

## طبقه‌بندی برنج سفید خالص و مخلوط با استفاده از طیف‌سنجی مرئی-مادون قرمز نزدیک و یادگیری ماشین

حامد زارع ثانی<sup>۱</sup>، امیرحسین افکاری سیاح<sup>۲</sup> و حماد ذرعی‌فروش<sup>۳\*</sup>

### چکیده

برنج، غذای اصلی حدود ۲/۵ میلیارد نفر از جمعیت جهان است. کیفیت این محصول تحت تأثیر عوامل مختلفی است. در نتیجه، کنترل کیفیت و تشخیص اختلاط از مسائل مهم صنعت برنج محسوب می‌شود. هدف از این تحقیق، بررسی و تشخیص اختلاط برنج با به کارگیری روش طیف‌سنجی Vis/NIR و روش‌های شیمی‌سنجی و یادگیری ماشین بود. برای انجام آزمایش‌ها، ۱۳ طبقه براساس میزان اختلاط رقم اصلی برنج (رقم هاشمی) با ارقام غیراصلی فجر، گیلانه، خزر و شیرودی تهیه شد. بدین صورت که طبقه‌های ناخالص حاصل اختلاط ۹۰، ۸۰ و ۷۰ درصد از برنج رقم هاشمی به ترتیب با ۱۰، ۲۰ و ۳۰ درصد از ارقام دیگر مورد اشاره بود. برای طبقه‌بندی نمونه‌های برنج از الگوریتم‌های پرسپترون چندلایه (MLP)، درخت تصمیم (DT) و ماشین‌های بردار پشتیبان (SVM) استفاده شد. برای ارزیابی عملکرد طبقه‌بندی‌های مورد مطالعه از شاخص‌های آماری شامل حساسیت (Se)، ویژه بودن (Sp)، صحت (Ac) و ریشه میانگین مربعات خطا (RMSE) در قالب توسعه ماتریس‌های اغتشاش طبقه‌بندی استفاده شد. نتایج مطالعه نشان داد که الگوریتم‌های درخت تصمیم از نوع J48، شبکه‌های عصبی مصنوعی پرسپترون چندلایه و ماشین بردار پشتیبان با تابع هسته RBF به ترتیب با مقادیر صحت و RMSE ۱۰۰ درصد و صفر، ۹۶/۹۲ درصد و ۰/۰۹۵۱ و ۹۲/۳۱ درصد و ۰/۲۴۸۳ توانستند نمونه‌های مختلف برنج سفید را طبقه‌بندی کنند.

**واژه‌های کلیدی:** برنج، درجه‌بندی، شبکه‌های عصبی مصنوعی، درخت تصمیم، طیف‌سنجی مرئی-مادون قرمز نزدیک.

**ارجاع:** زارع ثانی ح، افکاری سیاح ا، ح. و ذرعی‌فروش ح. ۱۴۰۳. طبقه‌بندی برنج سفید خالص و مخلوط با استفاده از طیف‌سنجی مرئی-مادون قرمز نزدیک و یادگیری ماشین. نشریه پژوهش‌های مکانیک ماشین‌های کشاورزی. ۳۲، ۴۴-۴۴. <https://doi.org/10.22034/JRMAM.2024.14748.704>

۱- دانشجوی دکتری تخصصی گروه مهندسی بیوسیستم، دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران.

۲- دانشیار گروه مهندسی بیوسیستم، دانشکده کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران.

۳- دانشیار گروه مهندسی بیوسیستم، دانشکده علوم کشاورزی، دانشگاه گیلان، رشت، ایران.

\* نویسنده مسئول: [hemad.zareiforush@guilan.ac.ir](mailto:hemad.zareiforush@guilan.ac.ir)

تاریخ دریافت: ۱۴۰۳/۰۲/۰۲ تاریخ پذیرش: ۱۴۰۳/۰۴/۱۱

## مقدمه

برنج گیاهی یک ساله، علفی، با ریشه افشان، ایستاده، قوی و به رنگ سفید بوده و از خانواده *Oryza* متعلق به گروه *Oryzae* است (Van Nguyen & Ferrero, 2006). این محصول غذای اصلی حدود ۲/۵ میلیارد نفر از جمعیت جهان بوده که حدود ۲۰ درصد از انرژی مورد نیاز آن‌ها را تأمین می‌کند (Qamar *et al.*, 2013). از تولیدکنندگان اصلی برنج می‌توان به کشورهای تایلند، برمه، ویتنام، لائوس، فیلیپین، اندونزی، پاکستان، آمریکا، هند، ژاپن، ایتالیا، چین، مصر، برزیل، مکزیک، کوبا و استرالیا اشاره کرد، که در این میان ارقام هاشمی، طارم و صدری از بهترین و مرغوب‌ترین ارقام برنج بومی کشور ایران به حساب می‌آیند (Rasooli Sharabiani & Khorramifar, 2022b).

به‌طور معمول برنج معطر به خاطر داشتن کیفیت خوب (مانند ظرافت، رنگ، شکل، مزه، طعم و عطر) توسط مصرف‌کنندگان ترجیح داده می‌شود (Choudhury *et al.*, 2001). کیفیت برنج معطر تحت تأثیر عوامل مختلفی است، مانند: شرایط آب و هوایی، محل کشت، فعالیت‌های ژنتیکی و پس از برداشت (Champagne, 2008). فعالیت‌های پس از برداشت مؤثر بر کیفیت برنج، شامل شرایط و مدت زمان انبارمانی، روش خشک شدن، فرآیند غنی‌سازی و مواد بسته‌بندی است که بایستی شرایط نگهداری، دما، زمان و رطوبت به منظور حفظ کیفیت برنج کنترل گردد (Wongpornchai *et al.*, 2004). مسائل مهم موجود در صنعت برنج عبارتند از: برچسب‌گذاری اشتباه، کنترل کیفیت، درجه‌بندی و اختلاط در انواع مختلف برنج. به همین خاطر، صنعت برنج از درجه‌بندی‌های استاندارد بر اساس معیارهای بازار برای شناسایی دانه استفاده می‌کند. استفاده از پانل‌های متخصص انسانی مرسوم‌ترین روشی است که برای ارزیابی کیفیت برنج معطر به کار می‌رود که برنج را بر اساس ویژگی عطر تشخیص داده می‌شود (Rasooli Sharabiani & Khorramifar, 2022b). البته این روش دارای اشکالاتی نیز هست، نظیر آموزش‌هایی که امکان دارد سال‌ها به طول انجامد و یا اگر تعداد نمونه‌ها افزایش یابد، افراد دچار خستگی می‌شوند و منجر به کاهش دقت نتایج می‌گردد (Pearce *et al.*, 2006). همچنین کیفیت برنج معطر را نیز می‌توان به کمک تجزیه و تحلیل GC-MS

(طیف‌سنجی جرمی کروماتوگرافی گازی) ارزیابی کرد، اما این روش پر هزینه و مفصل است.

چندین مرحله بازرسی (برای مشخص کردن نوع و کیفیت دانه) ممکن است در حین عملیات جابجایی انجام شود. بنابراین، بررسی خلوص رقم دانه‌ها نسبت به عوامل دیگر مانند طعم، عطر، اندازه، پاکیزگی و رنگ پیچیده‌تر و دشوارتر است. از این رو، یک روش غیرمخرب و سریع به منظور ارزیابی برنج و جابجایی آسان مورد نیاز است (Abdullah *et al.*, 2016).

