



UNIVERSIDAD NACIONAL DE TRUJILLO

UNT

FACULTAD DE CIENCIAS AGROPECUARIAS

ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA AGROINDUSTRIAL

**Validación de un método rápido para la determinación de humedad de
harina de arroz (*Oryza sativa* L.), mediante espectroscopía infrarroja por
transformada de Fourier (FTIR)**

Validation of a rapid method for the determination of moisture in rice flour (*Oryza
sativa* L.), using Fourier transform infrared spectroscopy (FTIR)

TESIS

**PARA OPTAR EL TÍTULO PROFESIONAL DE
INGENIERO AGROINDUSTRIAL**

AUTOR: Otiniano Burgos, Anthony Brando

ASESOR: MSc. Barraza Jáuregui, Gabriela del Carmen

TRUJILLO – PERÚ

2023

DEDICATORIA

A Dios:

*Por ser la guía en mi vida, por su cuidado,
bondad y amor para darme las fuerzas de
cumplir un objetivo más en mi vida. La
gloria siempre para él.*

A mis padres:

*Gladys y Américo, por ser la principal
motivación y apoyo en mis objetivos, por sus
sacrificios a lo largo de mi vida estudiantil.
Además de todos los valores inculcados
para ser una mejor persona cada día.*

A mi familia:

*Todos mis familiares cercanos que me
apoyaron a lo largo de mi vida
universitaria, a la memoria de mis abuelos
y tío, quienes siempre soñaron con verme
cumplir cada meta trazada.*

El autor

Sustentado y Aprobado ante el Honorable Jurado:

**Validación de un método rápido para la determinación de humedad de harina de
arroz (*Oryza sativa* L.), mediante espectroscopía infrarroja por transformada de
Fourier (FTIR)**

Presentado por:

Br. Otiniano Burgos Anthony Brando

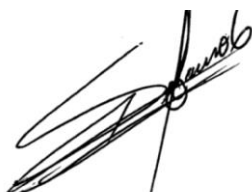
Aprobado por:



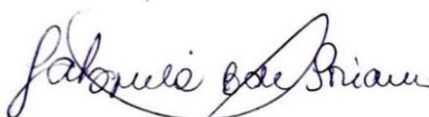
MSc. Leslie Cristina Lescano Bocanegra
Presidente



MSc. Freddy Waldir Gómez Escobedo
Secretario



Mg. Juan Carlos Solano Gaviño
Miembro Vocal



MSc. Gabriela del Carmen Barraza Jáuregui
Miembro Asesor

AGRADECIMIENTO

A la MSc. Gabriela del Carmen Barraza Jáuregui por su asesoría y apoyo en la elaboración de este trabajo, brindando acceso a las instalaciones del laboratorio, así como sus sugerencias y por su formación académica universitaria que permite desempeñarme óptimamente en esta etapa de mi vida.

A mis profesores de la escuela profesional de Ingeniería Agroindustrial por compartir sus conocimientos y experiencias formándome para mi vida profesional.

A mis compañeros de la promoción XXI, por todas las experiencias vividas, grupos que formamos académicamente y se volvieron buenos amigos, brindándome consejos y motivándome para siempre salir adelante a lo largo de los cinco años de carrera en la escuela de ingeniería agroindustrial.

INDICE

DEDICATORIA.....	ii
AGRADECIMIENTO.....	iv
INDICE	v
RESUMEN.....	vi
ABSTRACT	vii
I. INTRODUCCIÓN.....	1
II. MATERIALES Y METODOLOGÍA	4
2.1. Materiales.....	4
2.2. Metodología.....	4
2.2.1. Esquema experimental de la investigación	4
2.2.2. Preparación de muestras.....	4
2.2.3. Determinación del contenido de humedad de las muestras.....	6
2.2.4. Adquisición de curvas espectrales.....	6
2.2.5. Pre-procesamiento espectral.....	6
2.3. Análisis de datos	6
2.3.1. Minería de datos.....	6
2.3.2. Machine learning para construcción de un modelo de completo PLS-R (regresión de mínimos cuadrados parciales)	7
III. RESULTADOS Y DISCUSIÓN.....	8
3.1. Minería de datos para la identificación de espectros atípicos	8
3.2. Machine learning para construcción de un modelo de regresión completo empleando regresión de mínimos cuadrados parciales (PLS-R)	11
IV. CONCLUSIONES.....	16
V. REFERENCIAS BIBLIOGRÁFICAS	17
ANEXOS.....	

RESUMEN

La harina de arroz es higroscópica, por lo que es necesario un monitoreo constante de su contenido de humedad para garantizar su calidad. Por ello, diversos estudios indican que existe una tendencia en encontrar un método de análisis de alimentos, que sea práctico y sencillo. Por lo tanto, se buscó validar un método rápido para la determinación de humedad por espectroscopia infrarrojo medio en harina de arroz empleando PLSR. Se acondicionaron 5 muestras de harina de arroz por cada recipiente hermético con humedades relativas distintas: 0%, 33%, 55%, 76% y 86%, se realizó 13 repeticiones. Se obtuvo 325 espectros por FTIR (Thermo Scientific, Alemania), en un rango de longitud de onda de 1150 a 3750 nm. Se corrigió la línea base en los espectros y no se presentó valores negativos de absorbancia. Se separó espectros típicos de atípicos con minería de datos mediante distancia de Mahalanobis. Para la construcción del modelo PLSR se escogió 158 espectros para la calibración y 68 espectros para la validación externa. Se trabajó la primera derivada de Savitzky-Golay, debido a que obtuvo un menor valor de error cuadrático medio (0.569) y mayor coeficiente de determinación de validación cruzada (0.951). En la validación externa, el coeficiente de determinación (R^2_v), 0.953 y la relación de desviación de la predicción (RPD), 4.617 demostraron la capacidad de predicción del modelo. Por consiguiente, se validó un método rápido, no destructivo y no contaminante para la determinación de humedad por espectroscopia infrarrojo medio en harina de arroz.

Palabras clave: FTIR, harina de arroz, distancia de Mahalanobis, regresión de mínimo cuadrados parciales (PLSR), coeficiente de determinación, relación de desviación de la predicción.

ABSTRACT

Rice flour is hygroscopic, so constant monitoring of its moisture content is necessary to ensure its quality. Therefore, various studies indicate that there is a tendency to find a method of food analysis that is practical and simple. Therefore, it was sought to validate a rapid method for moisture determination by mid-infrared spectroscopy in rice flour using PLSR. 5 samples of rice flour were conditioned for each hermetic container with different relative humidities: 0%, 33%, 55%, 76% and 86%, 13 repetitions were carried out. 325 spectra were obtained by FTIR (Thermo Scientific, Germany), in a wavelength range from 1150 to 3750 nm. The baseline in the spectra was corrected and no negative absorbance values were presented. Typical spectra were separated from outliers with data mining by Mahalanobis distance. For the construction of the PLSR model, 158 spectra were chosen for calibration and 68 spectra for external validation. The first derivative of Savitzky-Golay was used, since it obtained a lower mean square error value (0.569) and a higher cross-validation coefficient of determination (0.951). In external validation, the coefficient of determination (R^2V), 0.953, and the prediction deviation ratio (RPD), 4.617 demonstrated the predictive ability of the model. Therefore, a fast, non-destructive and non-contaminating method for the determination of moisture by mid-infrared spectroscopy in rice flour was validated.

Keywords: FTIR, rice flour, Mahalanobis distance, partial least squares (PLSR) regression, coefficient of determination, prediction deviation ratio.

I. INTRODUCCIÓN

El arroz (*Oriza sativa* L.) es uno de los cereales más consumidos en nuestro país. En el último año, la región La Libertad fue uno de los cinco principales productores de arroz a nivel nacional, con una participación del 8.6% del total de la producción en el Perú (MIDAGRI, 2022).

Este cereal se industrializa y comercializa en diferentes tipos de calidad: arroz extra, arroz superior, arroz corriente. Así mismo, existen subproductos del proceso de pilado de arroz: harina de arroz, arroz partido en fracciones grandes (utilizado para sopas) y arroz partido en fracciones pequeñas (Fasabi, 2019).

Por otro lado, la harina de arroz se obtiene por molienda y tamizado de granos de arroz sanos, limpios, enteros o quebrados, con o sin cáscara, libre de impurezas y materia extraña que alteren su calidad. Además, posee un valor nutritivo semejante a otras harinas de cereales, pero no contiene gluten y es por ello que se recomienda en la dieta de personas celíacas, que son intolerantes al gluten del trigo y otros cereales. También se emplea en la preparación de: productos de panificación, salsas, sopas instantáneas, bebidas, productos cárnicos congelados, etc. (Wu et al., 2019).

No obstante, la harina tiene la capacidad de absorber o ceder humedad y ello provoca la alteración de la calidad. La humedad baja en la harina es considerada segura y fácil de conservar, pero cuando la humedad aumenta por encima del 17%, provoca reacciones químicas y bioquímicas, especialmente en el contenido de lípidos que alteran su calidad, debido a que la harina llega a ser susceptible al ataque de microorganismos que degradan el azúcar, produciendo fermentación ácida, alcohólica y acética, lo que ocasiona aumento de acidez en la harina (Anticona, 2017).

