

NIR-Based Detection of Starch Adulteration in Soft Cheese Using Machine Learning Models

Jimmy Oblitas-Cruz¹; Jhoana Uriarte-Valdivia¹; André Rodríguez-León²

^{1,3}Universidad Privada del Norte, Perú, jimmy.oblitas@upn.edu.pe, N00272826@upn.pe

²Universidad Nacional de Cajamarca, Perú, arodriguezl@unc.edu.pe

Abstract— The control of adulteration in dairy products requires rapid, accurate, and non-destructive methods to ensure authenticity and quality. This study aimed to detect starch adulteration levels in soft cheese using NIR spectra (700–1000 nm) acquired through hyperspectral imaging (HSI) and machine learning models. A total of 20 averaged spectra were collected across five adulteration levels, and eight representative models were trained, ranging from penalized regression approaches to non-linear methods. Results show that Elastic Net achieved the highest performance ($R^2_{\text{test}} = 0.986$, $\text{RMSE} = 0.628$), outperforming Lasso and Ridge, while more complex models such as Random Forest, Gradient Boosting, and MLP exhibited overfitting. After hyperparameter optimization via Grid Search, Ridge reached $R^2_{\text{test}} = 0.981$ and $\text{RMSE} = 0.739$, confirming the robustness of penalized linear methods when working with reduced datasets. These findings indicate that the combination of NIR spectroscopy and regularized regression provides an efficient and feasible tool for the rapid detection of starch adulteration in soft cheese, enabling future applications in in-plant quality control and routine monitoring.

Keywords—Adulteration, HSI, fresh cheese, machine learning, Elastic Net, Ridge, spectral multicollinearity.

Detección de adulteración con almidón en queso blanco utilizando modelos de aprendizaje automático basados en espectroscopía NIR

Jimmy Oblitas-Cruz¹, Jhoana Uriarte-Valdivia¹, André Rodríguez-León²

^{1,3}Universidad Privada del Norte, Perú, jimmy.oblitas@upn.edu.pe, N00272826@upn.pe

²Universidad Nacional de Cajamarca, Perú, arodriguezl@unc.edu.pe

Resumen– El control de adulteración en productos lácteos requiere métodos rápidos, precisos y no destructivos para garantizar autenticidad y calidad. Este estudio tuvo como objetivo detectar niveles de adulteración con almidón en queso blanco mantecoso utilizando espectros NIR de 700–1000 nm obtenidos mediante HSI y modelos de aprendizaje automático. Se adquirieron 20 espectros promediados en cinco niveles de adulteración y se entrenaron ocho modelos representativos, desde regresión penalizada hasta métodos no lineales. Los resultados muestran que Elastic Net alcanzó $R^2_{test} = 0.986$ y $RMSE = 0.628$, superando a Lasso y Ridge, mientras que modelos complejos como Random Forest, Gradient Boosting y MLP evidenciaron sobreajuste. Tras la optimización de hiperparámetros mediante Grid Search, Ridge obtuvo $R^2_{test} = 0.981$ y $RMSE = 0.739$, confirmando la robustez de los métodos lineales penalizados sobre datasets reducidos. Se concluye que la combinación de espectroscopía NIR y regresión regularizada constituye una herramienta eficiente y viable para la detección rápida de adulteración en queso blanco, permitiendo aplicaciones futuras en control de calidad en planta y monitoreo rutinario.

Palabras clave– Adulteración, HSI, queso, aprendizaje automático, ElasticNet, Ridge, multicolinealidad espectral.

I. INTRODUCCIÓN

La adulteración de productos lácteos mediante la incorporación de insumos de bajo costo, como el almidón de maíz, constituye una práctica creciente que compromete la autenticidad del producto, su valor nutricional y la confianza del consumidor [1]. En el caso del queso mantecoso (blando), esta problemática se ve agravada por su alta variabilidad composicional y por la limitada aplicación de métodos analíticos rápidos y no destructivos para su control rutinario [2].