روش‌های مرسوم ارزیابی کیفیت اغلب مخرب و ناکارآمد بوده و سامانه ارزیابی کیفیت باید از سرعت بالا، دقت خوب و هزینه کم برخوردار باشد. به علاوه، روش‌های نوین آزمون غیرمخرب مانند پردازش تصویر و بینایی ماشین هم ممکن است در مواردی مانند تشخیص اختلاط ارقام و تقلب در برنج سفید به خصوص در مورد ارقام برنج ایرانی که از نظر خواص فیزیکی و ظاهری بسیار نزدیک به هم هستند، از دقت قابل قبولی برخوردار نباشند. چنین اهدافی را می‌توان به کمک روش‌های طیف‌سنجی NIR محقق کرد، زیرا نیازی به آماده‌سازی نمونه ندارند و دارای مزایایی چون غیر مخرب بودن، سریع، دقیق، کم هزینه و بدون آلودگی هستند (Shahgholi *et al.*, 2022; Sun *et al.*, 2009). این روش که بر پایه جذب تابش در ناحیه فرو سرخ نزدیک طیف الکترومغناطیسی است، توانسته برای کنترل کیفیت محصولات کشاورزی و غذایی مورد استفاده قرار گیرد (Hu *et al.*, 2019; Quiñones *et al.*, 2018; Williams, 2019).

نخستین بار در سال ۱۹۶۸ از طیف‌سنجی NIR برای بررسی ترکیبات غلات استفاده شد (Ben-Gera & Norris, 1968). در تحقیقی با استفاده از طیف‌سنجی بازتابشی فروسرخ نزدیک به بررسی خصوصیات کیفی گوجه فرنگی مانند سفتی، SSC و pH آن پرداختند و این خصوصیات را با ضریب همبستگی بالا به صورت غیرمخرب پیش‌بینی کردند. ضریب همبستگی برای پیش‌بینی SSC برابر با ۰/۸۹ و میزان خطا نیز برابر با ۰/۳۷۷ درجه بریکس بود (Shao *et al.*, 2007).

در تحقیقی دیگر (Fu *et al.*, 2017) از NIR به همراه دستگاه درون-ترکیبی یک در مقابل یک حداقل مربع پشتیبانی ماشین بردار (IC-OVO-LS-SVM) برای طبقه‌بندی *Chinese Ganoderma lucidum* استفاده شد

برای هر نمونه، طیف‌سنجی با استفاده از نرم‌افزار Spectra-Wiz Spectrometer OS v5.33 (c) 2014 انجام شد و داده‌های به دست آمده بعد از میانگین‌گیری ثبت شدند. این نرم‌افزار داده‌های جذبی را به صورت مستقیم استخراج می‌کند و در نتیجه نیاز به تبدیل داده‌ها نیست (Khorramifar *et al.*, 2022c; Tarighi & Khorramifar, 2023). در شکل ۱ نمودار حاصل از طیف‌سنجی Vis/NIR برای طبقه خالص (رقم هاشمی) و طبقه‌های ناخالص پس از حذف نویزها قابل مشاهده است.

جدول ۱- مشخصات نمونه‌های برنج مورد مطالعه در تحقیق

ردیف	مشخصات ترکیب طبقه	کد
۱	رقم هاشمی با خلوص ۱۰۰ درصد	HASHEMI
۲	مخلوط رقم هاشمی ۹۰٪ + رقم فجر ۱۰٪	H90F10
۳	مخلوط رقم هاشمی ۸۰٪ + رقم فجر ۲۰٪	H80F20
۴	مخلوط رقم هاشمی ۷۰٪ + رقم فجر ۳۰٪	H70F30
۵	مخلوط رقم هاشمی ۹۰٪ + رقم گیلانه ۱۰٪	H90G10
۶	مخلوط رقم هاشمی ۸۰٪ + رقم گیلانه ۲۰٪	H80G20
۷	مخلوط رقم هاشمی ۷۰٪ + رقم گیلانه ۳۰٪	H70G30
۸	مخلوط رقم هاشمی ۹۰٪ + رقم خزر ۱۰٪	H90K10
۹	مخلوط رقم هاشمی ۸۰٪ + رقم خزر ۲۰٪	H80K20
۱۰	مخلوط رقم هاشمی ۷۰٪ + رقم خزر ۳۰٪	H70K30
۱۱	مخلوط رقم هاشمی ۹۰٪ + رقم شیرودی ۱۰٪	H90S10
۱۲	مخلوط رقم هاشمی ۸۰٪ + رقم شیرودی ۲۰٪	H80S20
۱۳	مخلوط رقم هاشمی ۷۰٪ + رقم شیرودی ۳۰٪	H70S30

### انتخاب ویژگی‌ها و طبقه‌بندی نمونه‌ها

با توجه به تعدد طول‌موج‌های مورد بررسی در پژوهش حاضر که از منظر یادگیری ماشین جزو ویژگی‌های ورودی به مسأله طبقه‌بندی نمونه‌های برنج محسوب می‌شوند، از روش انتخاب ویژگی مبتنی بر همبستگی (CFS) به منظور انتخاب طول موج‌هایی که ارتباط قوی‌تری با طبقه‌های مورد مطالعه داشتند، استفاده شد. الگوریتم انتخاب ویژگی مبتنی بر همبستگی (CFS) یکی از متداول‌ترین و معتبرترین روش‌های انتخاب ویژگی در مسائل تشخیص الگویی است. روش CFS یک رویکرد فیلتر ویژگی تحت نظارت است که می‌تواند براساس تابع همبستگی یک زیرمجموعه بهینه از ویژگی‌ها را از مجموعه داده‌های اولیه انتخاب نماید. در این روش زیرمجموعه‌ای از ویژگی‌ها براساس یک تابع ارزیابی ابتکاری مبتنی بر شاخص همبستگی رتبه‌بندی می‌شوند

و نتایج نشان داد که با اطمینان می‌توان از NIR و با استفاده از الگوریتم‌های یادگیری ماشین برای شناسایی و طبقه‌بندی در صنایع غذایی استفاده کرد.

با توجه به اهمیت تشخیص اختلاط ارقام در ارقام برنج سفید که در سال‌های گذشته به چالشی در زمینه بازار خرید و فروش این محصول ارزشمند تبدیل شده است و خطر از بین رفتن اصالت ارقام اصیل نظیر برنج رقم هاشمی (که از ارقام باکیفیت برنج تولید شده در کشور است) را به دنبال دارد. هدف از این پژوهش ارزیابی توانایی و دقت طیف‌سنجی Vis/NIR در ترکیب با الگوریتم‌های یادگیری ماشین برای تشخیص و تمایز یک طبقه رقم اصلی و مطلوب برنج از ۴ طبقه نمونه متشکل از ارقام غیرمطلوب بود.