Las empresas que procesan harinas o utilizan esta materia prima durante algún proceso en la elaboración de alimentos, toman en cuenta a la humedad como un parámetro importante de calidad, para lo cual se requiere un monitoreo constante y rápido que garantice un producto final en óptimas condiciones y así evitar pérdidas por el retiro de lotes que no cumplan con las especificaciones técnicas debido a un insuficiente control empleado para la determinación de humedad (Sampaio et al., 2020; Wu et al., 2019).

En vista de ello, los investigadores están en búsqueda de métodos analíticos más rápidos, selectivos, precisos y menos contaminantes. Existen métodos como la calorimetría diferencial de barrido, cromatografía de gases, cromatografía de líquidos de alta resolución, métodos basados en DNA y diversas técnicas de espectroscopia: UV-Vis (ultra violeta visible), colorimetría, difracción de rayos X o la espectroscopia de infrarrojo. Esta última, la técnica de adquisición de datos por transformada de Fourier en el espectro infrarrojo (FTIR) ha revolucionado los métodos de análisis y el manejo de las muestras, llegando a ser una técnica alternativa para el análisis de alimentos, de manera que resulta una herramienta muy útil y práctica para obtener resultados en menos tiempo (Feng et al., 2019).

Diversas investigaciones reportaron el uso de la espectroscopia infrarroja en la medición y caracterización de diferentes componentes en alimentos (Quimis y Salazar, 2017; Genkawa et al., 2016). Se evaluó la estructura de la proteína del gluten mediante FTIR durante el proceso de envejecimiento de los granos de trigo en función del contenido de humedad, el tiempo de almacenamiento y la temperatura (Aghababei et al., 2017). De igual manera, se realizó estudios para predecir el contenido de humedad de harina de quinua (González-Muñoz et al., 2016). Se empleó para la caracterización fisicoquímica del bagazo de caña de azúcar proveniente de la Sierra Falconiana, aceite de coco, miel de abeja y pasta de tomate (Ciursă et al., 2021; Jamwal et al., 2020; Zara et al., 2017). También se demostró

la validación y efectividad del modelo predictivo PLSR con un 97.5% de confiabilidad en la determinación de humedad en remolacha (Nesakumar et al., 2018).

Por lo tanto, en el presente trabajo busca validar un método rápido para la determinación de humedad, empleando espectroscopía Infrarroja por Transformada de Fourier (FTIR) en muestra de harina de arroz.

BIBLIOTECA DE CIENCIAS AGROPECUARIAS

II. MATERIALES Y METODOLOGÍA

2.1. Materiales

Se utilizaron tres kilos de harina de arroz sin cáscara; producida y envasada por comercializadora de productos naturales NUTRIMIX E.I.R.L. Posteriormente, las muestras fueron llevadas al Laboratorio de Análisis por Instrumentación del Departamento de Ciencias Agroindustriales de la Facultad de Ciencias Agropecuarias de la Universidad Nacional de Trujillo.

La harina de arroz presentó la siguiente composición proximal: humedad, $10.27 \pm 0.5\%$; proteínas, $11.36 \pm 1.5\%$; carbohidratos, 60.16% ; grasa, $1.48 \pm 0.2\%$; ceniza, $0.96 \pm 0.5\%$; energía, 299.40 kcal y fibra total, 1.76 %.

2.2. Metodología

2.2.1. Esquema experimental de la investigación

La Figura 1 muestra el esquema experimental de la validación del método rápido para determinar humedad por espectrofotometría infrarrojo por FTIR en muestras de harinas de arroz.

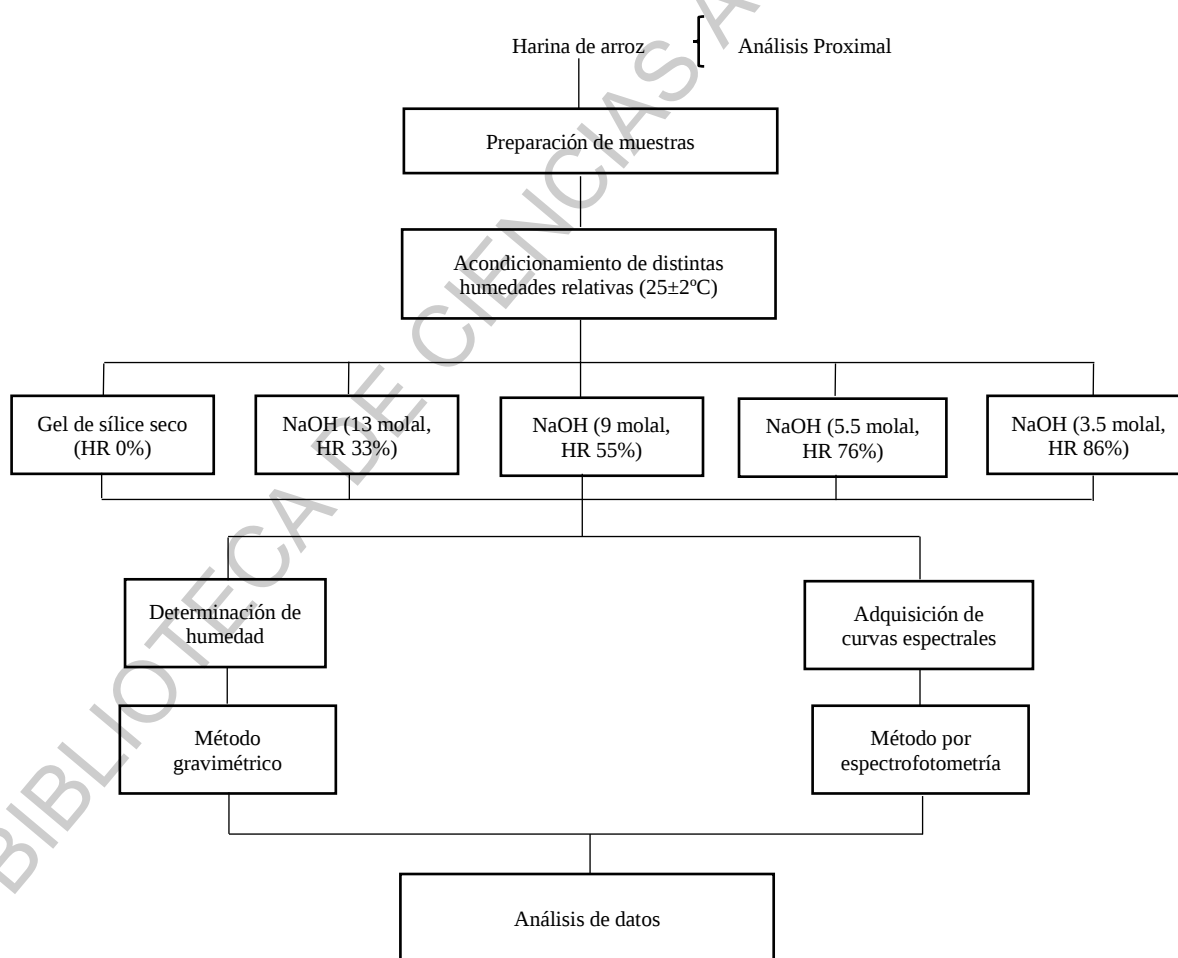
2.2.2. Preparación de muestras

La harina de arroz fue tamizada empleando tamiz N°60 (serie ASTM) de abertura de malla 250 μm para uniformizar el tamaño de partícula. La humedad inicial de la harina de arroz fue de $10.27\% \pm 0.005$. Posteriormente, se colocaron las muestras de harina de arroz en una estufa de aire forzado (Mettler, UF260 Plus, Alemania) a $50\text{ }^{\circ}\text{C}$ por 48 h, obteniendo una humedad en base seca de $5.79\% \pm 0.002$. Para lograr distintos contenidos de humedad relativa, se colocaron cinco vasos de precipitado de 50 ml conteniendo $4\text{g} \pm 0.001$ de muestra de harina de arroz cada uno y un vaso de precipitado de 100 ml conteniendo soluciones salinas no saturadas: NaOH (13 molal, HR 33%); NaOH (9 molal,

HR 55%), NaOH (5.5 molal, HR 76%) y NaOH (3.5 molal, HR 86%), además de sílica gel seca (HR 0%), en recipientes herméticos a 25 ± 2 °C (Barbosa-Cánovas et al., 2007). Se colocó cubetas transparentes de 1x1x4 cm conteniendo timol en los recipientes con humedades relativas de 55%, 76% y 86% para evitar el crecimiento de hongos. Se logró el equilibrio en las muestras de harina de arroz después de 15 días cuando no existió variación significativa en el peso (<0.1 gramos), medido con una balanza analítica (AS 82/220.X2, Radwag, Polonia) (Greenspan, 1977). Se realizaron 13 repeticiones.