La adulteración del queso abarca varias prácticas como son la incorporación de grasas vegetales, la adición de proteínas no lácteas, la dilución con agua y la tergiversación del origen geográfico [3]. En mercados artesanales y de pequeña escala, donde la supervisión es limitada y la variabilidad composicional del queso es alta, la detección de adulteración sigue dependiendo de métodos químicos tradicionales que son lentos, destructivos y poco adecuados para controles rutinarios [4]. La calidad del queso terminado es evaluada típicamente mediante análisis sensorial; sin embargo, este enfoque resulta costoso, subjetivo y limitado en

términos operativos, lo que ha impulsado el interés por técnicas espectroscópicas con características no destructivas y rápidas, como la espectroscopía de infrarrojo cercano (NIR), como alternativas analíticas emergentes para el control de calidad de alimentos. Este escenario es evidencia de la ausencia de métodos automatizables, rápidos y no destructivos que permitan identificar adulteración en tiempo real y con mínima manipulación de la muestra [5].

Esta problemática en la industria de alimentos ha generado estudios tanto en el sector empresarial como en los sectores relacionados a la regulación o investigación, los cuales han encontrado en las técnicas de espectroscopía una importante herramienta para la verificación de posibles adulteraciones [6]. Estas técnicas espectroscópicas han surgido como herramientas potentes, no destructivas, rápidas y, a menudo, rentables para detectar la adulteración en el sector lácteo y en específico en la industria del queso [7].

Aunque la espectroscopía en imágenes hiperespectrales ha demostrado eficacia en la discriminación de parámetros fisicoquímicos en matrices alimentarias, aún existe una limitada aplicación de modelos de aprendizaje automático para cuantificar adulteración en queso mantecoso (blando) bajo condiciones experimentales reducidas en poblaciones queseras como es la región de Cajamarca en Perú [8]. Esto abre la necesidad de evaluar si los algoritmos regulares y no lineales pueden capturar variaciones espectrales sutiles asociadas al almidón, especialmente en datasets pequeños y altamente correlacionados.

A fin de contextualizar la detección de adulteración con almidón en queso mantecoso (blando) mediante espectroscopía NIR, resulta relevante revisar investigaciones recientes con enfoques similares. Joolaei et al. [3] analizó la autenticidad de productos lácteos y concluyeron que los métodos basados en espectroscopía NIR combinados con PLS (Partial Least Squares) y SVM (Support Vector Machine) han logrado detectar adulteraciones por agua, grasa y aditivos como almidón y microcelulosa en más de veinte variedades de queso, destacando su capacidad para identificar fraudes incluso en matrices con alta variabilidad fisicoquímica. Otros trabajos refuerzan la utilidad del aprendizaje automático. Silveira et al. [9] evaluaron mezclas lácteas con FT-NIR (Fourier Transform Near-Infrared) y Mamgain et al. [10]

entrenaron una CNN (Convolutional Neural Network) para adulterantes en leche, alcanzando un alto valor de exactitud de clasificación. En conjunto, estos estudios muestran que NIR combinada con PLS, SVM o redes neuronales logra precisiones y exactitudes elevadas lo que respalda el objetivo de la presente investigación de hacer uso de métodos regulares para detectar adulteración en queso.

En este contexto, el uso combinado de espectros NIR obtenidos mediante (Hyperspectral Imaging) HSI y modelos penalizados ofrece una alternativa prometedora para desarrollar sistemas de control rápido, precisos y aplicables en entornos de producción artesanal o industrial.

II. MATERIALES Y MÉTODOS

En esta sección se describe el enfoque de aprendizaje automático empleado para la detección de adulteración con almidón en queso mantecoso (blando) mediante espectros NIR. El modelo fue desarrollado en Google Colab utilizando un conjunto de datos conformado por muestras con diferentes niveles de adulteración, registradas entre 700 y 1000 nm. Tal como se ilustra en la Figura 1, el proceso inicia con la adquisición y organización del dataset espectral, seguido de una fase de preprocesamiento NIR que asegura señales comparables y libres de sesgo instrumental. Posteriormente, los datos se dividen en conjuntos de entrenamiento y prueba, sobre los cuales se evalúan múltiples modelos de aprendizaje automático con el fin de identificar el algoritmo más robusto y con mayor capacidad predictiva. De la misma manera, como paso final, los modelos pasaron por un proceso de optimización mediante GridSearch a fin de comparar cuál de los modelos presentó mejor optimización. Este flujo metodológico permite demostrar que es posible detectar adulteración en queso mantecoso (blando) con alta exactitud utilizando técnicas no destructivas basadas en espectroscopía NIR.