### مواد و روش‌ها

#### تهیه و آماده‌سازی نمونه‌ها

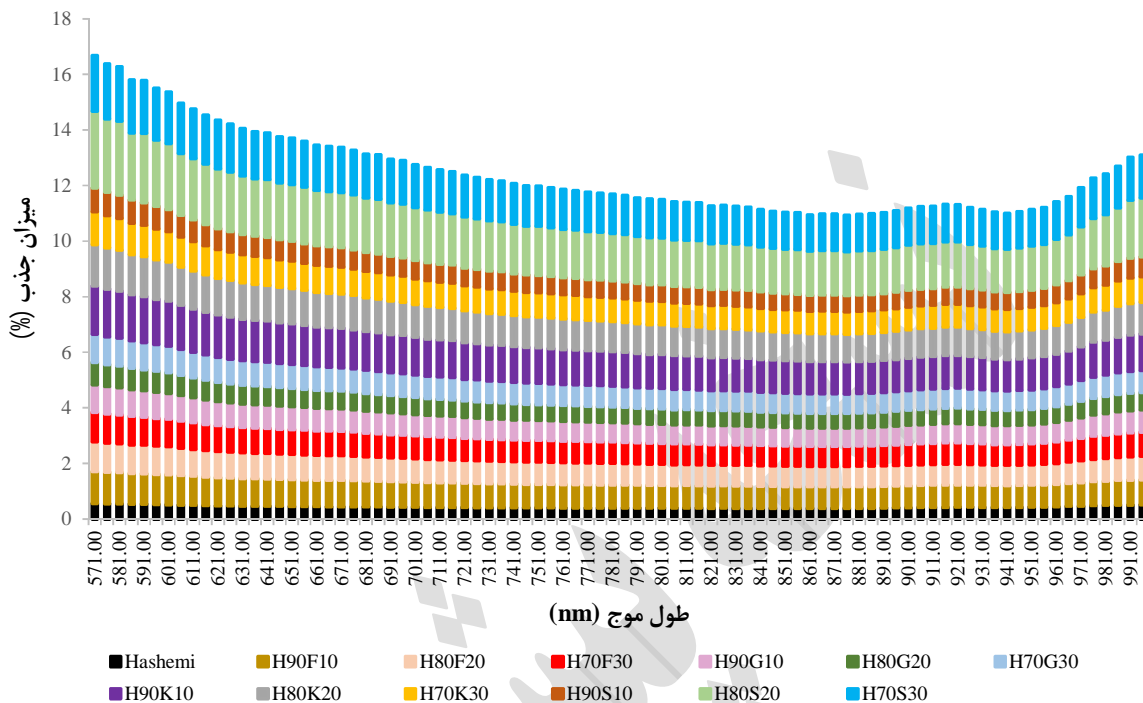
ابتدا ۵ رقم برنج از مؤسسه تحقیقات برنج کشور واقع در شهرستان رشت تهیه شد. این ۵ رقم شامل یک رقم برنج اصلی و مطلوب به نام هاشمی و ۴ رقم برنج غیرمطلوب به نام‌های فجر، گیلانه، خزر و شیرودی بود. بنابراین در آزمایش‌ها یک طبقه برنج خالص (شامل رقم هاشمی) و ۱۲ طبقه برنج ناخالص (شامل درصدهای مختلفی از اختلاط ارقام فجر، گیلانه، خزر و شیرودی با رقم هاشمی) تهیه شد، بدین صورت که طبقه‌های ناخالص حاصل اختلاط ۹۰، ۸۰ و ۷۰ درصد از برنج رقم هاشمی به ترتیب با ۱۰، ۲۰ و ۳۰ درصد از ارقام دیگر مورد اشاره بود. برای هر یک از طبقه‌ها ۲۵ نمونه تهیه شد. با توجه به این توضیحات، در مجموع ۱۳ طبقه نمونه برنج به شرح جدول ۱ تهیه شد.

#### طیف‌سنجی Vis/NIR

طیف‌سنجی Vis/NIR با استفاده از یک اسپکترومتر مدل PS-100 (Apogee Instruments, INC. Logan, Utah, USA) با آشکارساز CCD، ۲۰۴۸ پیکسل، با قدرت تفکیک ۱ نانومتر و منبع نور هالوژن-تنگستن در محدوده طول‌موج‌های ۱۱۰۰-۳۵۰ نانومتر انجام شد. همچنین برای کالیبراسیون از یک Reflectance standard استفاده شد (Rasooli Sharabiani & Khorramifar, 2022a).

جزو پرکاربردترین الگوریتم‌های طبقه‌بندی هستند که به-  
 طور وسیعی جهت تحلیل داده در کاربردهای کیفیت-  
 سنجی مواد غذایی استفاده می‌شوند (Anwar *et al.*,  
 2023; Zhu *et al.*, 2021).

(Bakhshipour & Zareiforoush, 2020b). در تحقیق  
 حاضر از الگوریتم‌های پرسپترون چندلایه (MLP)، درخت  
 تصمیم (DT) و ماشین‌های بردار پشتیبان (SVM) برای  
 طبقه‌بندی نمونه‌های برنج استفاده شد. این الگوریتم‌ها



شکل ۱- نمودار طیف‌سنجی Vis/NIR مربوط به نمونه‌های مختلف برنج

سازماندهی درخت‌های تصمیم به‌گونه‌ای است که  
 در هر لایه یک طبقه نامطلوب رد می‌شود. آخرین طبقه  
 باقیمانده در بخش پایینی درخت را به عنوان طبقه تعیین  
 شده در نظر می‌گیرند. شاخه‌های خروجی هر گره با  
 نتیجه احتمالی آزمایش در آن گره مطابقت دارد  
 (Zareiforoush *et al.*, 2016). در تحقیق حاضر الگوریتم-  
 های درخت تصمیم از نوع J48 (C4.5 decision tree  
 learner), REP (reduced-error pruning), LMT  
 (logistic model trees) که جزو الگوریتم‌های درخت  
 تصمیم مرسوم در عملیات طبقه‌بندی نظارت شده  
 هستند، برای طبقه‌بندی کیفی نمونه‌های برنج استفاده  
 شد. این نوع الگوریتم‌های درخت تصمیم در گذشته  
 به‌صورت موفقیت‌آمیز برای درجه‌بندی شاخص‌های کیفی  
 مواد غذایی و محصولات کشاورزی استفاده شده‌اند  
 (Bakhshipour *et al.*, 2018; Liakos *et al.*, 2018; )  
 (Wang *et al.*, 2022).

روش ماشین‌های بردار پشتیبان یک الگوریتم  
 یادگیری و طبقه‌بندی تحت نظارت بر اساس نظریه آماری  
 است (Khorramifar *et al.*, 2022a). یک SVM می‌تواند  
 داده‌های ورودی را که در یک فضای کم به‌طور خطی قابل  
 تفکیک نیستند، با استفاده از یک تابع هسته به صورت  
 غیرخطی در یک فضای بزرگ‌تر تفکیک کند. یک ابر  
 صفحه در فضای با ابعاد بالا ساخته می‌شود تا فاصله بین  
 دو طبقه را حداکثر کرده و داده‌ها را در فضای با ابعاد بالا  
 طبقه‌بندی کند. از آنجا که SVM خطاهای ساختاری را  
 به حداقل می‌رساند، به عنوان طبقه‌بندی کننده خوبی  
 برای داده‌های غیرخطی و اندازه نمونه‌های کوچک در نظر  
 گرفته می‌شود (Khorramifar *et al.*, 2022b). در این  
 مطالعه، توابع مختلف هسته شامل هسته چند جمله‌ای،  
 هسته چند جمله‌ای عادی، هسته مبتنی بر تابع Pearson  
 VII (PUK) و هسته RBF برای توسعه مدل‌های  
 طبقه‌بندی SVM استفاده شد و طبقه‌بند با بهترین  
 عملکرد بر اساس نوع تابع هسته در نتایج گزارش شد.

$$S_e = \frac{TP}{TP + FN} \times 100 \quad (3)$$

$$S_p = \frac{TN}{TN + FP} \times 100 \quad (4)$$

در این معادلات TP تعداد مواردی که به درستی در طبقه مورد نظر تشخیص داده شده‌اند، FP تعداد مواردی که به اشتباه در طبقه مورد نظر تشخیص داده نشده‌اند، FN تعداد مواردی که به اشتباه در طبقه مورد نظر تشخیص داده شده‌اند و TN تعداد مواردی که به درستی در طبقه مورد نظر تشخیص داده نشده‌اند، را بیان می‌کنند.