Figura 1

Esquema experimental de la investigación para validar un método rápido por Espectroscopía Infrarroja con Transformada de Fourier



2.2.3. Determinación del contenido de humedad de las muestras

El contenido de humedad de las muestras de harina se calculó en base seca de acuerdo con el método gravimétrico en una estufa de aire forzado a 105 °C por 30 horas (AOAC, 2005, como se citó en García et al., 2012).

2.2.4. Adquisición de curvas espectrales

Para la adquisición de las curvas espectrales de las muestras de harina de arroz se utilizó un espectrofotómetro en el rango de infrarrojo medio con transformada de Fourier (Thermo Scientific, Nicolet iS50, Alemania) en un rango de frecuencia de 4000 – 600 cm^{-1} . El análisis FTIR se realizó colocando las muestras de harina de arroz sobre la superficie del cristal ATR. Se usó una resolución de 4 cm^{-1} con 16 escaneos por espectro en el modo de reflectancia. Finalmente, los espectros FTIR fueron transformados en modo absorbancia usando el software OMNIC Spectra 9.7.39 (Ciursă et al., 2021).

2.2.5. Pre-procesamiento espectral

Se aplicó ingeniería de datos a los espectros de las muestras de harina de arroz, siendo la base de datos conformada por 325 valores de absorbancia correspondientes a cada longitud de onda. Posteriormente se estandarizó el rango de longitud de onda desde 1150 a 3750 nm (con número de 5393 intervalos); luego se realizó la corrección de la línea base para que los espectros no presentaran valores negativos de absorbancia, sumando 0.15 u.a. a cada espectro. Se utilizó el software Python 3.7.1 (Ciursă et al., 2021).

2.3. Análisis de datos

2.3.1. Minería de datos

Se evaluó la presencia de espectros atípicos en la base de datos mediante la distancia de Mahalanobis con la técnica de componentes principales; considerándose espectros atípicos aquellos que presentaron valores de raíz cúbica de Mahalanobis mayor o igual a 3;

99 espectros fueron considerados como outliers o atípicos, siendo depurados de la base de datos inicial (Aykas et al., 2020; Gallego et al., 2013; Leys et al., 2018; Xiang et al., 2008).

2.3.2. Machine learning para construcción de un modelo de completo PLS-R (regresión de mínimos cuadrados parciales)

Para la construcción y validación del modelo se realizó la partición de la base de datos en forma aleatoria. El 70% de los espectros se emplearon para la calibración (158 espectros) y el 30% (68 espectros) para la validación externa (Benoudjit et al., 2004). Para corregir y minimizar la variabilidad de los espectros se utilizó la primera y segunda derivada de Savitzky-Golay de orden polinomial 2, ancho de ventana o filtro de 19 unidades. Para determinar un modelo que tenga la capacidad de generalización para estimar el contenido de humedad de la harina de arroz, se utilizó PLS-R, el cual permitió encontrar una relación matemática entre dos conjuntos de datos, espectros (inputs), porcentajes de humedad (outputs). Para la calibración del modelo se empleó el algoritmo de validación cruzada dejando uno (CV-LOO) (número de muestreo = 10%); para determinar el número óptimo de variables latentes (VL) se usó como indicador el menor valor de error cuadrático medio de validación cruzada (ECM_{CV}) y para el rendimiento del modelo de calibración fue evaluado por el coeficiente de determinación de validación cruzada (R^2_{CV}).

Se determinó la capacidad de generalización del modelo, para ello se evaluó el error cuadrático medio de validación externa (ECM_V) y coeficiente de determinación de validación externa (R^2_V); además de la relación de desviación de predicción (RPD) cuyo valor debe ser mayor a 3 (Alfaro, 2016; Aykas et al., 2020; Killner et al., 2011; Pasquini y Bueno, 2007;). Adicionalmente se construyó un modelo reducido a partir de los coeficientes de regresión de mayor importancia (en valor absoluto) del modelo completo inicial, pasando de 5393 intervalos de longitudes de onda a 1140 intervalos. Para la analítica de datos se utilizó el software Python 3.7.1 con la plataforma de Google Colab. (Ciursă et al., 2021).

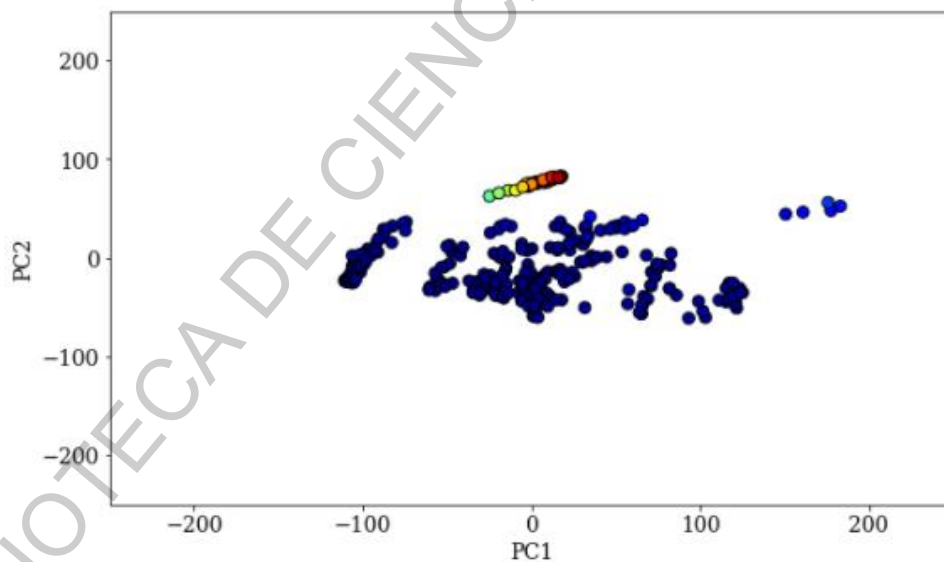
III. RESULTADOS Y DISCUSIÓN

3.1. Minería de datos para la identificación de espectros atípicos

La Figura 2 muestra la identificación de espectros atípicos empleando la distancia de Mahalanobis. La raíz cubica de la distancia de Mahalanobis cuyo valor fue menor a 3 son considerados espectros inliers y seleccionados para la construcción del modelo PLS-R. Por otro lado, la raíz cubica de la distancia de Mahalanobis cuyo valor fue mayor o igual a 3 son considerados outliers y descartados para el modelo. Se obtuvieron 325 espectros, el 69.54% se consideraron espectros inliers (tonalidad azul) y 30.46%, espectros outliers (tonalidad verde a roja), como se muestra en la Figura 2.

Figura 2

Espectros atípicos detectados empleando la distancia de Mahalanobis



La distancia de Mahalanobis (DM) es la separación métrica estadística entre grupo de datos que clasifica y selecciona espectros, además es un factor fundamental que ayuda a determinar el valor del coeficiente de determinación ($R^2 > 0.90$) para la construcción del modelo predicción PLSR (Aykas et al., 2020; Flores-Valdez et al., 2020). El valor que se le

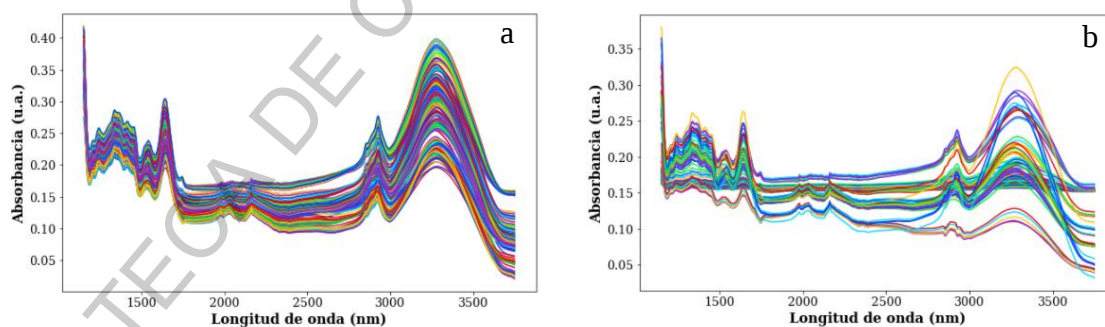
asigne a la distancia de Mahalanobis varía según el criterio de cada investigación (Gallego et al., 2013; Ghorbani, 2019).

Las investigaciones indican que obtuvieron curvas espectrales inliers y outliers aplicando la distancia de Mahalanobis, en la determinación de residuos de tetraciclina en la leche de vaca cruda; y también en la identificación y cuantificación de adulterantes en café (*Coffea arabica* L.) (Casarrubias-Torres et al., 2018; Flores-Valdez et al., 2020).

Las curvas espectrales de las muestras de harina de arroz que presentan el mismo comportamiento, es decir están superpuestas, en el rango de longitud de onda de 1150 a 3750 nm, se le denomina inliers, como se observa en la Figura 3a; estas curvas son consideradas para la construcción del modelo de predicción. Sin embargo, las curvas espectrales que presentan un comportamiento diferente, como se muestra en la Figura 3b, son considerados outliers; estos espectros son descartados para la construcción del modelo.

