

# امکان سنجی کاربرد روش غیرمخرب طیفسنجی فروسرخ نزدیک و مدلسازی PLSR در تخمین پروتئین دانه‌های گندم و پهنه‌بندی نقشه کیفی مزارع

بهنام فروزانی<sup>1</sup>، حسین باقرپور<sup>2\*</sup>، خلیل زابلی<sup>3</sup>

1- دانش آموخته کارشناسی ارشد مهندسی مکانیک بیوسیستم، دانشگاه بوعلی سینا، همدان

2- استادیار، گروه مهندسی مکانیک بیوسیستم، دانشکده کشاورزی، دانشگاه بوعلی سینا، همدان

3- استادیار، گروه علوم دامی، دانشکده کشاورزی، دانشگاه بوعلی سینا، همدان

(تاریخ دریافت: 98/01/28 تاریخ پذیرش: 98/11/26)

## چکیده

پروتئین دانه‌های گندم از شاخص‌های اصلی کیفی محصول هستند و نقش به سزایی در فراوری این محصول ایفا می‌کند. با توجه به فراوری محصولات متنوع از گندم، اندازه‌گیری سریع یا برخط کفی محصول گندم در کنترل فرآیندهای تولید آرد یا انتخاب رقم محصول گندم بسیار مهم می‌باشد. از طرفی در مفهوم جدید کشاورزی دقیق، با اندازه‌گیری مستقیم پروتئین و یا رطوبت محصول بر روی کمباین و تهیه نقشه عملکرد کیفی مزرعه، می‌توان وضعیت مزرعه در نقاط مختلف و یا مقایسه مزارع با همدیگر را نیز ارزیابی کرد. بنابراین هدف اصلی این پژوهش ارزیابی روش غیرمخرب طیفسنجی فروسرخ نزدیک حالت بازتابی جهت پیش‌بینی متغیر پروتئین و رطوبت دانه‌های سالم گندم بود. در این تحقیق 108 نمونه سالم گندم از سه رقم میهن، پیشگام و گاسکوژن در مرحله قبل از برداشت از مزارع روستای کرفس در استان همدان تهیه شدند. طیف‌گیری از نمونه‌ها به روش بازتابی و در محدوده دامنه 950-1650 نانومتر و قدرت تفکیک 5 نانومتر انجام شد. نتایج نشان داد که بهترین مدل با استفاده از روش PLSR و پیش‌پردازش ترکیبی  $SG+SNV+D1$  و  $MA+D2+SNV$  به ترتیب برای پروتئین و رطوبت بدست آمد. مقادیر ضریب رگرسیون  $(R^2)$ ، RMSE و SDR داده‌های اعتبارسنجی با تست به ترتیب برابر با 0/83، 0/84 و 2/54 برای تخمین پروتئین و 0/994، 0/96 و 5/34 برای تخمین رطوبت حاصل شد. نتایج نشان داد که هر چند تفاوت معنی داری بین میانگین پروتئین رقم‌ها یافت نشد، ولی اختلاف آماری معنی داری در سطح 5 درصد بین نقاط مختلف مزارع بدست آمد. نتایج نشان داد که روش طیفسنجی فروسرخ نزدیک و مدل PLSR روشی کارآمد و دارای پتانسیل قوی برای تشخیص سریع پارامترهای پروتئین و رطوبت دانه‌های گندم می‌باشد.

کلید واژگان: روش غیرمخرب، گندم، فروسرخ نزدیک، پروتئین

\* مسئول مکاتبات: h.bagherpour@basu.ac.ir

## 1- مقدمه

غلات و حبوبات نقش ویژه‌ای در تأمین غذای مردم، چه در کشورهای پیشرفته و چه در کشورهای در حال توسعه ایفا می‌کنند. همچنین به دلیل سرشار بودن از انواع پروتئین‌ها و مواد معدنی یک منبع غذایی مهم به شمار می‌روند. گندم در دنیا بالاترین سطح زیر کشت و تولید را به خود اختصاص داده و به‌طور متوسط 22 درصد کل پروتئین در رژیم غذایی انسان را فراهم می‌کند [1]. ترکیبات مهم دانه گندم معمولاً شامل: 7 تا 18 درصد پروتئین (بستگی به نوع گندم)، 60 تا 70 درصد نشاسته، تا 2/5 درصد سلولز (فیبر خام)، 1/5 تا 2 درصد چربی و مابقی مرکب از رطوبت و مواد کانی است.

وجود کمیت و کیفیت مختلف ترکیبات شیمیایی مثل نشاسته، پروتئین و غیره در گندم باعث می‌گردد تا بتوان محصولات مختلفی از انواع گندم تهیه کرد. پروتئین یکی از اجزای مهم دانه گندم بوده و به دلیل تأثیرگذاری آن بر کیفیت نان از اهمیت ویژه‌ای برخوردار است. پروتئین گندم حاوی اسیدآمینه‌های ضروری بدن است و بخش مهم و قابل توجهی از پروتئین مورد نیاز بدن را تأمین می‌کند. از این رو بررسی کمیت و کیفیت آن در فرآورده‌های تولید و یا پخت نان از اهمیت ویژه‌ای برخوردار است [2].

یکی از مهم‌ترین عامل‌ها در تعیین زمان برداشت محصول گندم، رطوبت آن است. پایین یا بالا بودن رطوبت، باعث افزایش خسارات به محصول در زمان برداشت و انبارداری شده، و از طرفی می‌تواند تا حدودی هزینه‌های فرآورده‌های پس از برداشت را زیاده‌تر کند. انبار گندم با رطوبت بیشتر از 13 تا 14 درصد، باعث تشدید واکنش‌های شیمیایی و بیوشیمیایی شده و همچنین شرایط مناسبی را برای رشد قارچ و حشرات در سیلوها فراهم می‌کند. از طرفی، برای کاهش تلفات زمان برداشت، معمولاً رطوبت دانه را بین 14 الی 16 درصد در نظر می‌گیرند [3].