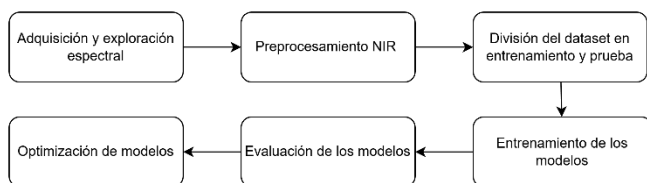


Fig. 1 Flujo de obtención de los modelos de predicción.

A. Descripción del dataset

El estudio empleó un conjunto experimental compuesto por 20 muestras de queso mantecoso (blando), elaboradas bajo un protocolo controlado para garantizar homogeneidad en humedad, contenido graso y estructura proteica. Las muestras fueron adulteradas con almidón de maíz en cinco niveles definidos: 0%, 3%, 6%, 9% y 12%, con cuatro réplicas independientes por nivel, lo cual permite capturar variabilidad real sin comprometer la estabilidad estadística del modelo. Cada unidad experimental fue procesada bajo idénticas

condiciones de temperatura, manipulación y empaquetado para evitar efectos de confusión.

Los espectros NIR fueron registrados entre 700 y 1000 nm, generando 37 longitudes de onda seleccionadas que incluyen regiones asociadas a bandas de combinación y sobretonos de enlaces O–H, C–H y C–O, altamente sensibles a cambios estructurales inducidos por la presencia de polisacáridos como el almidón.

B. Instrumentación y adquisición espectral

Se usó la cámara hiperespectral del modelo Pika L (400-1000 nm) de Resonon (Resonon, EE. UU.) para obtener una ampliación del espectro NIR. Para el presente estudio se tomó el espectro NIR de 700 nm a 1000 nm. El software utilizado para la extracción de la data fue Spectronon 2020. Se realizaron tres mediciones consecutivas por muestra, promediadas posteriormente para reducir ruido aleatorio y fluctuaciones de la fuente. El sistema operó bajo condiciones ambientales controladas (25 ± 1 °C).

C. Preprocesamiento espectral

El preprocesamiento se orientó a mejorar la calidad espectral y reducir artefactos asociados a la dispersión óptica del queso mantecoso (blando) en el rango NIR. En primera instancia, los espectros fueron suavizados mediante el algoritmo Savitzky–Golay, aplicando una derivada de primer orden, lo cual permitió resaltar cambios sutiles en la curvatura espectral derivados de la presencia de almidón. Esta operación eliminó componentes de baja frecuencia y mejoró la delimitación de pequeñas inflexiones entre 700 y 1000 nm, facilitando la identificación de patrones característicos de cada nivel de adulteración.

Posteriormente, se aplicó Standard Normal Variate (SNV) como corrección individual por dispersión, mitigando variaciones asociadas a diferencias físicas entre muestras, tales como humedad superficial, rugosidad y microestructura interna del queso. Esta corrección redujo efectos multiplicativos y aditivos propios de matrices lácteas heterogéneas, estabilizando la respuesta óptica en todas las longitudes de onda.

Ambas transformaciones fueron integradas mediante pipelines reproducibles, garantizando que cada etapa se ajustara únicamente con los datos de entrenamiento y preservando la integridad estadística del proceso de modelado.

D. División del conjunto de datos

El conjunto de datos fue particionado utilizando un esquema hold-out con el fin de evaluar el desempeño real de los modelos bajo condiciones de entrenamiento limitadas, propias de experimentos con matrices alimentarias perecibles. Se empleó una división 80/20, asignando 16 muestras al conjunto de entrenamiento y 4 al conjunto de prueba, garantizando que cada nivel de adulteración estuviera representado en ambas particiones. Para asegurar reproducibilidad, todas las particiones fueron generadas con la semilla fija `random_state=42`, evitando variaciones aleatorias que afecten la estabilidad del RMSE (Root Mean Square Error) y del coeficiente de determinación.

Dado el tamaño reducido del dataset, no se aplicó validación cruzada k-fold para evitar sobreajuste estructural ni fuga de información, un riesgo frecuente cuando existen pocas réplicas por nivel. En su lugar, se optó por pipelines cerrados donde el preprocesamiento se calcula únicamente sobre los datos de entrenamiento, preservando la independencia entre particiones. Este enfoque permite evaluar de manera más realista la capacidad de generalización del modelo ante nuevas muestras lácteas adulteradas y constituye una práctica recomendada en estudios espectrales con bajo N por clase.