Figura 3

(a) Espectros inliers (b) Espectro outliers



La Figura 4a muestra el Boxplot para los porcentajes de humedad gravimétrica asociados a los espectros inliers, donde se observa que la línea está muy cerca al centro del diagrama, indicando que existe una distribución normal de los datos (Pereira et al., 2009).

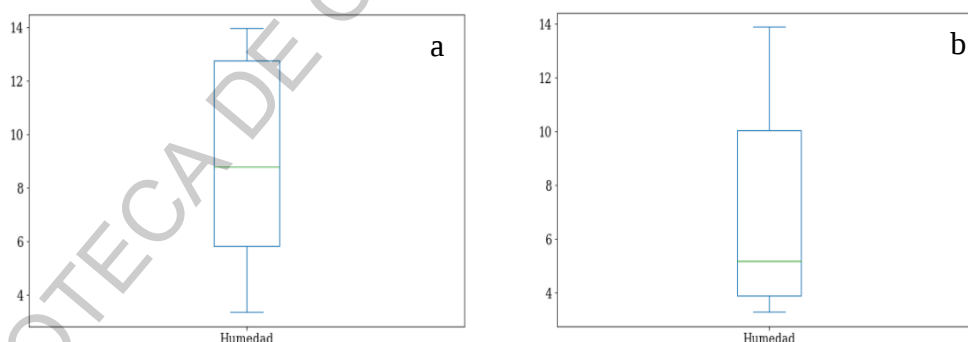
Los valores se encontraron entre 3.35% y 13.94%, con media de 9.60% y desviación estándar

de 3.44%; además el valor de rango intercuartílico (P_{25} : 5.81%; P_{75} : 12.75%), que contiene al 50% de los datos, fueron cercanos a la mediana (8.76%). Además, se menciona que si se encuentra una distribución normal de los datos es porque se aplicó adecuadamente la distancia de Mahalanobis en la evaluación de los espectros (Galeano et al., 2015; Ghorbani, 2019).

Por el contrario, en la Figura 4b se observa una asimetría para los datos de porcentajes de humedad gravimétrica asociados a los espectros considerados outliers; porque la línea del diagrama está ubicada en la parte inferior, señalando que existe una distribución no normal de los datos (Pereira et al., 2009). Los valores de humedad se encontraron entre 3.30 y 13.86%, con media de 6.81% y desviación estándar de 3.74%; además de rango intercuartílico (del percentil 25 al 75) que contiene al 50% de los datos de humedad cercanos a la mediana (5.18%) entre 3.90 y 10.02%.

Figura 4

(a) Diagrama de Boxplot para datos de inliers (b) Diagrama de Boxplot para datos de outliers



En la predicción de ácidos grasos de la composición de la leche y en la investigación del monitoreo de procesos de digestión anaeróbica, ambos empleando el modelo PLSR mediante FTIR, se reportó una distribución normal de los datos en el Boxplot (Rovere et al., 2021; Zeaiter et al., 2022).

Así mismo, el boxplot indicó una distribución normal en la investigación del contenido de nitrato en vegetales de hoja verde como col china (*Brassica rapa subsp. chinensis*), espinaca (*Ipomoea aquatica Forssk*), apio (*Apium graveolens L.*) y lechuga (*Lactuca sativa L.*) (Ma et al., 2021). De igual modo, en el estudio sobre la liberación de opsina en las neuronas corticales y aceleración del aprendizaje (Zhang et al., 2021).

3.2. Machine learning para construcción de un modelo de regresión completo empleando regresión de mínimos cuadrados parciales (PLS-R)

Para la calibración del modelo completo se contó con 158 espectros, con rangos de humedad observada entre 3.35 y 13.94%; en la Tabla 1 se presenta el modelo generado con el preprocesamiento de la primera derivada de Savitzky-Golay, con número óptimo de variables latentes de 7, error cuadrático medio de validación cruzada (ECM_{VC}) de 0.569 y coeficiente de determinación de validación cruzada (R^2_{CV}) de 0.951; así mismo se presenta el modelo de la segunda derivada de Savitzky-Golay, con número óptimo de variables latentes de 5, error cuadrático medio de validación cruzada (ECM_{VC}) de 1.003 y coeficiente de determinación de validación cruzada (R^2_{CV}) de 0.914; por lo que se determinó trabajar con el modelo completo de la primera derivada de Savitzky-Golay por presentar menor ECM_{VC} y mayor R^2_{CV} .

La capacidad de generalización del modelo completo fue evaluada empleando los datos de validación externa conformada por 68 espectros con rango de humedad entre 3.79 y 13.69%, donde el error cuadrático medio de validación externa (ECM_V) fue de 0.569 y coeficiente de determinación de validación externa (R^2_V) fue de 0.953; además presentó la relación de desviación de la predicción (RPD) de 4.617, siendo este valor superior a 3, para establecer exactitud de la capacidad de predicción del modelo (Aykas et al., 2020; Jamwal et al., 2020; Sen et al., 2021).

Tabla 1*Indicadores estadísticos del modelo completo de humedad para muestras de harina de arroz*

Pre-procesamiento	Calibración					Validación externa				
	Rango de humedad (%)	Nº	Variable latente	ECM _{CV}	R ² _{CV}	Rango de humedad (%)	Nº	ECM _V	R ² _V	RPD
1 ^{ra} derivada de Savitzky-Golay	3.35 - 13.94	158	7	0.569	0.951	3.79 - 13.69	68	0.569	0.953	4.617
2 ^{da} derivada de Savitzky-Golay	3.35 - 13.94	158	5	1.003	0.914	--	--	--	--	--

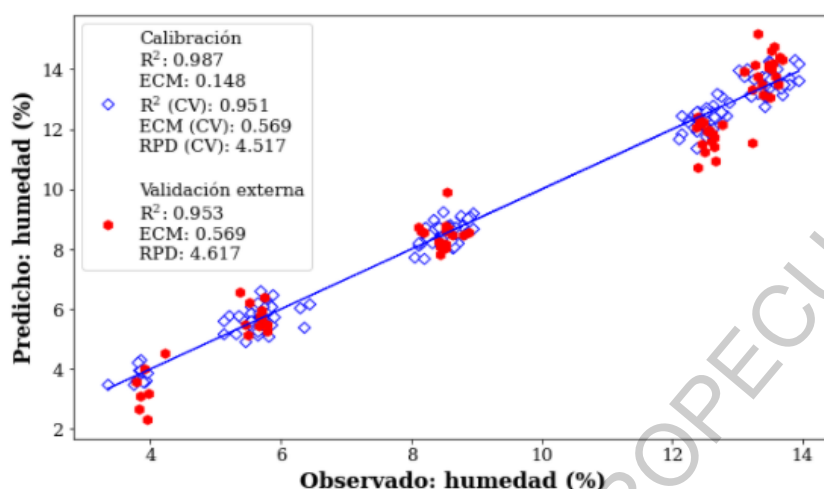
ECM_{CV} = Error cuadrático medio de validación cruzada; R²_{CV} = Coeficiente de determinación de validación cruzada; ECM_V = Error cuadrático medio de validación externa; R²_V = Coeficiente de determinación de validación externa; RPD = Relación de la desviación de predicción

El modelo de calibración generado se validó eligiendo un conjunto de muestras (30% del total de los espectros) que no se incluyeron en la calibración original. Este paso llamado validación externa fue de gran importancia para evaluar la solidez del modelo de calibración. Después de la evaluación de la calibración y validación externa, se observó que ECM_{CV} y ECM_V fueron iguales (0.569), esto indica que el modelo ha sido ajustado y optimizado al máximo, logrando su eficiencia y buen desempeño predictivo (Aykas et al., 2020; Jamwal et al., 2020).

En la Figura 5 se observa relación entre los resultados de la humedad observada y predicha. Los diamantes azules representan los valores de humedad de calibración y los hexágonos rojos representan el conjunto de validación externa; además la humedad observada y predicha se encuentran cercanas a la línea azul que representa el modelo.

Figura 5

Valores de humedad observados y predichos de las muestras de harina de arroz



Diversas investigaciones señalan que, en la validación externa el coeficiente de determinación (R^2) debe ser superior a 0.90 y el valor de la relación de desviación de la predicción (RPD) mayor a 3, ambos valores indican la validación y exactitud de la predicción del modelo utilizando PLSR (Aykas et al., 2020; Jamwal et al., 2020; Sen et al., 2021). La Figura 5 muestra el valor R^2 , 0.953 y RPD, 4.617, por lo tanto, se demuestra la validación del modelo PLSR para determinar humedad en muestras de harina de arroz.

En el análisis proximal de la pasta de tomate se obtuvo el valor R^2 , 0.99 para sólidos solubles y 0.97 para la glucosa; y valor RPD, 7.3 para sólidos solubles y 3.5 para glucosa. Además, sus valores se encuentran muy cercanos a la línea que representa el modelo como sucedió en la presente investigación (Aykas et al., 2020).

Jamwal et al. (2020), reportaron un valor R^2 , 0.99 y RPD, 11.28, demostrando la validación del modelo cuando realizó el estudio de la adulteración del aceite de coco puro por aceite frito de otros productos.