شاخص آماری RMSE نشان دهنده انحراف استاندارد تفاوت بین مقادیر پیش‌بینی شده و مقادیر مشاهده شده نمونه‌ها است. مقدار RMSE با استفاده از معادله (۵) محاسبه شد.

$$RMSE = \left[ \frac{1}{n} \sum_{i=1}^n (Y_{obs,i} - Y_{pred,i})^2 \right]^{0.5} \quad (5)$$

در معادله فوق،  $Y_{obs,i}$  و  $Y_{pred,i}$  به ترتیب عبارتند از آمین مقدار پیش‌بینی شده و آمین مقدار اندازه‌گیری شده (Bakhsipour & Zareiforoush, 2020a; Han et al., 2020; Huang & Gu, 2022; Liu et al., 2018). تجزیه و تحلیل داده‌ها و پیاده‌سازی الگوریتم‌های طبقه‌بندی با استفاده از نرم‌افزار weka-3-8-6 انجام شد.

## نتایج و بحث

### نمودار تغییرات طیف به دست آمده از طول موج‌های منتخب

نمودار مقایسه‌ای مقادیر بازتاب به دست آمده از طبقه‌های مختلف نمونه‌های برنج در ۱۰ طول موج برگزیده توسط الگوریتم انتخاب ویژگی CFS در شکل ۲ نشان داده شده است. در این نمودار اعداد واقع در مرزهای بیرونی نشان‌دهنده عدد موج‌های منتخب و اعداد روی خط عمودی درون نمودار نشان‌دهنده مقادیر بازتاب است. همان‌طور که مشاهده می‌شود، در تمام طول موج‌های منتخب، مقادیر بازتاب طبقه اصلی (رقم هاشمی خالص) به طور قابل توجهی کمتر از طبقه‌های دیگر بوده و با افزایش اختلاط ارقام غیراصلی، مقادیر بازتاب افزایش یافته است. در نتیجه، توجه به مقادیر بازتاب در روش

پرسپترون چند لایه (MLP)، یکی از رایج‌ترین انواع شبکه‌های عصبی مصنوعی برای انجام طبقه‌بندی است. MLPها شامل سه لایه اصلی (لایه ورودی، لایه مخفی و لایه خروجی) بوده و لایه‌ها متعلق به طبقه شبکه‌های پیشخور هستند (Hagbin et al., 2023; Nayak et al., 2023)، یعنی اطلاعات از گره‌های شبکه فقط در جهت رو به جلو عبور می‌کنند. به منظور طبقه‌بندی نمونه‌های برنج اصلی و غیراصلی، مدل MLP با استفاده از الگوریتم پس انتشار آموزش داده شد. این الگوریتم وزن تابع فعال‌سازی را برای هر نورون محاسبه می‌کند (Karray et al., 2004). در شبکه‌های پیش‌خور، روش‌های گوناگونی برای کمینه‌سازی خطا وجود دارد از جمله شیب نزول، گرادیان نزول با تکانه، Levenberg-Marquardt، گرادیان مزدوج و غیره (Omid et al., 2010). در این پژوهش از روش شیب نزول با رویکرد تکانه برای به حداقل رساندن خطا با ضریب تکانه ۰/۲ استفاده شد.

### ارزیابی عملکرد الگوریتم‌های طبقه‌بندی

بطور کلی ۶۰ درصد داده‌ها برای مرحله آموزش و ۴۰ درصد داده‌های برای مراحل آزمون و اعتبارسنجی طبقه‌بندی استفاده شد. عملکرد الگوریتم‌های فراابتکاری مورد استفاده در طبقه‌بندی نمونه‌های برنج با تشکیل ماتریس اغتشاش (CM) و محاسبه ویژگی‌های آماری شامل حساسیت ( $S_e$ )، ویژه بودن ( $S_p$ )، صحت ( $A_c$ ) و ریشه میانگین مربعات خطا (RMSE) مورد ارزیابی قرار گرفت. شاخص حساسیت توانایی طبقه‌بندی را برای شناسایی صحیح یک طبقه مشخص می‌کند. شاخص ویژه بودن نشان‌دهنده توانایی طبقه‌بندی برای حذف یک طبقه نامربوط به درستی است. این ویژگی که به آن «نرخ پاسخ‌های منفی درست» نیز می‌گویند، به معنی نسبتی از موارد منفی است که طبقه‌بندی آن‌ها را به درستی به عنوان نمونه منفی تشخیص داده است. برای محاسبه ویژگی‌های ذکر شده از معادلات (۱) الی (۴) استفاده شده است (Mollazade et al., 2012; Zareiforoush et al., 2016):

$$CM = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \quad (1)$$

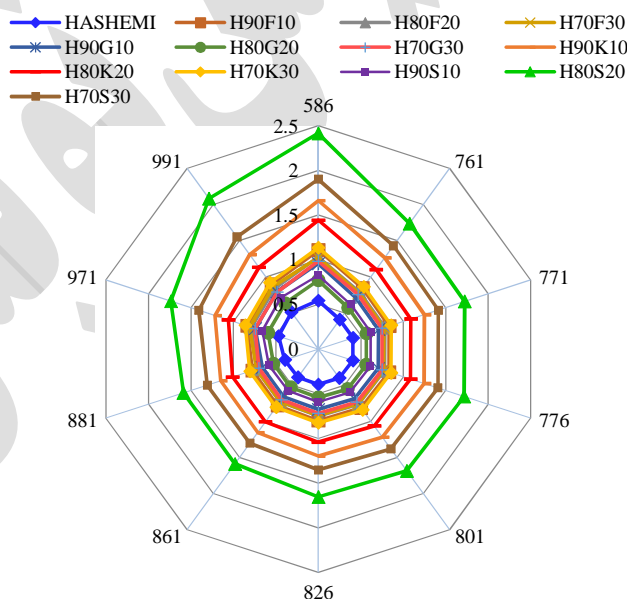
$$A_c = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \quad (2)$$

۰/۰۹۵۱ به‌دست آمد. در مطالعه‌ای، از سامانه تصویربرداری چندطیفی برای تشخیص اختلاط در برنج یاس تایلندی همراه با روش‌های شیمیایی از جمله تجزیه و تحلیل مؤلفه‌های اصلی (PCA)، حداقل مربعات جزئی (PLS) و LS-SVM و شبکه عصبی پس از انتشار (BPNN) استفاده شد. محققین گزارش کردند که چهار نوع از نمونه‌های برنج را می‌توان با دقت ۹۲ درصد توسط مدل BPNN طبقه‌بندی کرد و می‌توان به کمک فناوری تصویربرداری چندطیفی با روش‌های شیمیایی، تشخیص سریع و غیر مخرب اصالت برنج یاس تایلندی را انجام داد (Liu et al., 2021). در تحقیقی که به منظور شناسایی ارقام برنج از بینایی ماشین و روش MLP استفاده شد، دقت تشخیص ۸۴/۸۳ درصد گزارش شده است (Gujjar & Siddappa, 2013) که دارای دقت پایین‌تری نسبت به نتایج تحقیق حاضر است. نتایج به‌دست آمده در این تحقیق با نتایج تحقیق مشابهی که بر روی طبقه‌بندی برنج به کمک بینایی ماشین انجام شد و دقتی برابر با ۹۰ درصد را گزارش کردند، مطابقت دارد (Kaur & Singh, 2013).