روش‌های اندازه‌گیری کیفیت محصولات کشاورزی را می‌توان به دو بخش عمده‌ی مخرب و غیرمخرب طبقه‌بندی کرد. از مزایای اندازه‌گیری‌های مخرب می‌توان به دقت بالا و استاندارد بودن جهانی اشاره نمود؛ اما بدیهی است که روش‌های استاندارد و دقیق، به محیط و شرایط استاندارد و دقیق نیازمندند که این خود معایبی چون مشکل و زمان‌بر بودن آزمایش‌ها،

هزینه‌ی زیاد و تلفات اقتصادی ناشی از تخریب نمونه را به همراه دارد [4].

در سال‌های اخیر، روش‌هایی همچون ماشین بینایی، فراصوتی، طیف‌سنجی فروسرخ نزدیک (NIR)، رزونانس مغناطیسی هسته‌ای (NMR)، رامان و تصویربرداری فراطیفی و غیره، به عنوان روش‌های غیرمخرب در حال گسترش و توسعه می‌باشند. طیف‌سنجی NIR در مقایسه با برخی روش‌های غیرمخرب دیگر که تنها قادر به نمایش ساختار درونی مواد هستند، توانایی اندازه‌گیری نوع و مقدار ترکیبات مواد بیولوژیکی را نیز دارد [5]. این روش در مقایسه با روش‌های مخرب هزینه کمتری داشته و می‌توان با تدوین یک مدل کالیبراسیون، تعداد زیادی نمونه را بدون نیاز به آماده‌سازی قبلی مورد بررسی و آزمایش قرار داد [6].

پژوهش‌های انجام گرفته در زمینه‌ی کاربرد طیف‌سنجی NIR را می‌توان به دو دسته تقسیم کرد. دسته‌ی اول شامل تحقیقاتی است که هدف پژوهشگر، ارائه‌ی یک مدل رگرسیونی برای تخمین و یا پیشگویی میزان یک پارامتر می‌باشد. در بیشتر این تحقیقات از مدل‌های رگرسیون خطی شامل رگرسیون حداقل مربعات جزئی (PLSR) و رگرسیون مؤلفه‌های اصلی (PCR) استفاده می‌گردد [7]. دسته‌ی دوم از این تحقیقات، شامل پژوهش‌هایی است که در آن‌ها هدف، دسته‌بندی و یا طبقه‌بندی محصولات کشاورزی با استفاده از طیف‌سنجی NIR است. طیف NIR یک ترکیب می‌تواند اطلاعات مهمی در مورد ساختار مولکولی و طبیعت شیمیایی ماده ارائه کند؛ با این حال همیشه استخراج داده‌ها به طور مستقیم از طیف ممکن نیست؛ بنابراین، روش‌های آماری چند متغیره برای استخراج اطلاعات مفید از یک طیف NIR نیاز می‌باشد [8].

در پژوهشی که به کمک طیف‌سنجی NIR و در ترکیب با شبکه عصبی مصنوعی، پروتئین و رطوبت آرد گندم اندازه‌گیری شد، مدل شبکه عصبی توانست این دو متغیر را با مقدار  $R^2$  به ترتیب 0/95 و 0/92 تخمین بزند [9]؛ در این پژوهش، پارامترهای کیفی مختلفی مثل مقدار پروتئین، رطوبت، مقدار جذب آب، زمان آماده‌سازی خمیر، درجه نرمی خمیر، پایداری و سایر ویژگی‌های فیزیکی در محدوده طیف 400 - 2498 nm مورد بررسی و ارزیابی قرار گرفت. همچنین در پژوهشی دیگر، با بهره‌گیری از طیف‌سنج بازتابی فروسرخ نزدیک، میزان رطوبت، پروتئین، نشاسته و آمیلوز آرد گندم تعیین شد [10].

تعیین سریع مؤلفه‌های اندازه، وزن و رطوبت وجود دارد، اما هنوز اطلاعات کافی برای تشخیص سریع و غیرمخرب کیفیت محصول خام، به‌ویژه محصولات دانه‌ای، وجود ندارد. لذا در این پژوهش پتانسیل سنجی تشخیص مقدار پروتئین و رطوبت با بهره‌گیری از تکنیک طیف‌سنجی مادون قرمز نزدیک مورد بررسی قرار گرفت.

## 2- مواد و روش‌ها

### 2-1- تهیه نمونه

در این تحقیق برای بررسی دو متغیر رطوبت و پروتئین، نمونه‌های مورد آزمایش از مزارع روستای کرفس در استان همدان با عرض "04' 21' 35° و طول جغرافیایی "29' 19' 49° تهیه شدند. برای این در کل تعداد 108 نمونه از مزارع گندم در مرحله قبل از برداشت و با فاصله‌های 25 متری از همدیگر در 6 قسمت مختلف مزرعه جمع‌آوری شدند. نمونه‌ها از 6 مزرعه مختلف و از سه رقم میهن، پیشگام و گاسکوژن تهیه شد. پس از آن دانه‌های گندم از خوشه جدا شدند و نمونه‌ها از مواد خارجی و کاه و کلش پاک‌سازی شدند. برای نمونه‌هایی که فقط میزان رطوبتشان محاسبه گردید، به اندازه یک کیلوگرم گندم از یک رقم تهیه شد. اندازه‌گیری رطوبت با استفاده از استاندارد ملی (کد 2705) صورت گرفت. برای ایجاد نمونه‌هایی با رطوبت مختلف، نمونه گندم به تعداد 6 بخش مساوی تقسیم شد و با اضافه کردن مقدار مشخصی از آب، 6 سطح رطوبتی با سه تکرار شامل رطوبت‌های 4، 8، 11، 14، 18 و 22 درصد به‌دست آمد. برای این کار ابتدا رطوبت اولیه کل نمونه محاسبه گردید که برای نمونه‌های مختلف بین 10 الی 11% بدست آمد. برای ایجاد سطوح رطوبتی بالاتر، مقدار مشخص و محاسبه شده آب، به هر نمونه اضافه شد تا به میزان رطوبت مورد نظر برسند. رطوبت‌های پایین‌تر از شاهد نیز با گذاشتن در آون بدست آمد. برای هر سطح رطوبتی چهار تکرار در نظر گرفته شد.

