performance comparison of Random Forest (RF) and Multilayer Perceptron (ANN/MLP) classifiers based on Overall Accuracy and Macro F1-score evaluated on the independent test dataset (n=75) under two sensor array configurations: selected subset (4 sensors: MQ3, MQ9, MQ137, MQ138) versus full array (8 sensors).

F1 (eight sensors)	Accuracy (eight sensors)	F1 (four sensors)	Accuracy (four sensors)	Model
0/0347	0/0667	0/0681	0/1067	Random forest
0/0520	0/008	0/008	0/0533	MLP

ارزیابی تطبیقی کارایی آرایه حسگری در دو پیکربندی کاهش‌یافته(۴ حسگر) MQ3, MQ9, MQ137, MQ138 و آرایه کامل (۸ حسگر) در جدول ۷ ارائه شده است. نتایج نشان می‌دهد که افزایش ابعاد فضای ویژگی از ۴ به ۸ حسگر لزوماً منجر به ارتقای تعمیم‌پذیری مدل‌ها در آزمون مستقل ۱۵-کلاسه نگردیده است؛ در مقابل، در مدل پرسپترون چندلایه، با وجود بهبود نسبی صحت این رفتار واگرا در دو الگوریتم یادگیری ماشین اثبات می‌کند که چالش اصلی سیستم بویایی الکترونیکی در تفکیک آمیخته‌های سوختی پیچیده، ناشی از کمبود تعداد حسگرها نیست، بلکه ریشه در هم‌پوشانی ماهوی سیگنال‌های شیمیایی و فقدان ویژگی‌های تفکیک‌پذیر غیرخطی در پاسخ استاتیک سنسورهای نیمه‌رسانای اکسید فلزی دارد.

یافته محوری و بنیادین پژوهش حاضر یعنی عدم کفایت آرایه هشت‌حسگری اکسید فلزی نیمه‌رسانا در تفکیک ۱۵ کلاس آمیخته‌های سوختی دیزل، بیودیزل و روغن الکل فراتر از آستانه تصادف آماری در شرایط قرائت استاتیک/حالت‌پایا در وهله اول در تضاد ظاهری با بخش عمده‌ای از ادبیات تجربی حوزه بویایی الکترونیکی قرار می‌گیرد. با این حال، کالبدشکافی دقیق متریک‌های آماری نشان می‌دهد که این رفتار، نه یک خطای محاسباتی، بلکه بازتابی عینی از فیزیک حسگرها و ماهیت شیمیایی نمونه‌هاست. صحت آزمون ثبت‌شده برای هر دو طبقه‌بند غیرخطی، یعنی جنگل تصادفی و پرسپترون چندلایه، عملاً بر سطح احتمال تصادفی یک مسئله ۱۵کلاسه منطبق است. همچنین، ضریب کاپای کوهن که توافق فراتر از شانس را می‌سنجد،

در هر دو مدل به‌شدت به صفر میل می‌کند و آزمون مقایسه‌ای مک‌نمار نیز هیچ‌گونه برتری معنادار آماری را میان عملکرد این دو الگوریتم اثبات نمی‌کند. این همگرایی منفی و بروز بیش‌برازش فاحش در مدل جنگل تصادفی، نشان می‌دهد که پیچیدگی معماری یادگیری ماشین قادر به جبران «فقر اطلاعات تفکیک‌کننده» در فضای ورودی نیست. بر این مبنا، این ناهم‌خوانی در چارچوب سه محور ساختاری و متدولوژیک به بحث گذاشته می‌شود.

در سال‌های اخیر، پژوهش‌های متعددی کارایی سامانه بویایی الکترونیکی را در شناسایی سوخت‌های زیستی گزارش کرده‌اند؛ از جمله محمودی و همکاران (Mahmodi et al., 2022) که با تلفیق بینی الکترونیکی، ANN و روش سطح پاسخ (RSM) آمیخته‌های بیودیزل و دیزل را با دقت بالا تفکیک کردند، بصیرت پورو همکاران (Basiratpour et al., 2025) که الگوریتمی جهت پایش نسبت‌های اختلاط بیودیزل–روغن الکل توسعه دادند، وکاکایی و همکاران (Kakaei et al., 2025) که به شناسایی ناخالصی‌ها در فرآورده‌های فسیلی پرداختند، واگرایی آشکار نتایج مطالعه حاضر با این مراجع را باید در سه تفاوت ساختاری جست‌وجو کرد. در مطالعات موفق، الگوهای طبقه‌بندی معمولاً بر پایه متغیرهای دینامیکِ منحنی پاسخ نظیر ثابت زمانی ، شیب فاز جذب و واجذب ، انتگرال سطح زیر منحنی و زمان رسیدن به بیشینه بنا شده‌اند؛ درحالی‌که در پروتکل حاضر صرفاً مقادیر ولتاژ نقطه‌ای در فاز تعادل استاتیک در دسترس بوده است.