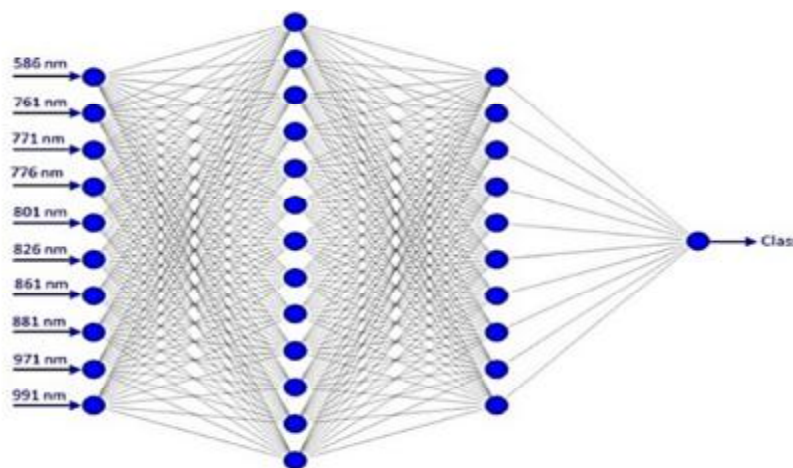
طیف‌سنجی Vis/NIR می‌تواند معیار مناسبی برای تشخیص برنج رقم خالص هاشمی از ارقام مخلوط و ناخالص باشد.

### طبقه‌بندی با الگوریتم پرسپترون چندلایه

برای شناسایی و تمایز طبقه‌های مختلف ارقام برنج بر اساس داده‌های حاصل از طیف سنجی Vis/NIR، از روش‌های پرسپترون چندلایه (MLP)، درخت تصمیم (DT)، شبکه بیزین (BN) و ماشین‌های بردار پشتیبان (SVM) استفاده شد. ساختار مربوط به طبقه‌بند پرسپترون چندلایه (MLP) در شکل ۳ نشان داده شده است. ساختارشناسی شبکه عصبی منتخب به صورت ۱-۱۰-۱۳-۱۰ بود که شامل لایه اول (طول موج‌های منتخب)، دو لایه پنهان با تعداد ۱۳ نورون در لایه مخفی اول و ۱۰ نورون در لایه مخفی دوم و لایه آخر (طبقه‌بندی) بود. ماتریس اغتشاش مربوط به طبقه‌بند پرسپترون چندلایه (MLP) به همراه شاخص‌های آماری مربوط به عملکرد این طبقه‌بند در جدول ۲ نشان داده شده است. مقادیر شاخص‌های آماری صحت و RMSE برای طبقه‌بند پرسپترون چندلایه به ترتیب برابر با ۹۶/۹۲ درصد و



شکل ۲- نمودار مقایسه‌ای مقادیر بازتاب به‌دست آمده از طبقه‌های مختلف نمونه‌های برنج در ۱۰ طول موج برگزیده توسط الگوریتم انتخاب ویژگی CFS



شکل ۳- ساختار مربوط به الگوریتم پرسپترون چندلایه برای طبقه‌بندی ارقام برنج

برابر با ۹۳، ۹۶/۵ و ۸۰ درصد بود. در این بین دقت روش LDA بسیار منطبق با دقت LDA در پژوهش ما بود. (Shi *et al.* (2023) نیز طی تحقیقی به بررسی امکان تمایز سریع برنج انبار شده به مدت یک سال از دو سال با استفاده از طیف سنجی NIR و الگوریتم‌های مختلف پرداخته و گزارش نمودند که روش‌های PLS-DA و SVM توانایی طبقه‌بندی عالی با دقت ۱۰۰ درصد در شناسایی تازگی برنج، با حساسیت و ویژه بودن ۱۰۰ درصد را از خود نشان دادند.

#### طبقه‌بندی با الگوریتم درخت تصمیم

ماتریس اغتشاش مربوط به طبقه‌بند درخت تصمیم از نوع J48 به همراه شاخص‌های آماری مربوط به عملکرد این طبقه‌بند در جدول ۴ نشان داده شده است. مقادیر شاخص‌های آماری صحت و RMSE برای طبقه‌بند درخت تصمیم از نوع J48 به ترتیب برابر با ۱۰۰ درصد و صفر به-دست آمد. نمودار درختی ایجاد شده توسط الگوریتم J48 نیز در شکل ۴ قابل مشاهده است. با به‌کارگیری یکی دیگر از الگوریتم‌های درخت تصمیم مورد مطالعه در این پژوهش یعنی الگوریتم REP نیز نتایج مشابهی از لحاظ شاخص‌های صحت و RMSE حاصل شد که به دلیل تشابه عملکردی، فقط نتایج الگوریتم J48 گزارش شده است.

همان‌طور که مشاهده می‌شود، الگوریتم درخت تصمیم در طبقه‌بندی نمونه‌های برنج دارای دقت بسیار بالایی است. Zareiforouh *et al.* (2016) در تحقیق خود برای طبقه-

#### طبقه‌بندی با الگوریتم ماشین‌های بردار پشتیبان (SVM)

ماتریس اغتشاش مربوط به طبقه‌بند ماشین‌های بردار پشتیبان (SVM) به همراه شاخص‌های آماری مربوط به عملکرد این طبقه‌بند در جدول ۳ نشان داده شده است. مقادیر شاخص‌های آماری دقت، حساسیت، ویژگی‌پذیری و RMSE برای طبقه‌بند ماشین‌های بردار پشتیبان با تابع هسته RBF به ترتیب برابر با ۹۲/۳۱ درصد و ۰/۲۴۸۳ به‌دست آمد. (Abdullah *et al.* (2016) با بویایی ماشین ۱۷ نمونه برنج را در ۴ دسته شناسایی و طبقه‌بندی کردند. طبق گزارش آن‌ها دقت روش SVM در این طبقه‌بندی برابر با ۱۰۰ درصد و بهترین روش طبقه‌بندی برای ارقام برنج بود. همچنین بیان کردند که روش KNN دومین روش برتر جهت طبقه‌بندی نمونه‌های برنج است. نتایج آن‌ها با نتایج یافته‌های پژوهش حاضر هم‌خوانی داشت. در تحقیقی دیگر (Liu *et al.* (2017) از تصویربرداری چندطیفی و روش‌های شیمی‌سنجی جهت تشخیص اختلاط ساکارز در رب گوجه فرنگی استفاده کردند. آن‌ها گزارش کردند که دقت روش‌های PLS و BPNN با ۹۳ درصد و کمتر از دو روش دیگر (LS-SVM و PCA) بود. همچنین دقت روش‌های LS-SVM و PCA نیز به ترتیب برابر با ۹۶ و ۹۸ درصد بوده و بنابراین روش PCA نسبت به روش‌های دیگر پاسخ خوبی داشت. همچنین (Jana *et al.* (2011) از روش بین‌الکترونیک با روش‌های ANN، PCA و LDA برای تشخیص برنج معطر و غیرمعطر استفاده کردند که دقت نتایج آن‌ها به ترتیب

طبقه‌بندی نمونه‌های سفید شده و قهوه‌ای واریته‌های مختلف برنج با استفاده از تصویربرداری فراطیفی داشتند؛ گزارش نمودند که بهترین دقت طبقه‌بندی برای برنج سفید شده و قهوه‌ای متعلق به روش SVM و به ترتیب برابر با ۹۵/۲ و ۹۳ درصد بوده است.