### 2-2- طیف‌سنجی از نمونه‌ها

طیف‌گیری از نمونه‌ها با استفاده از دستگاه اسپکترومتر مدل Perten DA7200 ساخت سوئیس و به روش بازتابی انجام شد. این طیف‌سنج دارای دامنه کاری 950-1650 نانومتر و قدرت تفکیک 5 نانومتر است که مجهز به آرایه‌های دیودی با

محققین در تحقیق دیگری با همین روش توانستند محتوای پروتئین غلات غذایی مختلف را در محدوده طیف nm 2500-1100 پیش‌بینی کنند [11]. در این پژوهش که بر روی چندین محصول مختلف انجام شد، با استفاده از مدل PLS، ارتباط مناسبی و با ضریب تعیین 0/97 و خطای استاندارد 0/09 درصد بین مقدار نیتروژن اندازه‌گیری شده و خروجی مدل کالیبراسیون بدست آمد. در تحقیقی که به منظور اندازه‌گیری غلظت پروتئین گندم با طیف‌سنجی عبوری NIR انجام شد، نتایج حاصل از مدل PLS بسیار مطلوب (0/95 <R<sup>2</sup> < 0/85 و 0/9 <SEP < 0/4) گزارش شد [12]. در پژوهش دیگر که بر روی گندم انجام شد، توسط طیف‌سنجی فروسرخ نزدیک، خواص پروتئین گندم آسیاب شده و رطوبت دانه گندم مورد ارزیابی قرار گرفت [13]. هدف اصلی این پژوهش مقایسه 11 مدل از دستگاه طیف سنج فروسرخ نزدیک در تشخیص مقدار پروتئین و رطوبت گندم بود. نتایج تحقیق آنها نشان داد که برای مدل‌های مختلف اگر دستگاه‌ها بخوبی تنظیم شوند، می‌توانند مقدار پروتئین را با دقت  $\pm 0/5\%$  نسبت به مقدار واقعی نشان دهند. از طرفی رطوبت بدست آمده با مدل‌های مختلف در مقایسه با روش خشک کردن با آون با میانگین دقت 0/39% گزارش گردید.

با توجه به اهمیت اندازه‌گیری میزان پروتئین دانه‌های گندم در پروسه‌های مختلف تولید و تهیه آرد و همچنین کاهش هزینه آزمایش‌ها و از طرفی کاهش خسارت‌های حاصل از رطوبت گندم، لازم است که این دو متغیر در مراحل مختلف از برداشت محصول تا رسیدن به یک محصول نهایی مورد سنجش و ارزیابی قرار گیرند. هر چند تحقیقاتی در زمینه سنجش پروتئین گندم به روش طیف‌سنجی انجام شده ولی اکثر پژوهش‌ها انجام شده بر روی آرد گندم بوده و از طرفی گزارشی در مورد اندازه‌گیری مستقیم پروتئین دانه‌های گندم در ایران یافت نگردید [7]. سنجش پروتئین گندم علاوه بر کنترل فرآورده‌های تولید، اگر در زمان برداشت و مستقیماً در روی کمباین انجام گیرد، می‌تواند گام بزرگی برای کشاورزان در جهت مدیریت مزرعه باشد. همچنین تشخیص مقدار رطوبت گندم بر روی کمباین، می‌تواند کمک شایانی در کاهش تلفات زمان برداشت و یا مدیریت بعدی فرآوری و انبار داری محصول داشته باشد. بنابراین تشخیص بلادرنگ رطوبت دانه‌ها از اهمیت بسیاری برخوردار است. با اینکه دانش کافی برای

(PCA) مشخص و حذف شدند. سپس برای مدل‌سازی، نمونه‌های باقی‌مانده به‌صورت تصادفی به دو دسته آموزش (70% نمونه‌ها) و آزمون (30% نمونه‌های باقی‌مانده) به ترتیب برای تدوین مدل واسنجی و پیش‌گویی تقسیم شدند. در این تحقیق روش اعتبارسنجی متقاطع کامل برای ارزیابی مدل‌های واسنجی و پیش‌گویی به کار رفت. مدل‌های تدوین شده توسط ضریب رگرسیون ( $R^2$ )، ریشه میانگین مربعات خطا (RMSE) و نسبت انحراف معیار به این خطا (SDR) ارزیابی شدند [16 و 17]. این سه ضریب به ترتیب بر اساس روابط 1، 2 و 3 محاسبه می‌شوند.

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad \text{رابطه (1)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad \text{رابطه (2)}$$

$$SDR = \frac{SD}{RMSEP} \quad \text{رابطه (3)}$$

که در این روابط RMSE ریشه میانگین مربعات خطای پیش‌بینی،  $y_i$  مقدار اندازه‌گیری شده صفت مورد نظر  $\hat{y}_i$  مقدار پیش‌بینی صفت مورد نظر برای نمونه  $i$  هنگامی که مدل بدون نمونه  $i$  ساخته شده است،  $N$  تعداد نمونه آزمون،  $y_m$  میانگین مقادیر اندازه‌گیری شده صفت، SDR نسبت انحراف معیار، SD انحراف معیار صفت مورد نظر هستند. مقدار SDR بین 2/5 تا 3 و بیشتر از 3 به ترتیب بیانگر دقت‌های پیش‌گویی خوب و عالی مدل است [17].