از دیدگاه فیزیکوشیمیایی حسگرهای MOS، تفکیک‌پذیری واقعی گازها ناشی از مقدار تعادلی مقاومت سطحی نیست، بلکه تابع مستقیم سینتیک انتقال بار و نرخ واکنش‌های ردوکس گونه‌های فرّار با اکسیژن‌های یونیزه‌شده سطحی است. ترکیبات آلی فرّار نرخ نفوذ و ضرایب جذب ناهمسانی دارند که تشکیل‌دهنده «امضای زمانی منحصربه‌فرد» آن‌هاست. تنزل سیگنال پیوسته به یک مقدار عددی اسکالر در فاز پایا، عملاً این ردپای سینتیکی را مخدوش ساخته و اطلاعات افتراقی سیستم را به‌شدت مستهلک می‌کند. این مسئله مشابه استدلال لیو و همکاران (Liu et al., 2026) است که نشان دادند غنای اطلاعاتی بازسازی‌های طیفی و زمانی، پیش‌شرط قطعی موفقیت شبکه‌های عصبی در تفکیک آمیخته‌های پیچیده الکل–دیزل است؛ بنابراین غنای مودالیته داده بر پیچیدگی معماری مدل تقدم دارد.

حسگرهای تجاری سری MQ به دلیل ساختار مبتنی بر لایه اکسید قلع ، گزینش‌پذیری ضعیفی نسبت به زنجیره‌های هیدروکربنی نزدیک به هم دارند. آمیخته‌های سوختی دیزل (عمدتاً آلکان‌های سنگین)، بیودیزل (استرهای متیل اسید چرب با قطبیت اندک) و روغن الکل(الکل‌های سنگین نظیر ایزوآمیل‌الکل) اشتراک بالایی در فضای بخار بالای نمونه دارند. واکنش سطحی این گازها با سنسور، هدایت الکتریکی مشابهی القا می‌کند که حاصل آن هم‌خطی شدید بردارهای پاسخ هشت حسگر است؛ به‌گونه‌ای که بعد مؤثر اطلاعاتی آرایه بسیار کمتر از بعد اسمی آن است.

اثبات ریاضی این پدیده در توزیع یکنواخت اهمیت ویژگی‌ها در جنگل تصادفی مشهود است؛ جایی که دامنه تغییرات شاخص ناخالصی بین حسگر مسلط (MQ8) و ضعیف‌ترین حسگر (MQ9) بسیار فشرده بوده و هیچ حسگری نقش تعیین‌کننده ایفا نمی‌کند. این وضعیت با پژوهش‌هایی که در آن‌ها فاصله فیزیکوشیمیایی کلاس‌ها بسیار عمیق بوده نظیر تمایز روغن‌های خوراکی بکر و صنعتی درکرمی و همکاران (Karami et al., 2020) یا ردیابی تقلب روغن زیتون درکراتشو و همکاران (Cerrato Oliveros et al., 2002) تفاوت بنیادین دارد؛ در آن دست مطالعات، حضور نشانگرهای فرّار اختصاصی مانع از درهم‌تنیدگی کامل ابعاد ویژگی می‌شود.

پیکربندی مسئله با ۱۵ کلاس تفکیکی و ۱۵ تکرار در هر کلاس (مجموعاً ۲۲۵ نمونه) چالش دیگری است. با تخصیص ساختاریافته دو–سوم نمونه‌ها به آموزش (n=150) و یک–سوم به آزمون مستقل (n=75و ۵ تکرار به‌ازای هر کلاس)، سیستم

از نظر تئوری یادگیری آماری در شرایط کم‌توان قرار می‌گیرد. در یک مجموعه آزمون ۵تایی به‌ازای هر کلاس، وقوع حتی یک طبقه‌بندی اشتباه، ضریب بازخوانی کلاس را ۲۰ واحد درصد کاهش می‌دهد.