E. Entrenamiento de los modelos

El entrenamiento se efectuó mediante una arquitectura basada en pipelines que integran el preprocesamiento (escalado estándar) con cada algoritmo, garantizando que las transformaciones se ajusten únicamente a los datos de entrenamiento. Se evaluaron ocho modelos representativos de las familias lineales, penalizadas, kernelizados, de ensamble y redes neuronales. Los modelos usados se muestran en la tabla I.

TABLA I
 MODELOS USADOS

Tipo	Modelo	Detalle
Modelos lineales y penalizados	Ridge Regression (alpha=1.0)	Utiliza regularización L2 para controlar la multicolinealidad inherente a los espectros NIR. El parámetro alpha regula la magnitud del castigo, estabilizando soluciones en matrices con alta correlación entre bandas.
	Lasso (alpha=0.01):	Incorpora penalización L1 para promover el enmascaramiento de coeficientes irrelevantes. Su sensibilidad a valores pequeños de alpha permite identificar longitudes de onda potencialmente informativas
	Elastic Net (alpha=0.1, l1_ratio=0.5):	Combina penalizaciones L1/L2 para mantener estabilidad numérica mientras realiza selección suave de características. La elección de l1_ratio=0.5 balancea ambos efectos, adecuado para espectros con correlación moderada.
Modelos no lineales basados en kernels	Support Vector Regression (RBF):	Entrenado con C=100, gamma=0.001, parámetros que controlan la flexibilidad del modelo y el alcance del núcleo Gaussiano. C elevado permite un ajuste más fino al conjunto de entrenamiento, mientras que gamma pequeño evita sobreajuste extremo en datasets pequeños.
Modelos de ensamble basados en árboles	Random Forest (n_estimators=500, max_depth=None):	Configurado para explorar múltiples particiones sin restricción de profundidad, capturando interacciones no lineales entre bandas.
	Gradient Boosting (n_estimators=300, learning_rate=0.05, max_depth=3):	Explora estructuras residuales en etapas sucesivas. El learning_rate moderado y la profundidad baja controlan la tendencia al sobreajuste típica en datasets reducidos.
	XGBoost (n_estimators=300, max_depth=3, subsample=0.8, eta=0.05)	Incorpora submuestreo para reducir varianza y mejorar generalización. Los parámetros seleccionados favorecen estabilidad en problemas con N pequeño.

Redes neuronales artificiales	MLPRegressor (hidden_layer_sizes=(50,50), activation='relu', alpha=0.001, max_iter=5000);	Arquitectura compacta de dos capas densas, diseñada para capturar no linealidades sin caer en sobreajuste severo. La regularización L2 (alpha) actúa como mecanismo de control de magnitud de pesos.
-------------------------------	---	--

En todos los casos, la optimización se realizó mediante entrenamiento hold-out, sin validación cruzada, para evitar inestabilidad en modelos paramétricamente complejos. Los parámetros fueron seleccionados para representar configuraciones razonables en espectros NIR con bajo tamaño muestral, priorizando comparabilidad entre algoritmos.

F. Evaluación del rendimiento

La evaluación comparativa inicial se llevó a cabo analizando los valores de R² y RMSE obtenidos por los modelos base.

G. Optimización de modelos

Durante la fase de optimización se aplicó un proceso de Grid Search bajo un esquema hold-out, evaluando sistemáticamente los hiperparámetros más influyentes de cada algoritmo.

III. RESULTADOS

La Figura 2 muestra los espectros NIR promedio registrados entre 700 y 1000 nm para los distintos niveles de adulteración. En este intervalo se observan variaciones suaves pero consistentes en la respuesta espectral, particularmente en zonas donde suelen manifestarse sobretonos y combinaciones débiles asociados a enlaces C–H y O–H presentes en la matriz láctea. El queso mantecoso (blando), compuesto principalmente por agua, proteínas y grasa, exhibe patrones característicos de absorción en este rango, los cuales sirven como referencia para detectar cambios inducidos por la incorporación de almidón.