انتخاب مدل بهینه از بین مدل‌های تدوین شده برای پیش‌گویی پارامتر پروتئین گندم بر پایه داشتن RMSEP کمتر،  $R^2$  و SDR بیشتر انجام شد.

تمام تحلیل‌های آماری مربوط به پیش‌پردازش طیف‌ها و تدوین مدل با استفاده از نرم‌افزار The Unscrambler X 10.4 انجام شد.

### 3- بحث و نتایج

به دلیل وجود پیک‌های پهن و پیک‌های جذبی کوچک‌تر که هر کدام بیانگر قسمتی از نمونه هستند کسب اطلاعات دقیق به صورت مستقیم از طیف خام دشوار است. همچنین بعضی پیک‌ها در طیف خام به دلیل وجود نویز قابل رؤیت نیستند که پس از اعمال پیش‌پردازش‌ها قابل رؤیت می‌شوند؛ بنابراین برای دستیابی به جزئیات جذبی بیشتر که در این طیف‌ها به

آشکار ساز ایندیم-گالیم آرسناید (InGaAs) می‌باشد. به منظور طیف‌گیری از نمونه‌ها، هیچ‌گونه آماده‌سازی بر روی نمونه‌ها انجام نشد و تنها بعد از جداسازی مواد اضافی از آنها، نمونه‌ها برای طیف‌سنجی آماده شدند. 108 نمونه موجود به ترتیب برای طیف‌سنجی شماره‌بندی، و در داخل کیسه‌های پلاستیکی قرار داده شدند. بعد از کالیبراسیون دستگاه، برای هر نمونه دو اسکن از آن گرفته شد که در نهایت میانگین طیف‌های اسکن شده توسط نرم‌افزار دستگاه به عنوان طیف آن نمونه لحاظ گردید [14]. پس از انجام مرحله طیف‌گیری، نمونه‌ها بلافاصله به آزمایشگاه تغذیه دام دانشگاه بوعلی سینا منتقل و نسبت به اندازه‌گیری ویژگی‌های کیفی نمونه‌ها اقدام شد. اندازه‌گیری پروتئین گندم به روش کج‌لدال و با استفاده از استاندارد ملی به شماره 19052 انجام شد.

### 2-3- پیش‌پردازش داده‌ها

داده‌های طیفی به دست آمده از دستگاه طیف‌سنج علاوه بر دارا بودن اطلاعات نمونه، شامل اطلاعات ناخواسته، پس‌زمینه‌ها، خواص فیزیکی و نویزها نیز هستند. با توجه به موارد ذکر شده، برای تدوین مدل‌های رگرسیونی قابل اعتماد نیاز به پیش‌پردازش داده‌های طیفی پیش از تدوین مدل‌های رگرسیونی است [13]. کاربرد روش‌های پیش‌پردازش برای کاهش تغییرات طیف در اثر تغییرات فیزیکی ناشناخته بسیار معمول است و در این تحقیق به‌منظور آماده‌سازی و پیش‌پردازش طیف‌ها، از ترکیب برخی روش‌های مختلف هموارسازی و کاهش نویز، مشتق‌گیری، نرمال‌سازی و افزایش قدرت تفکیک طیفی استفاده شد.

### 2-4- مدل‌سازی خطی

هدف از به‌کارگیری روش‌های مدل‌سازی رگرسیون چندمتغیره، ایجاد یک ارتباط بین ویژگی‌های کیفی اندازه‌گیری شده و داده‌های طیفی است. از این رو، در این پژوهش برای پیش‌گویی پروتئین و رطوبت از دو روش مدل‌سازی رگرسیون چندمتغیره خطی بر پایه کاهش تعداد متغیرهای طیفی شامل رگرسیون مؤلفه‌های اصلی (PCR) و حداقل مربعات جزئی (PLSR) استفاده شد. در این راستا، ابتدا برای کل طیف‌های گرفته شده از نمونه‌ها، نمونه‌های پرت از نظر داده‌های طیفی که ممکن است به دلیل مشکل‌های فنی مربوط به تجهیزات اسپکتروسکوپی، جمع‌آوری داده، نمونه‌گیری نادرست و غیره به وجود آمده باشند، با استفاده از روش تجزیه مؤلفه‌های اصلی

طور نامحسوس وجود دارند مشتق‌های اول و دوم و پیش‌پردازش‌های دیگر طیف‌ها بررسی شدند.

همان‌طور که پیش از این بیان شد، برای به دست آوردن مدل‌های کالیبراسیون قابل اعتماد با بالاترین دقت پیش‌گویی شاخص‌های پروتئین و رطوبت گندم، روش‌های پیش‌پردازش به‌صورت ترکیبی بر روی طیف‌ها اعمال شدند و در نهایت مدلی که بیشترین مقدار  $R^2$  و SDR و پایین‌ترین مقدار RMSE را نتیجه داد، به عنوان بهترین مدل گزارش شد. برخی از نتایج مدل‌های تدوین شده برای پروتئین و رطوبت با روش‌های PCR و PLSR بر پایه ترکیب پیش‌پردازش‌های مختلف به ترتیب در جداول (1) و (2) آورده شده است.