به همین علت، فاصله اطمینان ۹۵درصد ویلسون برای صحت کلی هر دو مدل، آستانه شانس (۰٫۰۶۶۷) را پوشش می‌دهد.افزون بر این، رشد ترکیباتی مرزهای تصمیم در فضای ۱۵-کلاسه با توزیع‌های فوق‌هم‌پوشان، هندسه جداسازی هایپرپلین‌ها را تخریب می‌کند. در مقابل، مطالعات متمرکز بر رگرسیون پیوسته مشخصه‌های سوختی (Doicin et al., 2025; Shirley et al., 2027) یا کلاس‌بندی‌های کلان، از پیوستگی ریاضی متغیر پاسخ بهره می‌برند و با پدیده فروپاشی فضایی کلاس‌ها مواجه نمی‌شوند.

واگرایی مشاهده‌شده در این تحقیق، رسالت متدولوژیک انتشار صادقانه نتایج منفی را در سامانه‌هاپررنگ می‌سازد. پدیده سوگیری انتشار به نفع مدل‌های با دقت بالا، نوعی برآورد بیش از حد و کاذب از توانمندی آرایه‌های حسگری ارزان‌قیمت در متون علمی پدید آورده است. گزارش مستند مواردی که یک فناوری در آن‌ها پاسخگوی هدف نیست، به‌اندازه موفقیت‌ها واجد ارزش علمی است، زیرا مرزهای عملیاتی سخت‌افزار را تعیین کرده و مانع اتلاف منابع پژوهشی می‌گردد.

کارایی و توان تفکیک ابزارهای هوشمند پیوند ناگسستنی با اصالت فیزیکوشیمیایی و غنای اطلاعاتی داده‌های ورودی دارد؛ به‌گونه‌ای که در پایش ترکیبات روغنی و بهینه‌سازی فرایندهای ترانس‌استریفیکاسیون، تکیه صرف بر الگوریتم‌ها بدون فراهم‌سازی بردار ویژگی بهینه، ضامن دستیابی به دقت بالا نخواهد بود. همچنین در تحلیل‌های کمومتریکس بر داده‌های طیفی و کاربردهای بویایی الکترونیکی در محصولات کشاورزی اثبات شده است که طبقه‌بندهای غیرخطی تنها زمانی به مزیت رقابتی دست می‌یابند که سیگنال خام ورودی از نسبت تفکیک به نویز بالایی برخوردار باشد. از این رو، افت عملکرد طبقه‌بندی در پژوهش حاضر نشان‌دهنده نقص ذاتی الگوریتم‌های جنگل تصادفی (RF) و شبکه عصبی مصنوعی (ANN) نیست، بلکه آشکارساز یک «سقف فیزیکی» ناشی از محدودیت ذاتی ترکیب حسگرهای نیمه‌رسانای اکسید فلزی (MOS) عمومی، استخراج ویژگی‌های استاتیک در حالت پایدار و پیچیدگی ذاتی ماتریس تفکیک ۱۵-کلاسه است (Adelkhani and Daneshkhah, 2025; Balabin and Lomakina, 2011; Kesharvani et al., 2024;) .(Lashgari and Mohamadi-Gol, 2016

به‌منظور عبور از محدودیت‌های آشکارشده در این تحقیق، راهکارهای روش‌مند و سازگار با چارچوب حسگری ارائه می‌شود:

۱. استخراج ویژگی‌های گذرای: پیاده‌سازی سخت‌افزارهای نمونه‌برداری پیوسته جهت ثبت کل منحنی پاسخ گاز و استخراج شاخص‌های مشتق اول، دوم و ضرایب بسط فوریه یا موجک، بعد زمانی از دست رفته را احیا کرده و هم‌راستا با روش‌شناسی Mahmodi و همکاران (۲۰۲۲) و Liu و همکاران (۲۰۲۶)، توان تمایز سینتیکی بخارات فرّار را بازیابی می‌کند.

۲. پالایش و کاهش گزینشی آرایه حسگری: همان‌طور که Rasekh و همکاران (۲۰۲۱) در بهینه‌سازی سیستم‌های بینی الکترونیکی نشان داده‌اند، افزایش بی‌رویه تعداد حسگرها لزوماً منجر به بهبود خروجی نمی‌شود و می‌تواند از طریق پدیده «نفرین ابعاد»، عملکرد را تضعیف کند. تمرکز بر زیرمجموعه گزینش‌پذیرتر (حسگرهای MQ8، MQ3 و MQ138) هزینه ساخت را کاسته و نویز همبستگی متقاطع را مهار می‌سازد.