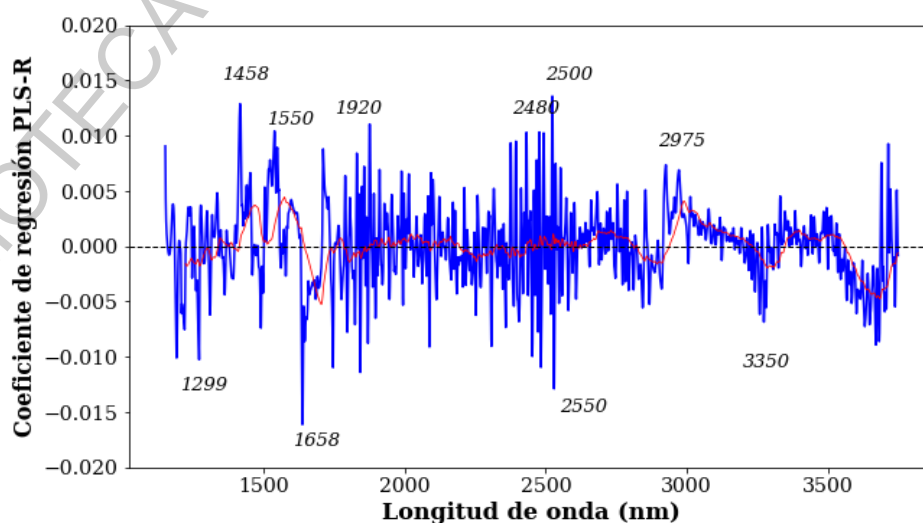
Ciursă et al. (2021), también reportaron un valor de R^2 , 0.96, demostrando la validación del modelo en el análisis de humedad, glucosa, fructuosa y sacarosa en muestras de miel de abeja.

Sen et al. (2021), demostraron la validación de su modelo debido a que el R^2 fue igual a 0.99, para los parámetros grasa, proteína, lactosa y sólidos no grasos; los valores de RPD fueron entre 8.9 y 11.6; esto en el estudio de la adulteración, detección y evaluación de la calidad en leche de búfala y cabra.

La Figura 6 muestra los coeficientes de regresión que permiten identificar los picos correspondientes al agua entre las longitudes de onda 1150 a 3750 nm. Se observa que el pico de 1658 nm, debido a su cercanía al pico de 1645 nm, hace referencia a la vibración del enlace H-O-H. Además del pico de 3350 nm, que se encuentra influenciado por un enlace diferente de O-H al de la molécula de agua; el O--H formado por la interacción entre sí de las moléculas de agua creando enlaces puentes de hidrógeno (Mondragón, 2020).

Figura 6

Modelo con coeficientes de regresión para estimar la humedad en muestras de harina de arroz



En función al modelo se puede construir un equipo con sensores que pueda leer los espectros a través de un software e identifique los picos correspondientes a esta longitud de onda.

En el Anexo A al K se encuentran los gráficos que completan la construcción del modelo (reducido) que no fue considerado en la investigación, porque presentó un menor valor de coeficiente R^2 .

BIBLIOTECA DE CIENCIAS AGROPECUARIAS

IV. CONCLUSIONES

La presente investigación demostró la validación de un método rápido para determinar humedad empleando PLSR mediante espectroscopía infrarroja media en muestras de harina de arroz. Se caracterizó picos que corresponden a grupos funcionales que representan el contenido de humedad en las curvas espectrales obtenidas. En la construcción del modelo PLSR, se relacionó los porcentajes de humedad gravimétrica y curvas espectrales, se determinó trabajar con la primera derivada porque presentó el menor valor de error cuadrático medio (ECM_{CV} , 0.569) y mayor valor R^2_{CV} , 0.951, con relación a la segunda derivada. Los valores de R^2_V , 0.953 y RPD, 4.617 determinaron que el modelo PLSR presenta una excelente capacidad de generalización en la humedad de harina de arroz. Por consiguiente, este método rápido puede ser utilizado para obtener resultados en menor tiempo en el análisis de esta muestra.

Para investigaciones que se desarrollen más adelante, se recomienda utilizar diversas muestras de harina de arroz u otra materia prima de diferentes regiones del país, logrando ampliar la base de datos de muestras de harina y otros alimentos; desarrollando una producción de modelo, esto contribuiría a obtener la información más completa y rápida de humedad de una muestra a través de un aplicativo de libre acceso.

V. REFERENCIAS BIBLIOGRÁFICAS

- Alfaro, J. (2016). El método NIR combinado con el análisis quimiométrico PLS-da para determinar la adulteración del aceite de oliva con aceite de girasol. *Pensamiento Actual*, 16(26), 163-172. <https://doi.org/10.15517/pa.v16i26.25764>
- Aykas, D. P., Rodrigues Borba, K., & Rodriguez-Saona, L. E. (2020). Non-Destructive Quality Assessment of Tomato Paste by Using Portable Mid-Infrared Spectroscopy and Multivariate Analysis. *Foods*, 9(9), 1300. <https://doi.org/10.3390/foods9091300>
- Aghababaei, A., Maftoonazad, N., Elhamirad, A., & Badii, F. (2017). Accelerated ageing of wheat grains: Part II-influence on thermal characteristics of wheat starch and FTIR spectroscopy of gluten. *Journal of cereal science*, 77, 157-165. <https://doi.org/10.1016/j.jcs.2017.08.010>
- Anticona, A. (2017). *Comparación físico-química y reológica de harinas: trigo (Triticum aestivum), centeno (Secale cereale) y triticale (x Triticosecale) en elaboración de pan*. Tesis de grado, Universidad Nacional Agraria La Molina, Facultad de Industrias Alimentarias, Lima, Perú.
- Repositorio institucional de la Universidad Nacional Agraria La Molina
<https://repositorio.lamolina.edu.pe/handle/20.500.12996/2967>
- Barbosa-Cánovas, G., Fontana, A., Schmidt, S., & Labuza, T. (2007). *Water Activity in Food*. ITF Press Editorial Advisory Board.
- Benoudjit, N., Cools, E., Meurens, M., & Verleysen, M. (2004). Chemometric calibration of infrared spectrometers: selection and validation of variables by non-linear models. *Chemometrics and intelligent laboratory systems*, 70(1), 47-53. <https://doi.org/10.1016/j.chemolab.2003.10.008>

- Casarrubias-Torres, L. M., Meza-Márquez, O. G., Osorio-Revilla, G., & Gallardo-Velázquez, T. (2018). Mid-infrared spectroscopy and multivariate analysis for determination of tetracycline residues in cow's milk. *Acta Veterinaria Brno*, 87(2), 181-188. <https://doi.org/10.2754/avb201887020181>
- Ciursă, P., Pauliuc, D., Dranca, F., Ropciuc, S., & Oroian, M. (2021). Detection of honey adulterated with agave, corn, inverted sugar, maple and rice syrups using FTIR analysis. *Food Control*, 130, 108266. <https://doi.org/10.1016/j.foodcont.2021.108266>
- Fasabi, C. (2019). *Agroindustrialización del arroz (Oryza Sativa L.) en la Empresa Agroindustrias San Hilarión S.A.C.* Tesis para optar el título profesional de Ingeniero Agroindustrial de la Universidad Nacional de San Martín. Tarapoto, Perú.
Repositorio institucional de la Universidad Nacional de San Martín
<https://repositorio.unsm.edu.pe/bitstream/11458/3739/1/FIAI%20%20Corina%20Doylith%20Fasabi%20Mozombite.pdf>
- Feng, L., Zhu, S., Chen, S., Bao, Y., & He, Y. (2019). Combining Fourier transform mid-infrared spectroscopy with chemometric methods to detect adulterations in milk powder. *Sensors*, 19(13), 2934. <https://doi.org/10.3390/s19132934>
- Flores-Valdez, M., Meza-Márquez, O. G., Osorio-Revilla, G., & Gallardo-Velázquez, T. (2020). Identification and quantification of adulterants in coffee (*Coffea arabica* L.) using FT-MIR spectroscopy coupled with chemometrics. *Foods*, 9(7), 851. <https://doi.org/10.3390/foods9070851>

Galeano, P., Joseph, E., & Lillo, R. E. (2015). The Mahalanobis distance for functional data with applications to classification. *Technometrics*, 57(2), 281-291.
<https://doi.org/10.1080/00401706.2014.902774>

Gallego, G., Cuevas, C., Mohedano, R., & Garcia, N. (2013). On the Mahalanobis distance classification criterion for multidimensional normal distributions. *IEEE Transactions on Signal Processing*, 61(17), 4387-4396.
<https://doi.org/10.1109/TSP.2013.2269047>

García, C., Dussán, S., & Gutiérrez, N. (2012). Uso de horno microondas en la determinación de contenido de humedad: yuca, ñame y plátano. *Bioteología en el Sector Agropecuario y Agroindustrial*, 10(1), 60-66.
<http://www.scielo.org.co/pdf/bsaa/v10n1/v10n1a08.pdf>