در اعتبار سنجی مدل‌های مربوط به پروتئین گندم، با مقایسه خروجی مدل با مقادیر مرجع داده‌های تست، بهترین مدل با استفاده از روش PLSR و پیش‌پردازش ترکیبی SG+SNV+D1 به دست آمد. در این مدل ضرایب  $R^2$ ، SDR و RMSEP برای مدل پیش‌بینی به ترتیب برابر با 0/84، 2/54 و 0/835 به دست آمد. اعتبارسنجی مدل‌ها نشان

داد که مدل‌هایی که با روش PLSR ایجاد شده‌اند نسبت به روش PCR، به دلیل بالاتر بودن مقادیر  $R^2$  و SDR و همچنین پایین‌تر بودن مقدار RMSEP (ریشه میانگین مربعات خطای داده‌های تست) قابل اعتمادتر هستند. با توجه به مقادیر شاخص SDR، در بیشتر مدل‌های رگرسیون چند متغیره تدوین شده به روش PLSR، مقدار SDR از 2 بالاتر بود که این بیان می‌کند امکان تشخیص و پیش‌بینی پروتئین نمونه‌ها با دقت بالا وجود دارد. نتایج جدول (1) نشان می‌دهد که مدل‌هایی که به روش PCR ایجاد شدند به دلیل اینکه مقدار شاخص SDR برای تعداد کمی از مدل‌ها بالاتر از 2 بود مدل‌هایی با مطلوبیت متوسط می‌باشند و در تخمین مقدار پروتئین نسبت به مدل PLSR ضعیف‌تر عمل می‌کنند. نتایج این تحقیق با نتایج ( $R^2=0.91$ ;  $SEP=3.1$ ) پژوهش مشابهی که با استفاده از روش طیف‌سنجی و مدل PLS، پروتئین دانه‌های گندم بر روی کمباین را اندازه‌گیری کردند [18]، همخوانی دارد.

Table 1 Models of NIR spectrometry to determine wheat protein

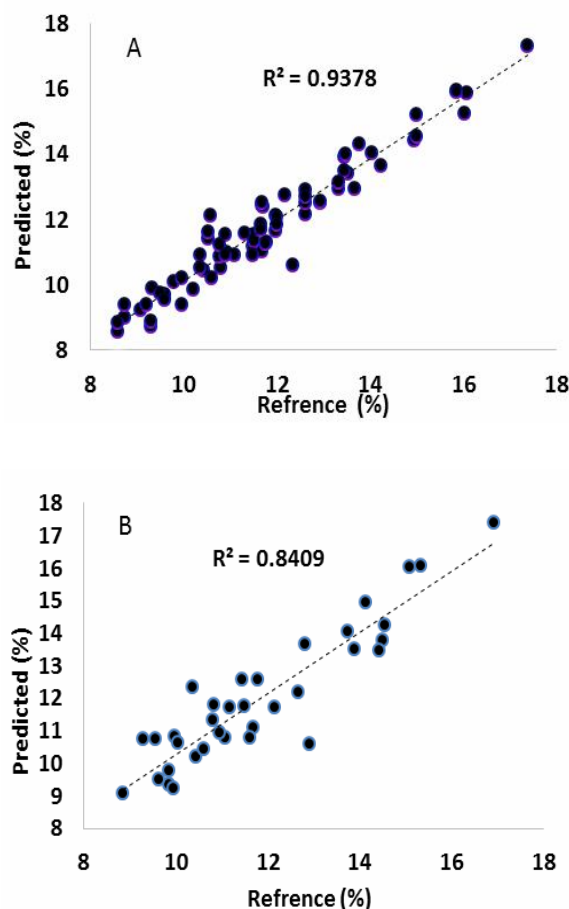
SDR	RMSEP	$R_p^2$	RMSEC	$R_c^2$	PCs	Model	Preprocessing
1.68	1.26	0.727	0.90	0.792	8	PCR	SG+SNV
2.02	0.98	0.768	0.78	0.851	8		MA+D1+SNV
1.59	1.38	0.719	0.93	0.786	8		SG+D2+SNV
2.10	0.97	0.774	0.73	0.861	6		MA+D1+MSC
1.62	1.10	0.758	0.78	0.845	6	PLSR	SG
1.91	1.01	0.774	0.79	0.840	5		SG+D1
2.38	0.87	0.833	0.63	0.897	8		MA+D1+SNV
2.44	0.84	0.839	0.60	0.908	8		D1+SNV
2.13	0.92	0.804	0.59	0.911	9		NOR+BL+D1
<b>2.54</b>	<b>0.83</b>	<b>0.840</b>	<b>0.49</b>	<b>0.937</b>	<b>8</b>		<b>SG+SNV+D1</b>

شکل 1 ضریب رگرسیونی طیف بهترین مدل برای پروتئین گندم را نشان می‌دهد. در مجموع، چندین پیک در طیف حاصل از مدل‌سازی مشاهده شد. نسبت دادن برخی از این طول‌موج‌ها به باندهای جذبی مربوط به پروتئین منطقی است. همان‌گونه که در شکل مشاهده می‌شود، محدوده طیفی

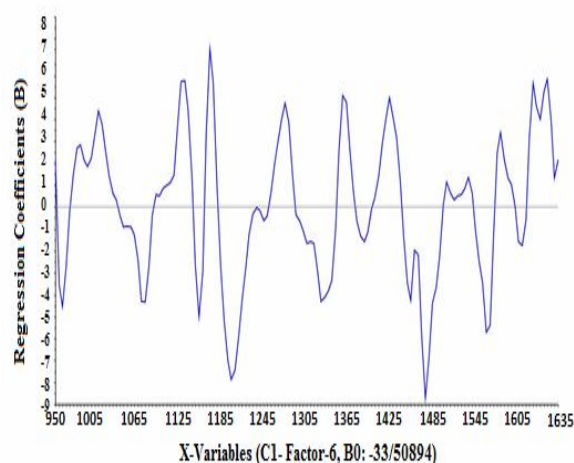
1465 نانومتر بالاترین ضریب رگرسیون را با مقادیر داده‌های پروتئین مرجع دارد که می‌تواند مربوط به باند جذبی پروتئین (عموماً در 1430-1571 nm) باشد. این محدوده با نتایج پژوهش دلویچ<sup>1</sup> (1998) هم‌خوانی دارد [12].