۳. تقلیل مقیاس به خوشه‌بندی‌های کلان فیزیکی: بازتعریف مسئله از ۱۵ کلاس ریزدانه به چند رده ماکروسکوپیک معنادار (نظیر تفکیک سوخت‌های پایه D100، B100 و F100، یا ردیابی آستانه‌ای نفوذ سوخت‌های زیستی در دیزل) تعمیم‌پذیری پایدارتری را محقق می‌سازد.

نتیجه‌گیری

در این پژوهش، قابلیت کاربردی و مرزهای عملیاتی یک آرایه بویایی الکترونیکی مبتنی بر هشت حسگر گاز نیمه‌رسانای اکسید فلزی (MOS سری MQ) در تفکیک ۱۵ کلاس متشکل از سوخت‌های خالص (دیزل D100، بیودیزل B100 و روغن الکل F100) و آمیخته‌های دوتایی و سه‌تایی آن‌ها در شرایط پاسخ استاتیک (حالت‌پایا) به‌صورت نظام‌مند مورد ارزیابی قرار گرفت. تحلیل ساختار داده‌های خام و ارزیابی پروتکل‌های یادگیری ماشین به نتایج کلیدی زیر انجامید:

مدل‌های غیرخطی جنگل تصادفی و شبکه عصبی مصنوعی پرسپترون چندلایه (ANN/MLP با ساختار ۸-۴-۱۵) روی مجموعه داده آزمون مستقل (n=75) به‌ترتیب به صحت کلی ۶٫۶۷درصد و ۸درصد (در پیکربندی ۸ حسگر) و ۱۰٫۶۷درصد و ۵٫۳۳درصد (در پیکربندی ۴ حسگر) دست یافتند. این مقادیر عملکردی از نظر آماری با آستانه تفکیک تصادفی هم‌پوشان بوده و فواصل اطمینان ۹۵درصد ویلسون و آزمون مک‌نمار ($p = ۱$) مؤید انطباق عملکرد طبقه‌بندها با حد شانس هستند. تحلیل مؤلفه‌های اصلی (PCA) نشان داد که با وجود تبیین ۷۴٫۹۶درصد از واریانس کل توسط دو مؤلفه نخست (PC1 و PC2)، خوشه‌های مربوط به ترکیبات سوختی در فضای نگاشت دوبعدی تداخل ساختاری شدیدی دارند. همچنین، ارزیابی اهمیت ویژگی‌ها در مدل جنگل تصادفی بیانگر توزیع یکنواخت شاخص کاهش ناخالصی (MDI بین ۰٫۱۰۰۸ تا ۰٫۱۴۶۵) میان تمامی حسگرها بود؛ امری که نشان‌دهنده حساسیت متقاطع بالا ، هم‌خطی شدید سیگنال‌ها بر پایه لایه فعال $SnO_۲$ و فقدان یک حسگر مرجع با گزینش‌پذیری اختصاصی برای بخارات آمیخته‌های سوختی است. مقایسه دو پیکربندی ۴ و ۸ حسگری آشکار ساخت که افزایش ابعاد بردار ورودی، به‌ویژه در مدل جنگل تصادفی، به افت صحت کلی و کاهش F1 کلان منجر می‌شود. این یافته اثبات می‌کند که در شرایط فقدان نشانگرهای شیمیایی تفکیک‌پذیر در فاز پایا، افزودن حسگرهای همبسته صرفاً ورود نویز محاسباتی و خطر بیش‌برازش را تشدید می‌نماید. در مجموع، این مطالعه به‌روشنی مرزهای متدولوژیک و محدودیت‌های فیزیکی استفاده از قرائت‌های نقطه‌ای/حالت‌پایا را در آرایه‌های حسگری کم‌هزینه ترسیم کرده و نشان داد که موفقیت سیستم‌های بینی الکترونیکی در متمایزسازی ریزدانه آمیخته‌های سوختی پیچیده، مشروط به غنای مودالیته سیگنال پایه و فراتر از قابلیت جبران‌سازی توسط الگوریتم‌های طبقه‌بندی است.