Aunque las diferencias entre curvas son moderadas, su consistencia a través de todo el rango permite que los modelos de aprendizaje automático identifiquen patrones distintivos asociados al incremento de polisacáridos. Asimismo, la separación visual entre los niveles de adulteración confirma que la presencia de almidón altera de manera medible la interacción de la luz NIR con la estructura interna del queso mantecoso (blando), proporcionando señales suficientes para técnicas de regresión multivariante.

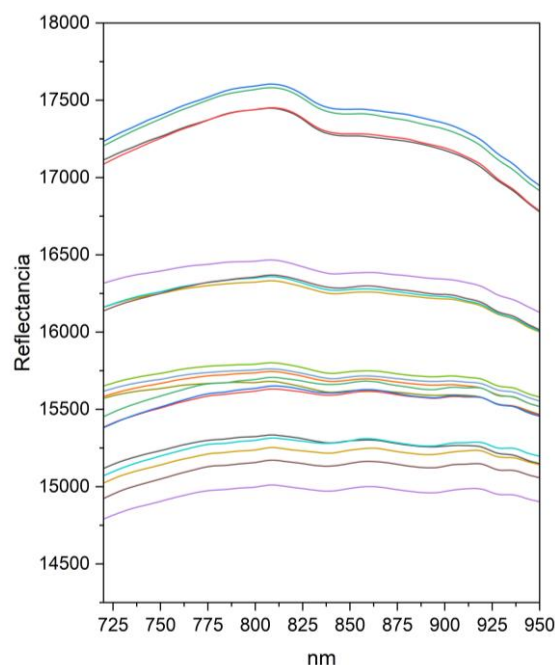


Fig.2 Espectros de quesos obtenidos.

Evaluación del rendimiento

La Tabla II presenta los valores de R^2 de los diferentes modelos entrenados, donde los modelos penalizados (ElasticNet, Lasso y Ridge) muestran los valores más altos y consistentes. Asimismo, al analizar el RMSE se muestra el mismo patrón: mientras que ElasticNet y Lasso mantienen errores bajos y estables, modelos complejos como Random Forest, Gradient Boosting y MLP (Multi-Layer Perceptron) presentan RMSE mayores, indicando sobreajuste debido al tamaño reducido del dataset ($N=20$).

TABLA II.
RESULTADOS COMPARATIVOS DE LOS MODELOS EVALUADOS

Modelo entrenado	R^2_{test}	$\text{RMSE}_{\text{test}}$
ElasticNet	0.99	0.628
Lasso	0.98	0.666
SVR-RBF	0.96	1.015
Ridge	0.96	1.123
RandomForest	0.94	1.268
GradientBoosting	0.84	2.134
XGBoost	0.94	1.269
MLP	0.65	3.154

La figura 3 muestra ambos indicadores a la vez para poder compararlos, donde se muestra que el modelo ElasticNet emerge como el mejor predictor global en su configuración base, alcanzando: $R^2_{\text{test}} = 0.986$ y $\text{RMSE} = 0.628$; superando a Lasso (0.985, $\text{RMSE}=0.666$) y Ridge (0.981, $\text{RMSE}=0.740$). Estos resultados demuestran que la combinación de penalización L1/L2 proporciona un equilibrio óptimo entre reducción de varianza y estabilidad frente a multicolinealidad espectral.

La Figura 4 muestra la curva de paridad del mejor modelo, donde las predicciones se alinean estrechamente con la diagonal ideal. Esto indica que el modelo reproduce adecuadamente las concentraciones reales de adulteración sin sesgos sistemáticos.