1. Delwiche

خطا (RMSE) به ترتیب برابر با 0/85 و 0/86 گزارش شده بود. در بررسی میزان پروتئین رقم‌های مختلف، با مقایسه میانگین‌ها که به روش دانکن انجام شد، نتایج نشان داد که اختلاف آماری معنی‌داری بین دو رقم پیشگام، میهن وجود نداشت ولی اختلاف معنی‌داری بین محصول گاسکوژن با دو رقم دیگر یافت شد. همچنین در بررسی اثر محل نمونه‌برداری بر میزان پروتئین، داده‌های آماری نشان داد که اختلاف معنی‌داری بین نقاط مختلف در یک مزرعه وجود دارد. این نتیجه از این نظر حائز اهمیت می‌باشد که بعضی از نقاط مزرعه به دلیل نوع خاک و یا مواد ریز مغذی تاثیر قابل توجهی بر عملکرد کیفی محصول داشته است. بنابراین می‌توان پیشنهاد کرد که در صورت نصب سکوی طیف‌سنجی مناسب بر روی کمباین و طیف‌سنجی نمونه‌ها، امکان تهیه نقشه عملکرد کیفی و مدیریت دقیق مزرعه وجود دارد.



**Fig 2** Relationship between predicted and measured value of wheat seed protein A) Calibration data B) Test data

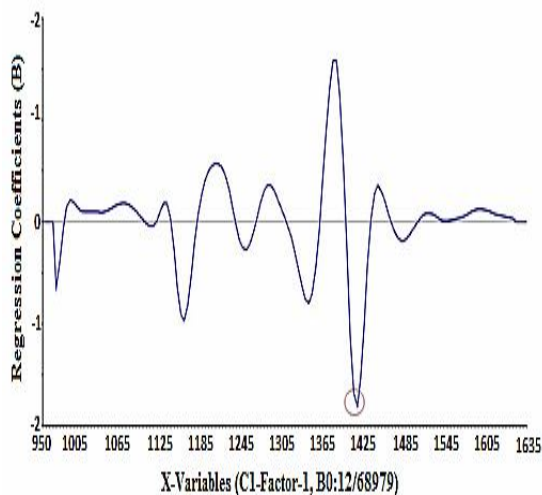


**Fig 1** Regression coefficient of PLSR model at different wavelength for prediction of protein

انتخاب تعداد مولفه‌های اصلی بر اساس تغییرات واریانس تبیین شده داده‌های ارزیابی (Validation Check) انجام شد. در این روش اگر بعد از تعداد مولفه اصلی مشخصی، مقدار واریانس تبیین برای چندین داده ارزیابی بعدی، دیگر تغییر نکند یا مقدار آن کاهش یابد، همان تعداد مولفه مشخص شده به عنوان مولفه اصلی برای مدل انتخاب می‌گردد. تعیین تعداد مولفه اصلی به کمک نرم افزار و تنظیمات آن انجام گردید. برای تخمین میزان پروتئین از 8 مولفه اصلی در این مدل استفاده شد که همین تعداد فاکتور توانست نزدیک به 90 درصد از واریانس‌های متغیر وابسته را توصیف کند.

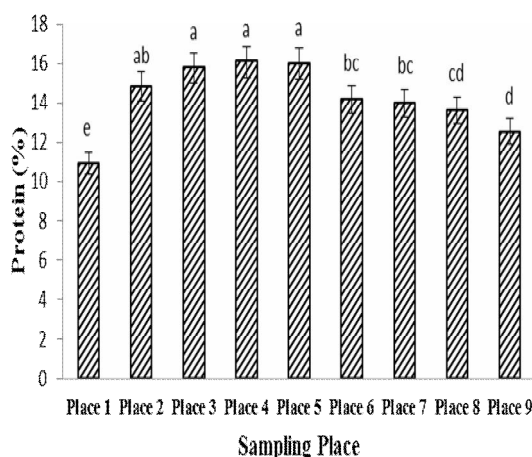
در شکل 2 ارتباط بین مقادیر واسنجی و پیش‌بینی، برای داده‌های کالیبراسیون و تست، در بهترین مدل انتخاب شده نشان داده شده است. همان‌طور که مشاهده می‌شود مدل مربوطه توانست مقدار پروتئین را با توانایی بالایی تخمین بزند. نتایج این پژوهش با نتایج موتلو و همکاران<sup>1</sup>، 2011 که با استفاده از مدل شبکه عصبی مصنوعی مقدار پروتئین و رطوبت آرد گندم را به ترتیب برابر با ضریب تعیین 0/95 و 0/93 گزارش کرده بودند مطابقت دارد [9]. از طرفی در مقایسه با روش‌های تصویربرداری فراطیفی که در آن مقدار پروتئین تک دانه‌های گندم، در محدوده طول موج 980-2500 nm تصویربرداری شده بود [19]، تخمین مناسب‌تری بدست آمد. در تصویربرداری حرارتی مقدار  $R^2$  و ریشه میانگین مربعات

نتایج حاصل از مدل رگرسیونی مربوط به نمودارهای کالیبراسیون و تخمین رطوبت در شکل 5 نشان داده شد است.