بندی کیفی دانه‌های برنج آسیاب شده از روش بینایی ماشین و الگوریتم‌های MLP، Bayesian و درخت تصمیم استفاده کردند که دقت طبقه‌بندی گزارش شده به ترتیب برابر با ۹۶/۶۷، ۹۷/۵۱ و ۹۷/۹۸ درصد بود. (Onmankhong *et al.* (2022) نیز طی مطالعه‌ای که برای

جدول ۲- ماتریس اغتشاش مربوط به الگوریتم پرسپترون چندلایه (MLP) برای طبقه‌بندی نمونه‌های برنج

A	B	C	D	E	F	G	H	I	J	K	L	M	طبقه مورد نظر
۱۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	A = HASHEMI
۰	۱۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	B = H90F10
۰	۰	۱۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	C = H80F20
۰	۰	۰	۱۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	D = H70F30
۰	۰	۰	۰	۱۰	۰	۰	۰	۰	۰	۰	۰	۰	E = H90G10
۰	۰	۰	۰	۰	۱۰	۰	۰	۰	۰	۰	۰	۰	F = H80G20
۰	۰	۰	۰	۰	۰	۱۰	۰	۰	۰	۰	۰	۰	G = H70G30
۰	۰	۰	۰	۰	۰	۰	۱۰	۰	۰	۰	۰	۰	H = H90K10
۰	۰	۰	۰	۰	۰	۰	۰	۱۰	۰	۰	۰	۰	I = H80K20
۰	۴	۰	۰	۰	۰	۰	۰	۰	۶	۰	۰	۰	J = H70K30
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱۰	۰	۰	K = H90S10
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱۰	۰	L = H80S20
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱۰	M = H70S30
۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۶۰	۱۰۰	۱۰۰	۱۰۰	حساسیت
۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۹۶/۷۷	۱۰۰	۱۰۰	۱۰۰	ویژه بودن

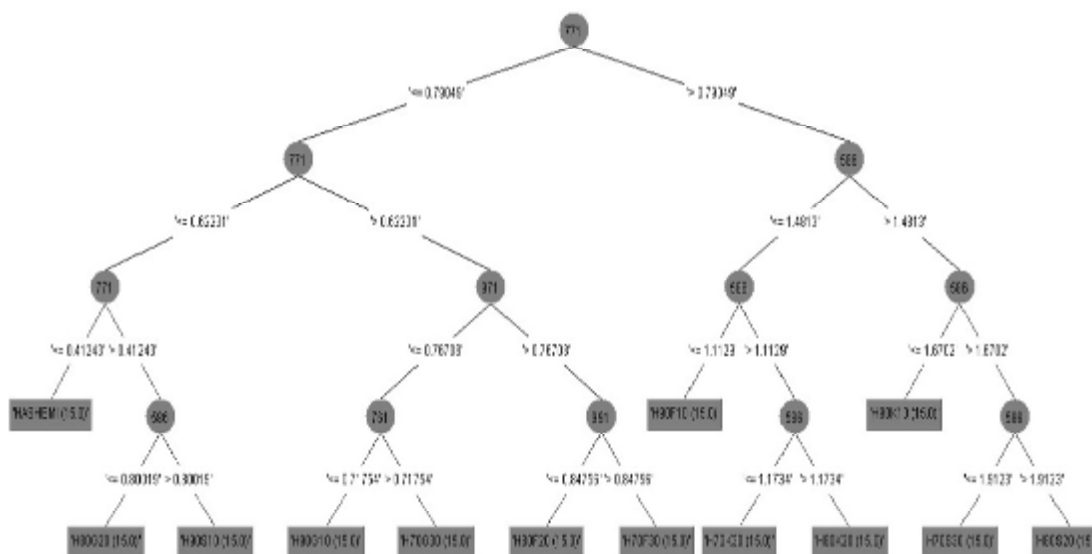
جدول ۳- ماتریس اغتشاش مربوط به الگوریتم ماشین‌های بردار پشتیبان (SVM) برای طبقه‌بندی نمونه‌های برنج

A	B	C	D	E	F	G	H	I	J	K	L	M	طبقه مورد نظر
۱۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	A = HASHEMI
۰	۱۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	B = H90F10
۰	۰	۱۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	C = H80F20
۰	۰	۰	۱۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	D = H70F30
۰	۰	۰	۰	۱۰	۰	۰	۰	۰	۰	۰	۰	۰	E = H90G10
۰	۰	۰	۰	۰	۱۰	۰	۰	۰	۰	۰	۰	۰	F = H80G20
۰	۰	۲	۰	۲	۰	۶	۰	۰	۰	۰	۰	۰	G = H70G30
۰	۰	۰	۰	۰	۰	۰	۱۰	۰	۰	۰	۰	۰	H = H90K10
۰	۰	۰	۰	۰	۰	۰	۰	۱۰	۰	۰	۰	۰	I = H80K20
۰	۶	۰	۰	۰	۰	۰	۰	۰	۴	۰	۰	۰	J = H70K30
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱۰	۰	۰	K = H90S10
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱۰	۰	L = H80S20
۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱۰	M = H70S30
۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۶۰	۱۰۰	۱۰۰	۴۰	۱۰۰	۱۰۰	۱۰۰	حساسیت
۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۹۶/۶۱	۱۰۰	۱۰۰	۹۵/۰۸	۱۰۰	۱۰۰	۱۰۰	ویژه بودن



جدول ۴- ماتریس اغتشاش مربوط به الگوریتم درخت تصمیم از نوع J48 برای طبقه‌بندی نمونه‌های برنج

A	B	C	D	E	F	G	H	I	J	K	L	M	طبقه مورد نظر
۱۰	.	.	.	.	.	.	.	.	.	.	.	.	A = HASHEMI
.	۱۰	.	.	.	.	.	.	.	.	.	.	.	B = H90F10
.	.	۱۰	.	.	.	.	.	.	.	.	.	.	C = H80F20
.	.	.	۱۰	.	.	.	.	.	.	.	.	.	D = H70F30
.	.	.	.	۱۰	.	.	.	.	.	.	.	.	E = H90G10
.	.	.	.	.	۱۰	.	.	.	.	.	.	.	F = H80G20
.	.	۲	.	۲	.	۱۰	.	.	.	.	.	.	G = H70G30
.	.	.	.	.	.	.	۱۰	.	.	.	.	.	H = H90K10
.	.	.	.	.	.	.	.	۱۰	.	.	.	.	I = H80K20
.	۶	.	.	.	.	.	.	.	۱۰	.	.	.	J = H70K30
.	.	.	.	.	.	.	.	.	.	۱۰	.	.	K = H90S10
.	.	.	.	.	.	.	.	.	.	.	۱۰	.	L = H80S20
.	.	.	.	.	.	.	.	.	.	.	.	۱۰	M = H70S30
۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	حساسیت
۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	ویژه بودن



شکل ۴- نمودار درختی ایجاد شده توسط الگوریتم J48 برای طبقه‌بندی نمونه‌های برنج

### نتیجه‌گیری

پشتیبان به ترتیب با مقادیر صحت ۱۰۰، ۹۶/۹۲ و ۹۲/۳۱ درصد با دقت بالایی موفق به طبقه‌بندی کیفی و تشخیص اختلاط در برنج سفید بودند. با توجه به نتایج به‌دست آمده، هر سه الگوریتم یادگیری ماشین مورد مطالعه از دقت بالایی برای تشخیص و تمایز نمونه‌های برنج با درصد‌های اختلاط مختلف برخوردار بودند. با توجه به آماره‌های ارزیابی، الگوریتم درخت تصمیم از نوع J48 به‌عنوان بهترین روش برای طبقه‌بندی نمونه‌های برنج با