Genkawa, T., Ahamed, T., Noguchi, R., Takigawa, T., & Ozaki, Y. (2016). Simple and rapid determination of free fatty acids in brown rice by FTIR spectroscopy in conjunction with a second-derivative treatment. *Food Chemistry*, 191, 7-11.
<https://doi.org/10.1016/j.foodchem.2015.02.014>

Ghorbani, H. (2019). Mahalanobis distance and its application for detecting multivariate outliers. *Facta Univ Ser Math Inform*, 34(3), 583-95.
<https://doi.org/10.22190/FUMI1903583G>

González-Muñoz, A., Montero, B., Enrione, J., & Matiacevich, S. (2016). Rapid prediction of moisture content of quinoa (*Chenopodium quinoa* Willd.) flour by Fourier transform infrared (FTIR) spectroscopy. *Journal of Cereal Science*, 71, 246-249.
<https://doi.org/10.1016/j.jcs.2016.09.006>

Greenspan, L. (1977). Humidity fixed points of binary saturated aqueous solutions. *Journal of research of the National Bureau of Standards. Section A, Physics and chemistry*, 81(1), 89. <https://doi.org/10.6028/jres.081A.011>

Jamwal, R., Kumari, S., Kelly, S., Cannavan, A., & Singh, D. K. (2020). Rapid detection of pure coconut oil adulteration with fried coconut oil using ATR-FTIR spectroscopy coupled with multivariate regression modelling. *LWT*, 125, 109250. <https://doi.org/10.1016/j.lwt.2020.109250>

Killner, M. H., Rohwedder, J. J., & Pasquini, C. (2011). A PLS regression model using NIR spectroscopy for on-line monitoring of the biodiesel production reaction. *Fuel*, 90(11), 3268-3273. <https://doi.org/10.1016/j.fuel.2011.06.025>

Leys, C., Klein, O., Dominicy, Y., & Ley, C. (2018). Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. *Journal of experimental social psychology*, 74, 150-156. <https://doi.org/10.1016/j.jesp.2017.09.011>

Ma, F., Du, C., Zheng, S., & Du, Y. (2021). In Situ Monitoring of Nitrate Content in Leafy Vegetables Using Attenuated Total Reflectance– Fourier-Transform Mid-infrared Spectroscopy Coupled with Machine Learning Algorithm. *Food Analytical Methods*, 14(11), 2237-2248. <https://doi.org/10.1007/s12161-021-02048-7>

MIDAGRI. 2022. Ministerio de Desarrollo Agrario y Riego.

<https://siea.midagri.gob.pe/portal/publicacion/boletines-anuales/4-agricola>

Mondragón, P. (2020). Principios y aplicaciones de la espectrofotometría infrarrojo en el análisis de alimentos y bebidas.

<https://ciatej.repositorioinstitucional.mx/jspui/bitstream/1023/720/1/Infrarrojo%20Cap%201.pdf>

Nesakumar, N., Baskar, C., & Kesavan, S. (2018). Analysis of Moisture Content in Beetroot using Fourier Transform Infrared Spectroscopy and by Principal Component Analysis. *Scientific Reports*, 1-10. <https://doi.org/10.1038/s41598-018-26243-5>

Pasquini, C., & Bueno, A. F. (2007). Characterization of petroleum using near-infrared spectroscopy: Quantitative modeling for the true boiling point curve and specific gravity. *Fuel*, 86(12-13), 1927-1934. <https://doi.org/10.1016/j.fuel.2006.12.026>

Pereira, A. C., Reis, M. S., & Saraiva, P. M. (2009). Quality control of food products using image analysis and multivariate statistical tools. *Industrial & Engineering Chemistry Research*, 48(2), 988-998. <https://doi.org/10.1021/ie071610b>

Quimis, K. & Salazar, M. (2017). *Propuesta de nuevas aplicaciones culinarias del polvo de arveja* . Tesis para obtener la Licenciatura en Gastronomía de la Facultad de Ingeniería Química de la Universidad de Guayaquil. Guayaquil, Ecuador. <http://repositorio.ug.edu.ec/handle/redug/22434>

Rovere, G., de Los Campos, G., Lock, A. L., Worden, L., Vazquez, A. I., Lee, K., & Tempelman, R. J. (2021). Prediction of fatty acid composition using milk spectral data and its associations with various mid-infrared spectral regions in Michigan Holsteins. *Journal of Dairy Science*, 104(10), 11242-11258. <https://doi.org/10.3168/jds.2021-20267>

Sampaio, P. S., Castanho, A., Almeida, A. S., Oliveira, J., & Brites, C. (2020). Identification of rice flour types with near-infrared spectroscopy associated with PLS-DA and

SVM methods. *European Food Research and Technology*, 246(3), 527-537.

<https://doi.org/10.1007/s00217-019-03419-5>

Sen, S., Dundar, Z., Uncu, O., & Ozen, B. (2021). Potential of Fourier-transform infrared spectroscopy in adulteration detection and quality assessment in buffalo and goat milks. *Microchemical Journal*, 166, 106207.

<https://doi.org/10.1016/j.microc.2021.106207>

Wu, T., Wang, L., Li, Y., Qian, H., Liu, L., Tong, L., ... & Zhou, S. (2019). Effect of milling methods on the properties of rice flour and gluten-free rice bread. *Lwt*, 108, 137-144.

<https://doi.org/10.1016/j.lwt.2019.03.050>

Xiang, S., Nie, F., & Zhang, C. (2008). Learning a Mahalanobis distance metric for data clustering and classification. *Pattern recognition*, 41(12), 3600-3612.

<https://doi.org/10.1016/j.patcog.2008.05.018>

Zara, J., Yegres, F., Vargas, N., Cubillan, L., Navas, P., & Márquez-Riquel, M. (2017). Empleo de la Espectroscopia Infrarroja (FT-IR-ATR) como herramienta para la Caracterización del bagazo de caña proveniente de la Sierra Falconiana. *Química Viva*, 16(3), 17-24. <https://www.redalyc.org/pdf/863/86354619003.pdf>

Zeaiter, M., Latrille, É., Gras, P., Steyer, J. P., Bellon-Maurel, V., & Roger, J. M. (2022). Improvements in the Robustness of Mid-Infrared Spectroscopy Models against Chemical Interferences: Application to Monitoring of Anaerobic Digestion Processes. *AppliedChem*, 2(2), 117-127.

<https://doi.org/10.3390/appliedchem2020008>

Zhang, J., He, Y., Liang, S., Liao, X., Li, T., Qiao, Z., ... & Chen, X. (2021). Non-invasive, opsin-free mid-infrared modulation activates cortical neurons and accelerates

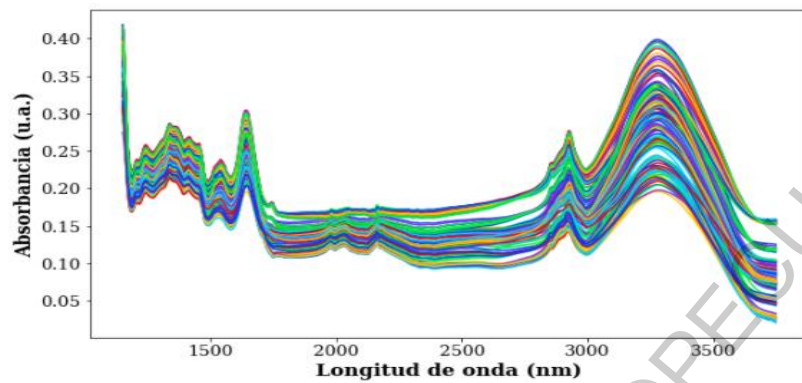
associative learning. *Nature communications*, 12(1), 1-9.
<https://doi.org/10.1038/s41467-021-23025-y>