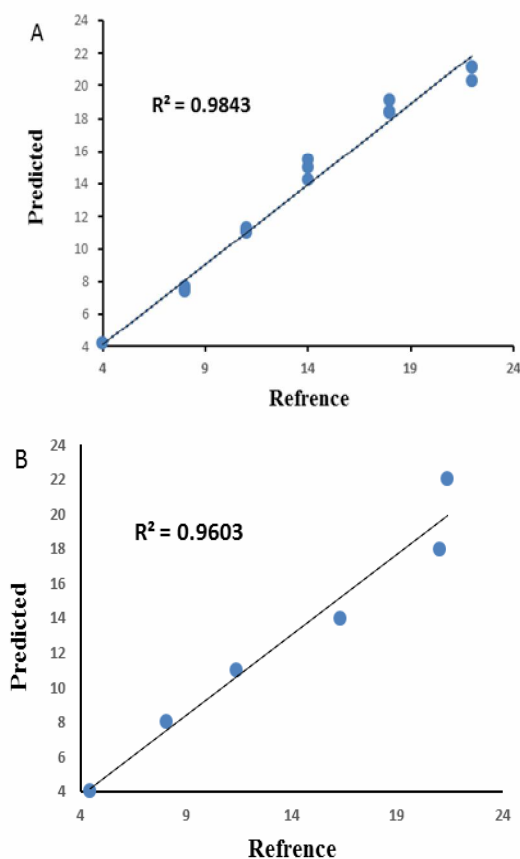


**Fig 4** Regression coefficient of PLSR model at different wavelength for prediction of moisture

یک نمونه از تغییرات میزان پروتئین برای رقم گاسکوژن در نقاط مختلف مزرعه در شکل 3 نشان داده شده است. در تحقیقی مشابهی که به صورت بر خط مقدار پروتئین نمونه‌ها اندازه‌گیری شد [18]، تغییرات قابل توجهی بین پروتئین نقاط مختلف مزرعه یافت گردید.



**Fig 3** Comparison of protein content at different points in field space



**Fig 5** Relationship between predicted and measured value of wheat seed moisture A) Calibration data B) prediction data

همان‌گونه که در جدول (2) نشان داده شده است، با اعتبارسنجی مدل‌های مربوط به رطوبت گندم، بهترین مدل با روش PLSR و پیش‌پردازش ترکیبی MA+D2+SNV حاصل شد. در این پیش‌بینی، مقدار ضرایب  $R^2$  و RMSE به ترتیب برابر با 0/96 و 0/99 بدست آمد. ضریب SDR برای مدل انتخابی، 5/34 و همچنین برای تمام مدل‌ها از 3 بالاتر بود که نشان می‌دهد مدل مربوطه از دقت عالی در پیش‌گویی رطوبت برخوردار است.

نمودار ضریب رگرسیون (شکل 4) نشان می‌دهد که بهترین طول موج برای تخمین میزان رطوبت در محدوده طول موج 1405 نانومتر که مربوط به پیوند O-H می‌باشد قرار داشته و دارای بیشترین ضریب رگرسیونی نسبت به بقیه طول موج‌ها است.

این نتایج با گزارش مربوط به جدول استاندارد مربوط به پیوندهای اتمی کاملاً هم‌خوانی دارد [4].

**Table 2** Models of NIR spectrometry to determine of the wheat moisture

RMSEP	$R_p^2$	RMSEC	$R_c^2$	Factor	Model	Preprocessing
0.99	0.960	1.02	0.984	2	PLSR	MA+D2+SNV
1.12	0.960	1.14	0.981	3		MF+D1+SNV
1.54	0.954	1.43	0.977	2		MA+D2
1.28	0.955	1.2	0.972	2	PCR	MA+D1+SNV
1.33	0.940	1.24	0.981	3		MA+D2+SNV
1.33	0.952	1.26	0.979	2		MF+D2+MSC

#### 4- نتیجه‌گیری

هرچند روش طیف‌سنجی بیشتر برای تشخیص پروتئین گندم فرآوری شده مثل آرد بکار رفته شده است، ولی نتایج تحلیل روش‌های رگرسیونی به منظور تشخیص غیرمخرب شاخص میزان پروتئین گندم نشان داد که از طیف‌سنجی فروسرخ نزدیک در محدوده 950-1650 نانومتر نیز می‌توان برای تشخیص و پیش‌بینی دانه‌های گندم استفاده نمود. اعتبارسنجی مدل‌های رگرسیون چند متغیره خطی بر پایه ترکیب روش‌های مختلف پیش‌پردازش طیف‌ها نشان داد که مدل PLSR توانایی خوبی در پیشگویی پروتئین دارد. با انتخاب پیش‌پردازش ترکیبی  $SG+SNV+D1$  مدل تخمین پروتئین دارای بالاترین میزان ضریب رگرسیون (0/84) و SDR (2/54) و کمترین مقدار RMSEP (0/835) بود. تغییرات معنی‌دار پروتئین در مکان‌های مختلف نشان داد که تهیه نقشه عملکرد کیفی مزرعه‌ای توجیح پذیر می‌باشد و با نصب دستگاه طیف‌سنج بر روی کمباین، امکان تهیه نقشه کیفی و مدیریت مناسبی برای افزایش عملکرد مزرعه وجود دارد. همچنین روش طیف‌سنجی NIR برای تشخیص میزان رطوبت دانه گندم نیز موفق عمل کرد؛ به گونه‌ای که با اعتبارسنجی مدل‌های رگرسیون چند متغیره خطی و ترکیب پیش‌پردازش‌ها، مدل‌های تدوین شده توانستند شش سطح رطوبتی را با ضریب رگرسیون بالایی (0/96) تشخیص دهند.