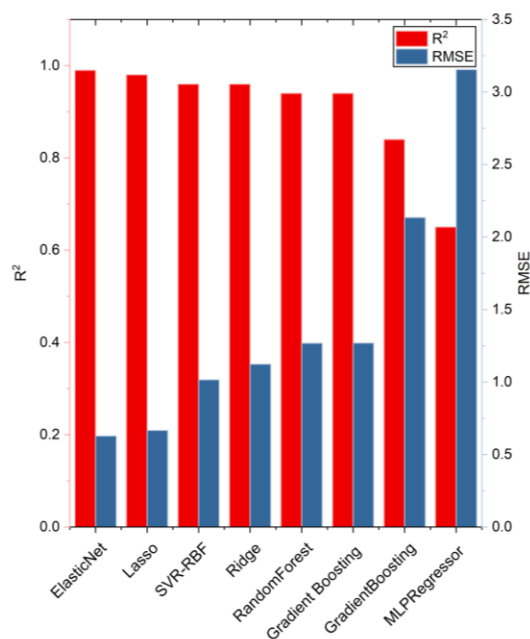


Fig.3 Comparativa del valor de R^2 y RMSN de los modelos entrenados

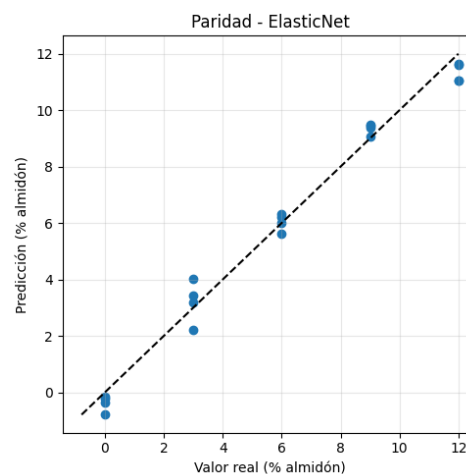


Fig.4 Curva de paridad para el modelo Elastic Net

Optimización de modelos

Para los modelos penalizados, se exploraron distintos niveles de regularización. En Ridge se utilizaron valores de α entre 0.01 y 100, mientras que en Lasso se consideraron α s desde 0.001 hasta 1 con el fin de analizar su capacidad para inducir sparsidad en matrices altamente correlacionadas. En Elastic Net se ajustaron simultáneamente los valores de α (0.01, 0.1 y 1) y las proporciones ll_ratio (0.1, 0.5 y 0.9), buscando un balance adecuado entre estabilidad numérica y selección de bandas espectrales relevantes.

En el modelo SVR con núcleo RBF (Radial Basis Function) se evaluaron combinaciones de C en $\{1, 10, 100\}$ y γ en $\{0.0001, 0.001, 0.01\}$, cubriendo configuraciones desde márgenes suaves hasta ajustes de alta sensibilidad. Para los métodos basados en árboles, Random Forest se optimizó variando el número de estimadores entre 200 y 500 y la profundidad máxima entre None, 5 y 10, mientras que en Gradient Boosting se ajustaron conjuntamente los valores de $n_estimators$ (200 y 300), $learning_rate$ (0.01 y 0.05) y max_depth (2 y 3), con el fin de evaluar la interacción entre complejidad del modelo y estabilidad frente al sobreajuste.

El modelo MLPRegressor se configuró evaluando arquitecturas densas de dos capas ocultas con tamaños de (30,30) y (50,50), combinadas con valores de regularización α entre 0.001 y 0.01 y un máximo de 5000 iteraciones, garantizando convergencia estable en un dataset pequeño. En el caso de XGBoost, la optimización consideró un conjunto más amplio de hiperparámetros, incluyendo $n_estimators$ entre 300 y 900, profundidades máximas (max_depth) de 2 a 4, tasas de aprendizaje ($learning_rate$) en el rango 0.03–0.80, así como valores de subsample entre 0.8 y 1.0 y $colsample_bytree$ entre 0.7 y 1.0. Esta estrategia permitió explorar configuraciones con diferentes niveles de complejidad y regularización, adecuadas para evitar sobreajuste en problemas espectrales con tamaño muestral reducido. Los resultados del proceso de optimización se presentan en la Tabla III.

TABLA III.
Resultados comparativos de los modelos evaluados

Modelo Optimizado	R^2_{test}	RMSE_test
Ridge	0.981	0.739
ElasticNet	0.979	0.765
Lasso	0.978	0.777
SVR-RBF	0.974	0.847
GradientBoosting	0.946	1.237
RandomForest	0.944	1.266
MLP	0.868	1.943
XGBoost	0.867	1.947

La figura 5 muestra los indicadores de los modelos optimizados, donde el modelo Ridge optimizado alcanzó el mejor desempeño posterior al ajuste, con $R^2_{\text{test}} = 0.981$ y $\text{RMSE}_{\text{test}} = 0.739$, demostrando mayor estabilidad y menor

sensibilidad a variaciones del conjunto de datos en comparación con ElasticNet optimizado. Del mismo modo en la Figura 6 se muestra la curva de paridad, destacándose la cercanía de los valores reales a la curva de predicción.