BIBLIOTECA DE CIENCIAS AGROPECUARIAS

ANEXOS

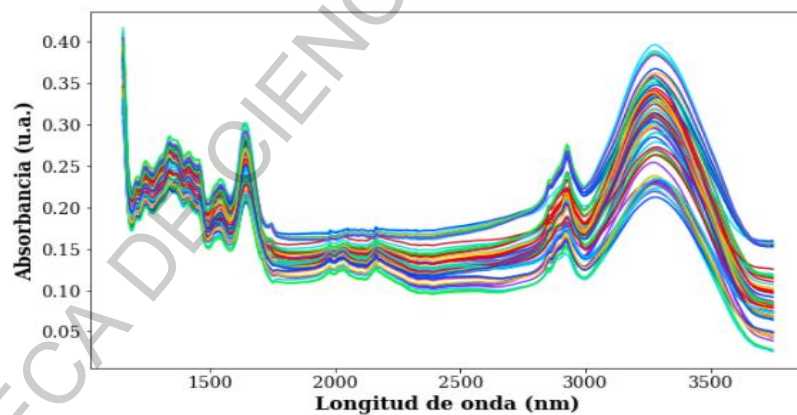
Anexo A

Espectros de calibración para muestras de arroz



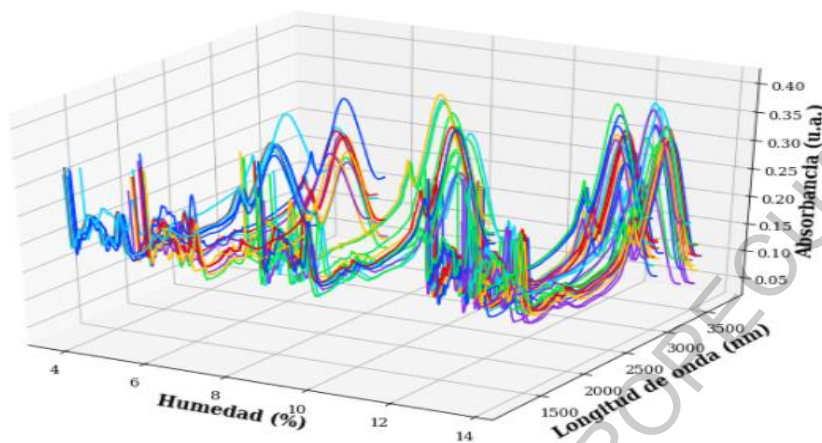
Anexo B

Espectros de validación externa para muestras de arroz



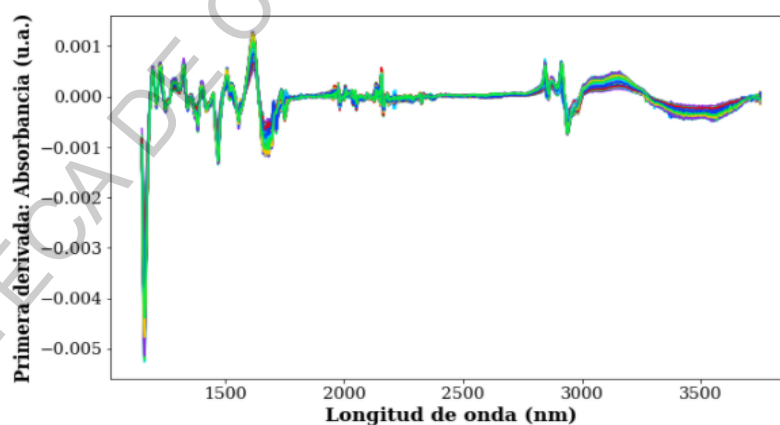
Anexo C

Humedad por el método espectral y gravimétrico para muestras de arroz



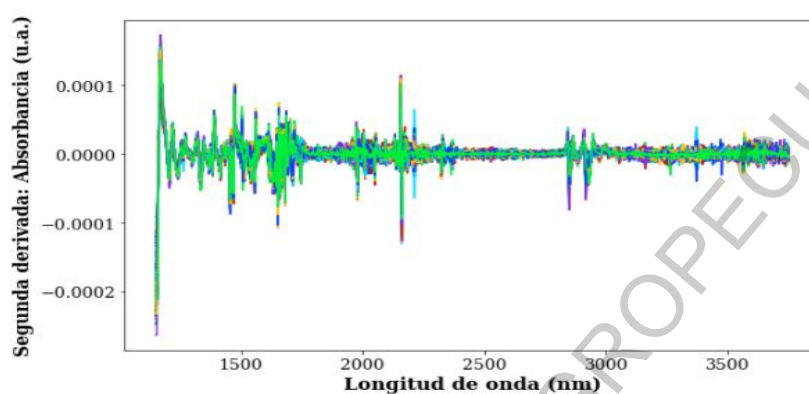
Anexo D

Primera derivada de Savitzky-Golay aplicada a la humedad espectral del Modelo completo en muestras de harina de arroz



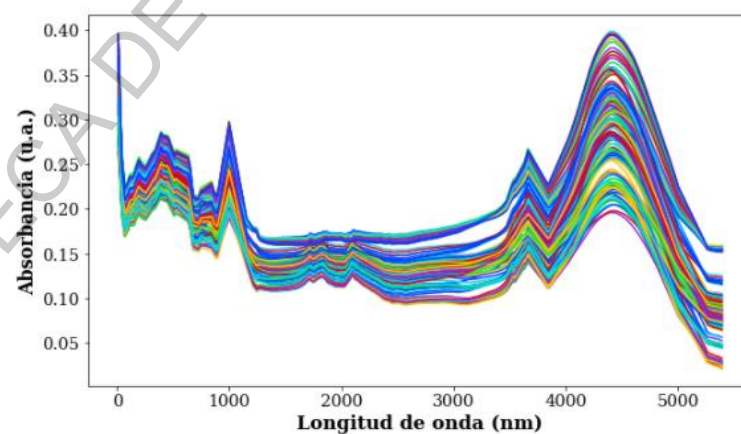
Anexo E

Segunda derivada de Savitzky-Golay aplicada a la humedad espectral del Modelo completo en muestras de harina de arroz



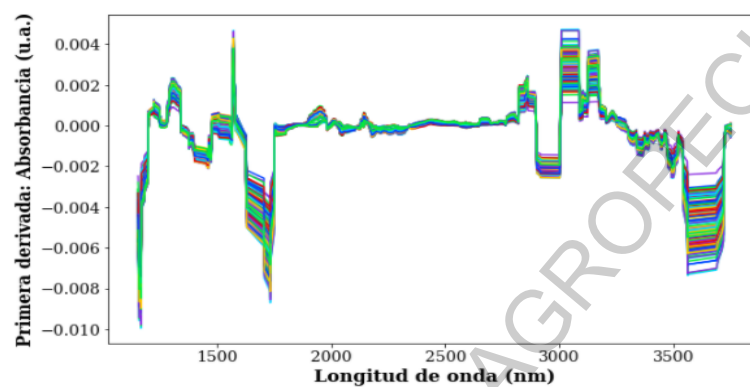
Anexo F

Espectros reducidos para la determinación de humedad por PLS-R para muestras de arroz



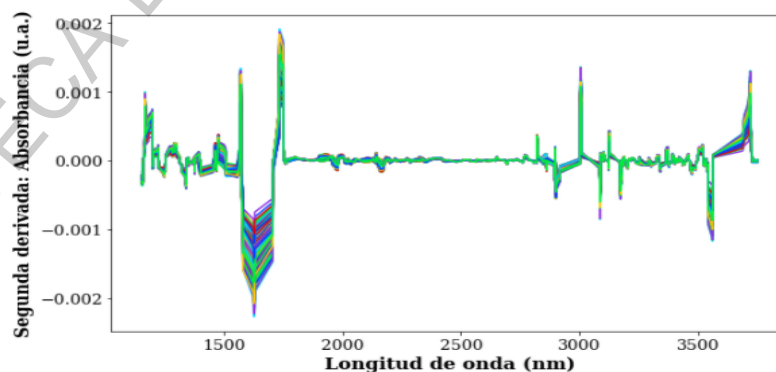
Anexo G

Pre-procesamiento con la primera derivada de Savitzky-Golay del Modelo reducido en muestras de arroz



Anexo H

Pre-procesamiento con la segunda derivada de Savitzky-Golay del Modelo reducido en muestras de arroz



Anexo I

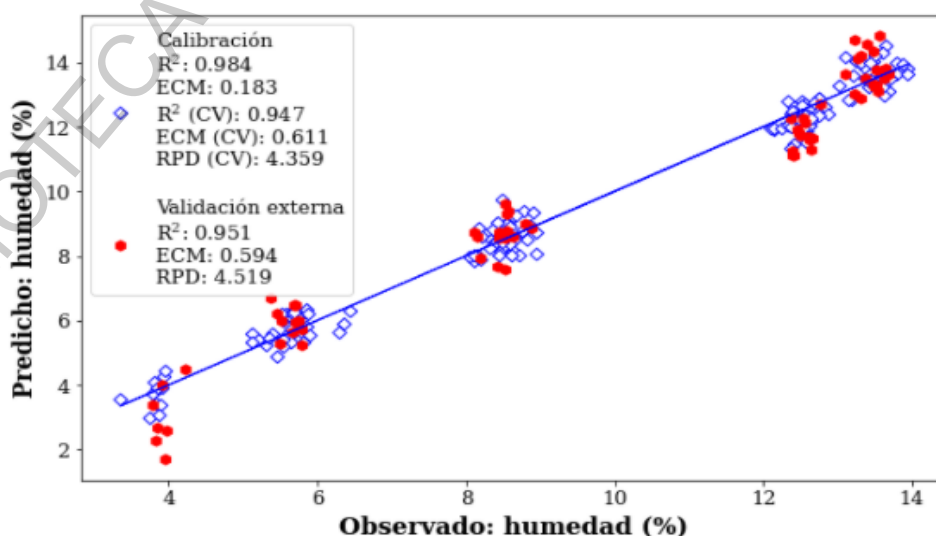
Indicadores estadísticos del modelo reducido de humedad para muestras de harina de arroz

Pre-procesamiento	Calibración					Validación				
	Rango de humedad (%)	Nº	Variable latente	ECM _{CV}	R ² _{CV}	Rango de humedad (%)	Nº	ECM _v	R ² _v	RPD
1 ^{ra} derivada de Savitzky-Golay	3.35 - 13.94	158	12	0.611	0.947	3.79 - 13.69	68	0.594	0.951	4.519
2 ^{da} derivada de Savitzky-Golay	3.35 - 13.94	158	4	0.925	0.920	--	--	--	--	--