References

Adelkhani, A., & Daneshkhah, E. (2025). Investigation of engine lubrication oil quality using a support vector machine and electronic nose. Machines, 13(2), Article 121. https://doi.org/10.3390/machines13020121

Balabin, R. M., & Lomakina, E. I. (2011). Support vector machine regression (SVR/LS-SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data. Analyst, 136(8), 1703–1712. https://doi.org/10.1039/c0an00387e

Basiratpour, A., Safieddin Ardebili, S. M., Mostafaei, M., & Zakidizaji, H. (2025). Development of an algorithm to detect ratio of the blend of biodiesel-fusel oil with diesel fuel using an e-nose system. Biosystems Engineering and Renewable Energies, 1(2), 115–120. https://doi.org/10.22069/bere.2025.24064.1029

Beurey, C., Gozlan, B., Carré, M., Bacquart, T., Morris, A., Moore, N., Arrhenius, K., Meuzelaar, H., Persijn, S., Rojo, A., & Murugan, A. (2021). Review and survey of methods for analysis of impurities in hydrogen for fuel cell vehicles according to ISO 14687:2019. Frontiers in Energy Research, 8, 1–20. https://doi.org/10.3389/fenrg.2020.615149

Cabaneros Lopez, P., Feldman, H., Mauricio-Iglesias, M., Junicke, H., Huusom, J. K., & Gernaey, K. V. (2019). Benchmarking real-time monitoring strategies for ethanol production from lignocellulosic biomass. Biomass and Bioenergy, 127, Article 105296. https://doi.org/10.1016/j.biombioe.2019.105296

Calam, A., Içingür, Y., Solmaz, H., & Yamk, H. (2015). A comparison of engine performance and the emission of fusel oil and gasoline mixtures at different ignition timings. International Journal of Green Energy, 12(8), 767–772. https://doi.org/10.1080/15435075.2013.849256

Cerrato Oliveros, M. C., Pérez Pavón, J. L., García Pinto, C., Fernández Laespada, M. E., Moreno Cordero, B., & Forina, M. (2002). Electronic nose based on metal oxide semiconductor sensors as a fast alternative for the detection of adulteration of virgin olive oils. Analytica Chimica Acta, 459(2), 219–228. https://doi.org/10.1016/S0003-2670(02)00119-8

Cinar, C., Uyumaz, A., Solmaz, H., Sahin, F., Polat, S., & Yilmaz, E. (2015). Effects of intake air temperature on combustion, performance and emission characteristics of a HCCI engine fueled with the blends of 20% n-heptane and 80% isooctane fuels. Fuel Processing Technology, 130, 275–281. https://doi.org/10.1016/j.fuproc.2014.10.026

Doicin, B., Duşescu-Vasile, C. M., Onuţu, I., Băjan, M., Bomboş, D., & Vasilievici, G. (2025). Estimation of properties of petrodiesel–biodiesel mixtures using an artificial neural network. Processes, 13(6), Article 1769. https://doi.org/10.3390/pr13061769

Hassan Pour, A., Safieddin Ardebili, S. M., & Sheikhdavoodi, M. J. (2018). Multi-objective optimization of diesel engine performance and emissions fueled with diesel-biodiesel-fusel oil blends using response surface method. Environmental Science and Pollution Research, 25(35), 35429–35439. https://doi.org/10.1007/s11356-018-3459-z

Hu, G., Ahmed, M., & L'Abbé, M. R. (2023). Natural language processing and machine learning approaches for food categorization and nutrition quality prediction compared with traditional methods. American Journal of Clinical Nutrition, 117(3), 553–563. https://doi.org/10.1016/j.ajcnut.2022.11.022

Kakaei, A., Mostafaei, M., & Naderloo, L. (2025). Identification and classification of adulteration in some fossil fuel products using an electronic nose. Fuel, 396, Article 135405. https://doi.org/10.1016/j.fuel.2025.135405

Karami, H., Rasekh, M., & Mirzaee-Ghaleh, E. (2020). Application of the e-nose machine system to detect adulterations in mixed edible oils using chemometrics methods. Journal of Food Processing and Preservation, 44(9), Article e14696. https://doi.org/10.1111/jfpp.14696

Kesharvani, S., Katre, S., Pandey, S., Dwivedi, G., Verma, T. N., & Baredar, P. (2024). Optimizing biodiesel production from Karanja and algae oil with nano catalyst: RSM and ANN approach. Energy Engineering, 121(9), 2363–2388.