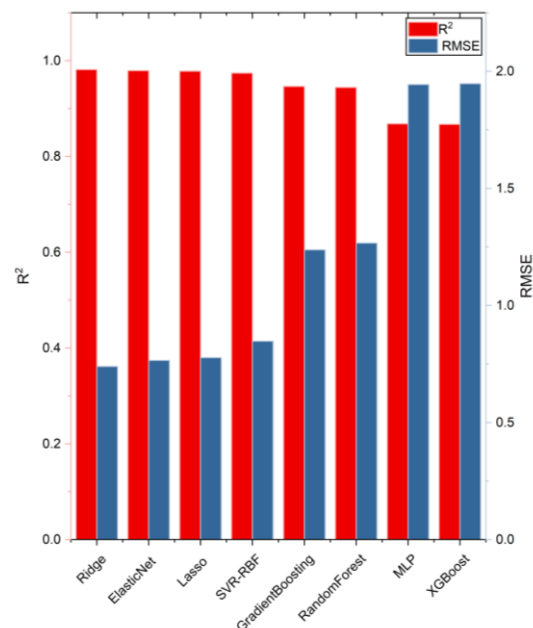


Fig.5 Comparativa del valor de R^2 y RMSN de los modelos optimizados

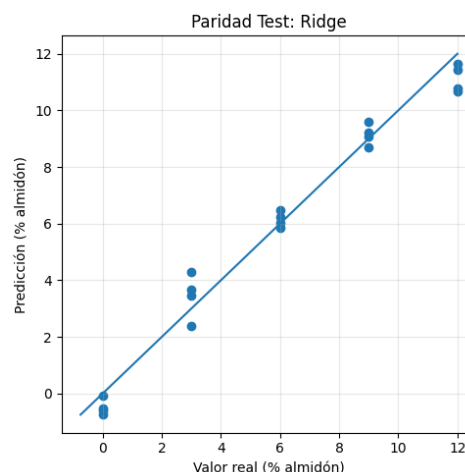


Fig.6 Curva de paridad para el modelo Ridge optimizado.

IV. DISCUSIÓN

Los resultados obtenidos muestran que los modelos penalizados lograron los coeficientes de determinación e índice de error más precisos en la predicción del nivel de adulteración, alcanzando R^2_{test} entre 0.98 y 0.99, valores coherentes con la magnitud de variación espectral inducida por el almidón en el rango 700–1000 nm. Esta superioridad frente a algoritmos más complejos se explica por la alta colinealidad inherente a los espectros NIR y por el tamaño

muestral reducido, condiciones donde las penalizaciones L1/L2 estabilizan el ajuste y controlan la varianza.

Resultados similares fueron reportados por Barreto et al. [11], quienes emplearon PLSR para cuantificar almidón en queso mantecoso (blando) y obtuvieron R^2 superiores a 0.99, demostrando que los cambios ópticos asociados al incremento de polisacáridos son detectables incluso a bajas concentraciones. Asimismo, los valores alcanzados se encuentran dentro de los rangos señalados por da Silva et al. [12], quienes reportaron precisiones superiores al 94% en detección de adulterantes en quesos mediante NIR portátil. Las similitudes entre estudios indican que la respuesta NIR del queso está fuertemente influenciada por aumentos en humedad, variaciones en la fase proteica y la absorción característica de almidón, lo que facilita la predicción por métodos multivariados [13]. La diferencia respecto a modelos neuronales observada en otras investigaciones se debe a que estos requieren mayor variabilidad composicional para evitar sobreajuste, mientras que los modelos regularizados capturan de manera más eficiente las relaciones espectrales lineales en matrices lácteas con colinealidad marcada.

Los valores obtenidos tras la optimización confirman que la regresión penalizada mantiene un rendimiento estable en matrices lácteas, donde la señal NIR refleja principalmente la interacción entre agua, proteínas y polisacáridos. ElasticNet, Lasso y Ridge conservaron R^2 cercanos a 0.98, lo que indica que estas técnicas se adaptan adecuadamente a la estructura espectral del queso mantecoso (blando), caracterizada por bandas amplias, altamente correlacionadas y sensibles a diferencias mínimas de composición. Esto coincide con Joolaei et al. [14], quienes en su revisión sugieren que el incremento de humedad y la capacidad del almidón para retener agua modifican la absorción óptica del queso, generando patrones detectables incluso con pocas bandas. La coherencia entre estudios refuerza la idea de que la adulteración con almidón altera la matriz láctea a nivel físico-químico de forma suficiente para ser modelada mediante regresión multivariante regularizada.