در این پژوهش، از روش طیف‌سنجی Vis/NIR به‌منظور طبقه‌بندی نمونه‌های مختلف برنج سفید براساس درصد‌های اختلاط مختلف با ارقام ناخالص استفاده شد. سه نوع از متداول‌ترین روش‌های طبقه‌بندی با الگوریتم‌های یادگیری ماشین مورد ارزیابی قرار گرفتند. نتایج نشان داد که الگوریتم‌های مبتنی بر روش‌های درخت تصمیم، شبکه‌های عصبی مصنوعی و ماشین بردار

- Champagne, E. T. (2008). Rice aroma and flavor: a literature review. *Cereal Chemistry*, 85(4): 445-454.
- Choudhury, P., Kohli, S., Srinivasan, K., Mohapatra, T., & Sharma, R. (2001). Identification and classification of aromatic rices based on DNA fingerprinting. *Euphytica*, 118(3): 243-251.
- Fu, H., Yin, Q., Xu, L., Wang, W., Chen, F., & Yang, T. (2017). A comprehensive quality evaluation method by FT-NIR spectroscopy and chemometric: Fine classification and untargeted authentication against multiple frauds for Chinese *Ganoderma lucidum*. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 182: 17-25.
- Gujjar, H. S., & Siddappa, D. M. (2013). A method for identification of basmati rice grain of india and its quality using pattern classification. *International Journal of Engineering Research and Applications*, 3(1): 268-273.
- Haghibin, N., Bakhshipour, A., Mousanejad, S., & Zareiforush, H. (2023). Monitoring *Botrytis cinerea* infection in kiwifruit using electronic nose and machine learning techniques. *Food and Bioprocess Technology*, 16(4), 749-767.
- Han, X., Lü, E., Lu, H., Zeng, F., Qiu, G., Yu, Q., & Zhang, M. (2020). Detection of spray-dried porcine plasma (SDPP) based on electronic nose and near-infrared spectroscopy data. *Applied Sciences*, 10(8), 2967.
- Hu, L., Yin, C., Ma, S., & Liu, Z. (2019). Vis-NIR spectroscopy Combined with Wavelengths Selection by PSO Optimization Algorithm for Simultaneous Determination of Four Quality Parameters and Classification of Soy Sauce. *Food analytical methods*, 12(3): 633-643.
- Huang, C., & Gu, Y. (2022). A machine learning method for the quantitative detection of adulterated meat using a MOS-based E-nose. *Foods*, 11(4): 602.
- Jana, A., Bandyopadhyay, R., Tudu, B., Roy, J. K., Bhattacharyya, N., Adhikari, B., Mukherjee, S. (2011). Classification of aromatic and non-aromatic rice using electronic nose and artificial neural network. *2011 IEEE Recent Advances in Intelligent Computational Systems*.
- Karray, F., Karray, F. O., & De Silva, C. W. (2004). Soft computing and intelligent systems design: theory, tools, and applications. Pearson Education.
- Kaur, H., & Singh, B. (2013). Classification and grading rice using multi-class SVM. *International Journal of Scientific and Research Publications*, 3(4): 1-5.
- درصدهای ناخالصی مختلف انتخاب شد. با توجه به نتایج حاصل از پژوهش می‌توان نتیجه گرفت که در حالت کلی به‌کارگیری داده‌های حاصل از طیف‌سنجی Vis/NIR در ترکیب با الگوریتم‌های فراابتکاری یادگیری ماشین از توانمندی بسیار مناسبی برای شناسایی و تمایز برنج اصیل (هاشمی) از ارقام برنج مخلوط شده (اختلاط هاشمی با ارقام دیگر) با دقت بالا برخوردار است. با توجه به آن که هم اکنون اختلاط ارقام مختلف برنج سفید با اهداف مختلفی در بازارهای داخلی این محصول در شمال کشور انجام می‌شود و تشخیص اختلاط به کمک روش‌های دیگر هم‌چون بینایی ماشین در مورد ارقام متنوع برنج سفید در شمال کشور کار راحتی نیست و از طرفی روش‌های بررسی ترکیبات شیمیایی در آزمایشگاه‌های استاندارد نیز زمان‌بر است، به‌کارگیری نتایج حاصل از پژوهش حاضر می‌تواند منجر به توسعه روش‌ها و ابزارهای سریع، دقیق و قابل حمل جهت تشخیص اختلاط در ارقام برنج سفید شود.

#### منابع

- Abdullah, A., Rahim, N., Masnan, M., Sa'ad, F., Zakaria, A., Shakaff, A., & Omar, O. (2016). Rice and the electronic nose. In *Electronic Noses and Tongues in Food Science* (pp. 103-113). Elsevier.
- Anwar, H., Anwar, T., & Murtaza, S. (2023). Review on food quality assessment using machine learning and electronic nose system. *Biosensors and Bioelectronics: X*, 100365.
- Bakhshipour, A., & Zareiforush, H. (2020a). Development of a fuzzy model for differentiating peanut plant from broadleaf weeds using image features. *Plant Methods*, 16(1): 1-16.
- Bakhshipour, A., & Zareiforush, H. (2020b). Development of a fuzzy model for differentiating peanut plant from broadleaf weeds using image features. *Plant Methods*, 16: 1-16.
- Bakhshipour, A., Sanaeifar, A., Payman, S. H., & de la Guardia, M. (2018). Evaluation of data mining strategies for classification of black tea based on image-based features. *Food Analytical Methods*, 11(4): 1041-1050.
- Ben-Gera, I., & Norris, K. H. (1968). Direct spectrophotometric determination of fat and moisture in meat products. *Journal of Food Science*, 33(1): 64-67.