ECM_{CV} = Error cuadrático medio de validación cruzada; R²_{CV} = Coeficiente de determinación de validación cruzada; ECM_v = Error cuadrático medio de validación externa; R²_v = Coeficiente de determinación de validación externa; RPD = Relación de la desviación de predicción

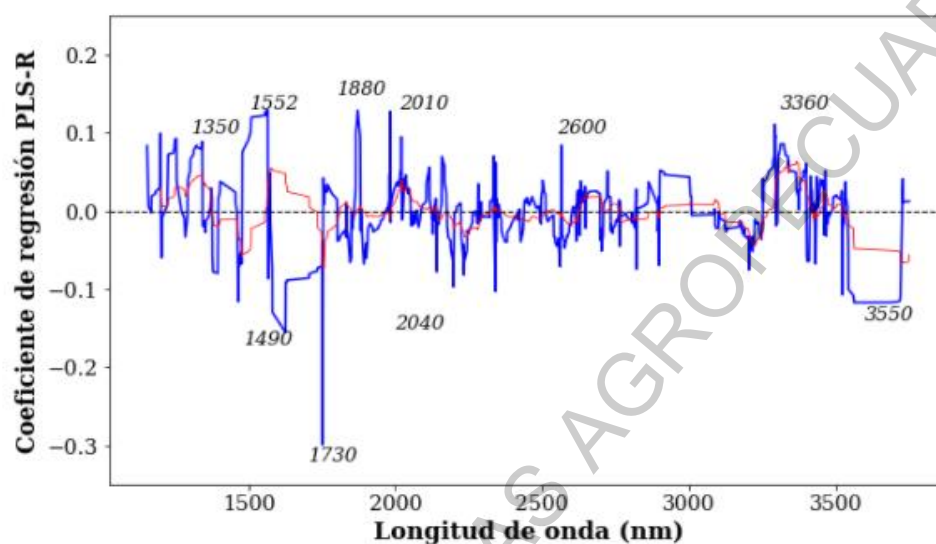
Anexo J

Valores observados y predichos de humedad del Modelo reducido (PLS-R) en muestras de harina de arroz



Anexo K

Modelo reducido con coeficientes de regresión para estimar la humedad en muestras en harina de arroz



Anexo L

Fotos del desarrollo experimental



(a)



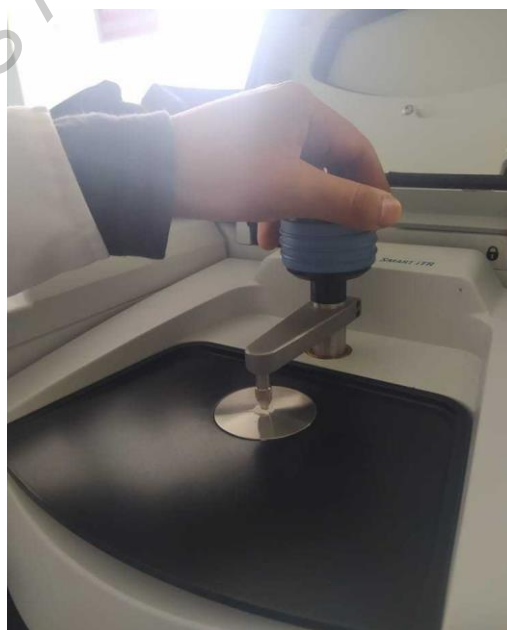
(b)



(c)



(d)



(e)

(a) Colocando las muestras de harina de arroz en la estufa para medir la humedad por el método gravimétrico (b) Diferentes condiciones de la muestra (c) Pesado de la muestra de harina de arroz, previamente acondicionada, para hacer la lectura de sus espectros (d) y (e) Colocando las muestras en el equipo espectrofotométrico infrarrojo-medio para obtener los espectros.



UNIVERSIDAD NACIONAL DE TRUJILLO

UNT

FACULTAD DE CIENCIAS AGROPECUARIAS

Escuela Profesional de Ingeniería Agroindustrial

UNIVERSIDAD NACIONAL DE TRUJILLO

DECLARACIÓN JURADA

Los AUTORES suscritos en el presente documento DECLARAMOS BAJO JURAMENTO que somos los responsables legales de la calidad y originalidad del contenido del Proyecto de Investigación Científica, así como, del Informe de la Investigación Científica realizado.

TÍTULO:

Validación de un método rápido para la determinación de humedad de harina de arroz (*Oryza sativa* L.), mediante espectroscopia infrarroja por transformada de Fourier (FTIR)

PROYECTO DE INVESTIGACIÓN CIENTÍFICAINFORME FINAL DE INVESTIGACIÓN CIENTÍFICA

PROYECTO DE TRABAJO DE INVESTIGACIÓN (PREGRADO)

()

TRABAJO DE INVESTIGACIÓN (PREGRADO)

()

PROYECTO DE TESIS PREGRADO

()

TESIS PREGRADO

(X)

PROYECTO DE TESIS MAESTRÍA

()

TESIS MAESTRÍA

()

PROYECTO DE TESIS DOCTORADO

()

TESIS DOCTORADO

()

Equipo Investigador Integrado por:

N°	Apellidos y Nombres	Facultad	Departamento Académico	Categoría Docente Asesor	Código Docente Asesor Número Matrícula del Estudiante	Autor Coautor asesor
01	Otiniano Burgos, Anthony Brando	Ciencias Agropecuarias	Ciencias Agroindustriales	---	1052400215	Autor
02	Barraza Jauregui, Gabriela del Carmen	Ciencias Agropecuarias	Ciencias Agroindustriales	Asociado	5577	Asesora
03						

Trujillo, 24 de Enero de 2023.

OTINIANO BURGOS, ANTHONY BRANDO

70474025
DNI DEL TESISISTA

MSc. BARRAZA JAUREGUI, GABRIELA DEL CARMEN

08715119
DNI DEL ASESOR

Este formato debe ser llenado, firmado, adjuntado al final del documento del PIC, del Informe de Tesis, Trabajo de Investigación respectivamente.

Av. Juan Pablo II S/N – Trujillo – Perú.

e-mail: agroindustrial@unitru.edu.pe
www.unitru.edu.pe



UNIVERSIDAD NACIONAL DE TRUJILLO

UNT

FACULTAD DE CIENCIAS AGROPECUARIAS

Escuela Profesional de Ingeniería Agroindustrial

UNIVERSIDAD NACIONAL DE TRUJILLO

CARTA DE AUTORIZACIÓN DE PUBLICACIÓN DE TRABAJO DE INVESTIGACION EN REPOSITORIO DIGITAL RENATI-SUNEDU

Trujillo, 24 de Enero de 2023

Los autores suscritos del INFORME FINAL DE INVESTIGACIÓN CIENTÍFICA

Titulado:

Validación de un método rápido para la determinación de humedad de harina de arroz (*Oryza sativa* L.), mediante espectroscopía infrarroja por transformada de Fourier (FTIR)

AUTORIZAMOS SU PUBLICACIÓN EN EL REPOSITORIO DIGITAL INSTITUCIONAL, REPOSITORIO RENATI-SUNEDU, ALICIA-CONCYTEC, CON EL SIGUIENTE TIPO DE ACCESO:

- A. Acceso Abierto: ☒ (datos del autor y resumen del trabajo)
 B. Acceso Restringido ☐
 C. No autorizo su Publicación ☐

Si eligió la opción restringido o NO autoriza su publicación sírvase justificar

ESTUDIANTES DE PREGRADO:	TRABAJO DE INVESTIGACIÓN	☐	TESIS	☒
ESTUDIANTES DE POSTGRADO:	TESIS MAESTRIA	☐	TESIS DOCTORADO	☐
DOCENTES:	INFORME DE INVESTIGACION	☐	OTROS	☐

El equipo investigador Integrado por:

Nº	Apellidos y Nombres	Facultad	CONDICION (NOMBRADO, CONTRATADO, EMERITO, estudiante, OTROS)	Código Docente Número Matrícula del Estudiante	Autor Coautor asesor
01	Otiniano Burgos, Anthony Brando	Ciencias Agropecuarias	----	1052400215	Autor
02	Barraza Jauregui, Gabriela del Carmen	Ciencias Agropecuarias	Nombrado	5577	Asesora
03					

OTINIANO BURGOS, ANTHONY BRANDO

70474025
DNI DEL TESISISTA

MSc. BARRAZA JAUREGUI, GABRIELA DEL CARMEN

08715119
DNI DEL ASESOR

'Este formato debe ser llenado, firmado y adjuntado en el Informe de Tesis y/o Trabajo de Investigación respectivamente 'Este formato en el caso de Informe de investigación científica docente debe ser llenado, firmado, escaneado y adjuntado en el sistema de www.picfedu.unitru.edu.pe

Av. Juan Pablo II S/N – Trujillo – Perú.

e-mail: agroindustrial@unitru.edu.pe
www.unitru.edu.pe