Kelanic, R. A. (2016). The petroleum paradox: Oil, coercive vulnerability, and great power behavior. Security Studies, 25(2), 181–213. https://doi.org/10.1080/09636412.2016.1171966

Khorramifar, A., Karami, H., Wilson, A. D., Sayyah, A. H. A., Shuba, A., & Lozano, J. (2022). Grape cultivar identification and classification by machine olfaction analysis of leaf volatiles. Chemosensors, 10(4), Article 125. https://doi.org/10.3390/chemosensors10040125

Lashgari, M., & Mohammadigol, R. (2016). Discrimination of Golab apple storage time using acoustic impulse response and LDA and QDA discriminant analysis techniques. Iran Agricultural Research, 35(2), 65–70. https://doi.org/10.22099/iar.2016.3799

Li, Y., Tang, W., Chen, Y., Liu, J., & Lee, C. F. (2019). Potential of acetone-butanol-ethanol (ABE) as a biofuel. Fuel, 242, 673–686. https://doi.org/10.1016/j.fuel.2019.01.063

Liu, S., Zhang, X., Wang, L., Chen, X., Liu, X., & Fang, L. (2026). NIR spectrogram reconstruction—augmented CNN for high-accuracy analysis of alcohol-diesel blend. Chemometrics and Intelligent Laboratory Systems, 273, Article 105687. https://doi.org/10.1016/j.chemolab.2026.105687

Loutfi, A., Coradeschi, S., Mani, G. K., Shankar, P., & Rayappan, J. B. B. (2015). Electronic noses for food quality: A review. Journal of Food Engineering, 144, 103–111. https://doi.org/10.1016/j.jfoodeng.2014.07.019

Mahmodi, K., Mostafaei, M., & Mirzaee-Ghaleh, E. (2019). Detection and classification of diesel-biodiesel blends by LDA, QDA and SVM approaches using an electronic nose. Fuel, 258, Article 116114. https://doi.org/10.1016/j.fuel.2019.116114

Mahmodi, K., Mostafaei, M., & Mirzaee-Ghaleh, E. (2022). Detecting the different blends of diesel and biodiesel fuels using electronic nose machine coupled ANN and RSM methods. Sustainable Energy Technologies and Assessments, 51, Article 101914. https://doi.org/10.1016/j.seta.2021.101914

Panoutsou, C., Germer, S., Karka, P., Papadokostantakis, S., Kroyan, Y., Wojcieszyk, M., Maniatis, K., Marchand, P., & Landalv, I. (2021). Advanced biofuels to decarbonise European transport by 2030: Markets, challenges, and policies that impact their successful market uptake. Energy Strategy Reviews, 34, Article 100633. https://doi.org/10.1016/j.esr.2021.100633

Rasekh, M., Karami, H., Wilson, A. D., & Gancarz, M. (2021). Classification and identification of essential oils from herbs and fruits based on a MOS electronic-nose technology. Chemosensors, 9(6), Article 142. https://doi.org/10.3390/chemosensors9060142

Ren, J., & Sovacool, B. K. (2015). Prioritizing low-carbon energy sources to enhance China's energy security. Energy Conversion and Management, 92, 129–136. https://doi.org/10.1016/j.enconman.2014.12.044

Safieddin Ardebili, S. M., Solmaz, H., İpci, D., Calam, A., & Mostafaei, M. (2020a). A review on higher alcohol of fusel oil as a renewable fuel for internal combustion engines: Applications, challenges, and global potential. Fuel, 279, Article 118516. https://doi.org/10.1016/j.fuel.2020.118516

Safieddin Ardebili, S. M., Taghipoor, A., Solmaz, H., & Mostafaei, M. (2020b). The effect of nano-biochar on the performance and emissions of a diesel engine fueled with fusel oil-diesel fuel. Fuel, 268, Article 117356. https://doi.org/10.1016/j.fuel.2020.117356

Shahbeik, H., Rafiee, S., Shafizadeh, A., Jeddi, D., Jafary, T., Lam, S. S., ... & Aghbashlo, M. (2022). Characterizing sludge pyrolysis by machine learning: Towards sustainable bioenergy production from wastes. Renewable Energy, 199, 1078–1092. https://doi.org/10.1016/j.renene.2022.09.022

Shirley, W. H., Voisinet, M. M., Waddell Smith, R., & McGuffin, V. L. (2027). Improvements in multivariate regression models to predict blend level of multiple-feedstock biodiesels. Fuel, 428(Part D), Article 140418. https://doi.org/10.1016/j.fuel.2026.140418

Wojnowski, W., Majchrzak, T., Dymerski, T., Gębicki, J., & Namieśnik, J. (2017). Portable electronic nose based on electrochemical sensors for food quality assessment. Sensors, 17(12), Article 2715. https://doi.org/10.3390/s17122715

Xu, M., Li, X., Jia, P., & Zhang, L. (2023). Classification techniques of electronic nose: A review. International Journal of Bio-Inspired Computation, 21(1). https://doi.org/10.1504/ijbic.2023.10059921