#### 5- منابع

- [1] Anonymous. (2018). Food and Agricultural Organization of United Nation. Available in: [www.fao.org](http://www.fao.org). (Accessed January, 2018).
- [2] Rastgar, M.A. 2015. General Agriculture. Tehran, Berahmand, p. 440.
- [3] Bahraei, S. (2003). Bread wheat quality evaluation based on the high molecular weight glutenin subunits. Iranian Journal of Crop Science. 5(3): 204-215.
- [4] Bagherpour, H., S. Minaei and M. Abdollahian Noghabi. (2014). Non-destructive determination of sugar content in root beet by near infrared spectroscopy (NIRS). Journal of Food Science and Technology, 6(1): 13-22.
- [5] Jha, S. N. and T. Matsuoka. (2000). Non-destructive techniques for quality evaluation of intact fruits and vegetables. Food Science and Technology Research. 6(4): 248-251.
- [6] Cen, H. and Y. He. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. Trends in Food Science and Technology. 18(2): 72-83.
- [7] Mao, X., Sun, L., Hui, G., and Xu, L. 2014. Modeling research on wheat protein content measurement using near-infrared reflectance spectroscopy and optimized radial basis function neural network. Journal of food and drug analysis, 22(2): 230-235.
- [8] Chen, H., Sun, Y., Wortmann, A., Gu, H., and Zenobi, R. (2007). Differentiation of maturity and quality of fruit using noninvasive extractive electrospray ionization quadrupole time-of-flight mass spectrometry. 79(4): 1447-1455.
- [9] Mutlu, A.C., Boyaci, I.H., Genis, H.E., Ozturk, R., Basaran-Akgul, N., Sanal, T. and Evlice, A.K. (2011). Prediction of wheat quality parameters using near-infrared spectroscopy and artificial neural networks. European food research and technology. 233(2): 267-274.
- [10] Hong, J.-H., K. Ikeda, I. Kreft and K. Yasumoto. (1996). Near-infrared diffuse reflectance spectroscopic analysis of the amounts of moisture, protein, starch,



- by means of NIR spectroscopy: A review". *Postharvest Biology and Technology*, 46(2): 99-118.
- [16] Kovalenko, I. V., Rippke, G. R., and Hurburgh, C. R. (2006). Determination of amino acid composition of soybeans (*Glycine max*) by near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry*. 54(10): 3485-3491.
- [17] Liu, Y., Sun, X., Zhou, J., Zhang, H., and Yang, C. (2010). Linear and nonlinear multivariate regressions for determination sugar content of intact Gannan navel orange by Vis-NIR diffuse reflectance spectroscopy. *Mathematical and Computer Modeling*. 51(11-12): 1438-1443
- [18] Long, D.S., Engel, R.E. and Siemens, M.C. (2008). Measuring grain protein concentration with in-line near infrared reflectance spectroscopy. *Agronomy Journal*, 100(2), 247-252.
- [19] Caporaso, N., Whitworth, M. B., and Fisk, I. D. 2018. Protein content prediction in single wheat kernels using hyperspectral imaging. *Food chemistry*. 240(1): 32-42.
- amylose, and tannin in buckwheat flours. *Journal of nutritional science and vitaminology*. 42(4): 359-366.
- [11] Kays, S. E., F. E., Barton and W. R., Windham. (2000). Predicting protein content by near infrared reflectance spectroscopy in diverse cereal food products. *Journal of Near Infrared Spectroscopy*. 8(1): 35-43.
- [12] Delwiche, S. (1995). Main content area Single wheat kernel analysis by near-infrared transmittance: protein content. *Cereal chemistry*. 72(1): 11-16.
- [13] Osborne, B. G. and T. Fearn. (1983). Collaborative evaluation of near infrared reflectance analysis for the determination of protein, moisture and hardness in wheat. *Journal of the Science of Food and Agriculture*. 34(9): 1011-1017.
- [14] Totschnig, G., Winter, F., Pustogov, V., Faist, J., and Müller, A. (2002). Mid-infrared external-cavity quantum-cascade laser. *Optics letters*, 27(20), 1788-1790
- [15] Nicolai, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., and Lammertyn, J. (2007). "Nondestructive measurement of fruit and vegetable quality

## Feasibility of Using Near Infrared Spectroscopy and PLSR Model for Prediction of Wheat Grains Qualities and Mapping Quality Yield Map

Foroozani, B. <sup>1</sup>, Bagherpour, H. <sup>2\*</sup>, Zaboli, Kh. <sup>3</sup>

1. MSc, graduated student, Department of Biosystems Engineering, Bu-Ali Sina University, Hamedan, Iran

2. Assistant Professor, Department of Biosystems Engineering, Bu-Ali Sina University, Hamedan, Iran

3. Assistant Professor, Department of Animal Science, Bu-Ali Sina University, Hamedan, Iran

(Received: 2019/04/17 Accepted: 2020/02/15)

Protein as an important ingredient in wheat plays main role in the production of wheat's products. Because of the production of various products from wheat, fast and online measuring of wheat grain quality is very important to control of flour production process and choosing an appropriate variety. Also in precision farming, combination of quantity and quality maps lets farmers to evaluate and control the plant production, well. Therefore, the purpose of this study was to evaluate the use of infrared spectroscopy in reflectance mode to predict protein and moisture content of wheat grain. In this study about 108 samples were collected from three varieties namely Mihan, Gazkojhen and Pishgam in the region near Hamedan province in Iran. Grain proteins content were measured with a DA7200 near infrared spectroscopy apparatus. This spectroscopy collects reflectance over a wavelength range of 650-1650 nm in 5 nm increments. Results show that the best models were obtained using the PLSR method and its preprocessing SG+SNV+D1 and MA+D2+SNV for protein and moisture content, respectively. The correlation coefficient ( $R^2$ ), root mean square error of prediction (RMSEP) and Standard Deviation Ratio (SDR) were obtained 0.84, 0.835 and 2.54 for protein content, whereas 0.96, 0.994 and 5.34 for moisture content, respectively. Results showed that there are no significant differences among proteins of three varieties. But the sampling places have a significant effect on the protein content at the significant level of 5%. These results indicated that the infrared spectroscopy method is an efficient method and has a strong potential for rapid detection of protein and moisture content of wheat grains

**Keywords:** Non-destructive methods, Wheat, Near infrared, Protein.

---

\* Corresponding Author E-Mail Address: h.bagherpour@basu.ac.ir