- of rice varieties: A comparison of machine learning and deep learning approaches in analysing long-wave near-infrared hyperspectral images of brown and milled samples. *Infrared Physics & Technology*, 123: 104100.
- Pearce, T. C., Schiffman, S. S., Nagle, H. T., & Gardner, J. W. (2006). Handbook of machine olfaction: electronic nose technology. *John Wiley & Sons*.
- Qamar, M., Siyah Poush, M. R., & Hasibi, P. (2013). Salinity tolerance assessment of rice sucrose transporter antisense lines (OsSUT1) at seedling stage (*Oryza sativa* var. TaiPai). *Agricultural Biotechnology Journal*, 5(3): 87-98.
- Quiñones, M. d. C. S., Martínez, L. A. O., Herrera, S. M. G., Quiñones, O. M. R., Laredo, R. F. G., & y Bioquímica, Q. (2018). Near-Infrared Spectroscopy (NIRS) applied to legume analysis: A Review. *Spectroscopy*, 8(4).
- Rasooli Sharabiani, V., & Khorramifar, A. (2022a). Quantitative analysis of organic acids in potatoes using NIR spectroscopy with PLS and ANN methods. *Journal of Environmental Science Studies*, 7(4): 5541-5549.
- Rasooli Sharabiani, V., & Khorramifar, A. (2022b). Recognition and classification of pure and adulterated rice using the electronic nose. *Journal of Environmental Science Studies*, 7(2): 4904-4910.
- Shahgholi, G., Moeinfar, A., & Khorramifar, A. (2022). Close infrared spectroscopy to estimate the sugar content of different potato cultivars. *Journal of Environmental Science Studies*, 7(3): 5121-5128.
- Shao, Y., He, Y., Gómez, A. H., Pereir, A. G., Qiu, Z., & Zhang, Y. (2007). Visible/near infrared spectrometric technique for nondestructive assessment of tomato 'Heatwave' (*Lycopersicon esculentum*) quality characteristics. *Journal of Food Engineering*, 81(4): 672-678.
- Shi, S., Feng, J., Yang, L., Xing, J., Pan, G., Tang, J., Jiang, Y. (2023). Combination of NIR spectroscopy and algorithms for rapid differentiation between one-year and two-year stored rice. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 291: 122343.
- Sun, T., Xu, H. R., & Ying, Y. B. (2009). Progress in application of near infrared spectroscopy to nondestructive on-line detection of products/food quality. *Spectroscopy and Spectral Analysis*, 29(1), 122-126.
- Tarighi, J., & Khorramifar, A. (2023). Evaluation ability of the NIR-Spectroscopy
- Khorramifar, A., Karami, H., Wilson, A. D., Sayyah, A. H. A., Shuba, A., & Lozano, J. (2022a). Grape Cultivar Identification and Classification by Machine Olfaction Analysis of Leaf Volatiles. *Chemosensors*, 10(4): 125.
- Khorramifar, A., Rasekh, M., Karami, H., Covington, J. A., Derakhshani, S. M., Ramos, J., & Gancarz, M. (2022b). Application of MOS gas sensors coupled with chemometrics methods to predict the amount of sugar and carbohydrates in potatoes. *Molecules*, 27(11): 3508.
- Khorramifar, A., Sharabiani, V. R., Karami, H., Kisalaei, A., Lozano, J., Rusinek, R., & Gancarz, M. (2022c). Investigating Changes in pH and Soluble Solids Content of Potato during the Storage by Electronic Nose and Vis/NIR Spectroscopy. *Foods*, 11(24): 4077.
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8): 2674.
- Liu, C., Hao, G., Su, M., Chen, Y., & Zheng, L. (2017). Potential of multispectral imaging combined with chemometric methods for rapid detection of sucrose adulteration in tomato paste. *Journal of Food Engineering*, 215: 78-83.
- Liu, Q., Zhao, N., Zhou, D., Sun, Y., Sun, K., Pan, L., & Tu, K. (2018). Discrimination and growth tracking of fungi contamination in peaches using electronic nose. *Food chemistry*, 262, 226-234.
- Liu, W., Xu, X., Liu, C., & Zheng, L. (2021). Nondestructive Detection of Authenticity of Thai Jasmine Rice Using Multispectral Imaging. *Journal of Food Quality*, 2021: 6642220.
- Mollazade, K., Omid, M., & Arefi, A. (2012). Comparing data mining classifiers for grading raisins based on visual features. *Computers and Electronics in Agriculture*, 84: 124-131.
- Nayak, P. P., Nandi, S., Bhunia, K., & Datta, A. K. (2023). Modelling the extraction process parameters of amorphous silica-rich rice husk ash using hybrid RSM- BPANN- MOGA optimization technique. *Materials Chemistry and Physics*, 293, 126944.
- Omid, M., Mahmoudi, A., & Omid, M. H. (2010). Development of pistachio sorting system using principal component analysis (PCA) assisted artificial neural network (ANN) of impact acoustics. *Expert Systems with Applications*, 37(10): 7205-7212.
- Onmankhong, J., Ma, T., Inagaki, T., Sirisomboon, P., & Tsuchikawa, S. (2022). Cognitive spectroscopy for the classification

- for detecting the ripening time of walnuts. *Journal of Environmental Science Studies*, 8(1): 5850-5856.
- Van Nguyen, N., & Ferrero, A. (2006). Meeting the challenges of global rice production. In (Vol. 4, pp. 1-9): *Springer*.
- Wang, X., Bouzembrak, Y., Lansink, A. O., & van der Fels-Klerx, H. (2022). Application of machine learning to the monitoring and prediction of food safety: A review. *Comprehensive Reviews in Food Science and Food Safety*, 21(1): 416-434.
- Williams, P. (2019). Near Infrared Technology: Getting the best out of light. *AfricanSun Media*.
- Wongpornchai, S., Dumri, K., Jongkaewwattana, S., & Siri, B. (2004). Effects of drying methods and storage time on the aroma and milling quality of rice (*Oryza sativa* L.) cv. Khao Dawk Mali 105. *Food chemistry*, 87(3): 407-414.
- Zareiforush, H., Minaei, S., Alizadeh, M. R., & Banakar, A. (2016). Qualitative classification of milled rice grains using computer vision and metaheuristic techniques. *Journal of food science and technology*, 53: 118-131.
- Zhu, L., Spachos, P., Pensini, E., & Plataniotis, K. N. (2021). Deep learning and machine vision for food processing: A survey. *Current Research in Food Science*, 4: 233-249.

Research paper

## Classification of pure and mixed white rice cultivars using VIS-NIR spectroscopy and machine learning

H. Zaresani<sup>1</sup>, A.H. Afkari-Sayyah<sup>2</sup> and H. Zareiforoush<sup>3\*</sup>

### Abstract

Rice is the staple food of about 2.5 billion people in the world. Various factors influence the quality of this product. As a result, quality control and mixing detection are important issues in the rice industry. This research aimed to investigate and detect mixing in white rice cultivars using Vis/NIR spectroscopy, chemometrics, and machine learning. To conduct experiments, 13 classes were prepared based on the amount of mixing of the main variety of rice (Hashmi variety) with non-main varieties of Fajr, Gilaneh, Khazar, and Shiroudi so that the gross classes resulting from mixing 90, 80, and 70% of Hashmi variety rice respectively with 10, 20 and 30% of other figures mentioned. Multilayer perceptron (MLP), decision tree (DT), and support vector machines (SVM) algorithms were used to classify rice samples. To evaluate the performance of the studied classifiers, statistical indices, including sensitivity (Se), specificity (Sp), accuracy (Ac), and root mean square error (RMSE), were used in the form of developing asymmetry matrices of the classifiers. The results of the study showed that J48 DT, MLP, and SVM with RBF kernel function with accuracy and RMSE values of 100% and zero, 96.92% and 0.0951, 92.31%, and 0.2483%, respectively, were able to classify different samples of white rice.

**Keywords:** Rice, Grading, Artificial Neural Networks, Decision Tree, Visible-Near-Infrared Spectroscopy.

**Citation:** Zaresani H. Afkari-Sayyah A. H. and Zareiforoush H. 2024. Classification of pure and mixed white rice cultivars using VIS-NIR spectroscopy and machine learning. Journal of Researches in Mechanics of Agricultural Machinery. 32: ??-??. <https://doi.org/10.22034/JRMAM.2024.14748.704>

1- Ph.D. Candidate, Department of Biosystems Engineering, Faculty of Agriculture, University of Mohaghegh Ardabili, Ardabil, Iran.

2- Assistant Professor, Department of Biosystems Engineering, Faculty of Agriculture, University of Mohaghegh Ardabili, Ardabil, Iran.

3- Assistant Professor, Department of Biosystems Engineering, Faculty of Agricultural Sciences, University of Guilan, Rasht, Iran.

\* Corresponding Author: [hemad.zareiforoush@guilan.ac.ir](mailto:hemad.zareiforoush@guilan.ac.ir)

Received: 2024/04/21

Accepted: 2024/07/01

<https://doi.org/10.22034/JRMAM.2024.14748.704>