La comparación con investigaciones previas muestra que, aunque el presente estudio usó un rango espectral más reducido (700–1000 nm), los modelos penalizados alcanzaron niveles de precisión comparables a los obtenidos mediante HSI de 200–1000 nm o equipos FT-NIR. Esto sugiere que los cambios inducidos por el almidón, como incremento de humedad, modificación del índice de refracción y variaciones en la estructura interna del cuajo, generan contrastes suficientes para ser capturados incluso en regiones NIR de baja energía. La consistencia en los valores de R^2 entre diferentes autores indica que la adulteración con almidón modifica tanto la dispersión como la absorción de la luz, afectando la respuesta óptica de enlaces O–H y C–H del queso, una propiedad que favorece el uso de métodos lineales y penalizados.

La diferencia observada entre modelos penalizados y algoritmos más complejos como MLP o XGBoost se explica porque estos últimos requieren un mayor número de muestras

y de variabilidad experimental para estabilizar sus parámetros internos. En matrices lácteas, donde la colinealidad es elevada y el rango espectral es limitado, los modelos penalizados resuelven con mayor eficiencia el problema de dimensionalidad y reducen el riesgo de sobreajuste. Este tipo de avances en el uso de métodos quimiométricos para explicar comportamientos biológicos en muestras de queso se dan en diversas soluciones como determinaciones de vida útil [15] y ahora con algoritmos más robustos que muestran la estabilidad como ElasticNet, Lasso y Ridge confirma que las relaciones espectrales asociadas a la presencia de almidón son esencialmente lineales y consistentes, lo que respalda su aplicación en sistemas de control en planta.

V. CONCLUSIONES

Los resultados evidencian que la espectroscopia NIR combinada con modelos de regresión penalizada permite predecir el nivel de adulteración con almidón en queso mantecoso (blando) con alta precisión. Los coeficientes de determinación obtenidos ($R^2_{\text{test}} = 0.96\text{--}0.99$) y los valores reducidos de error ($\text{RMSE}_{\text{test}} = 0.62\text{--}1.01$) demuestran que ElasticNet, Lasso y Ridge capturan adecuadamente las variaciones espectrales asociadas al almidón incluso en concentraciones bajas. Estos resultados confirman que los métodos lineales regularizados constituyen una alternativa robusta frente a modelos más complejos que requieren mayor cantidad de muestras para evitar sobreajuste.

La optimización mediante Grid Search permitió mejorar la estabilidad de los parámetros y confirmar la superioridad de los modelos penalizados, particularmente ElasticNet ($\alpha = 0.01$, $\text{ll_ratio} = 0.7$), cuyo rendimiento se mantuvo consistente en todas las métricas de validación y prueba. Esta coherencia respalda su aplicabilidad en escenarios reales donde la variabilidad composicional del queso mantecoso (blando) es alta y los volúmenes de datos suelen ser reducidos.

Los resultados también se alinean con la tendencia reportada en estudios previos, donde modelos quimiométricos como PLSR y SVM han alcanzado valores de R^2 superiores a 0.98 en la detección de adulterantes en lácteos. De este modo, el empleo de modelos penalizados ofrece un desempeño equivalente, con la ventaja adicional de interpretar mejor las bandas relevantes gracias a la estructura de regularización.

Para futuras implementaciones industriales, se recomienda complementar los modelos penalizados con técnicas de selección de bandas específicas (por ejemplo, VIP, CARS o intervalos adaptativos NIR), lo que permitiría reducir la dimensionalidad, disminuir la demanda computacional e incrementar la interpretabilidad del sistema antes de su despliegue en líneas de procesamiento o inspección rápida.

REFERENCES

- [1] N. Vásquez, C. Magán, J. Oblitas, T. Chuquizuta, H. Avila-George, y W. Castro, «Comparison between artificial neural network and partial least squares regression models for hardness modeling during the ripening

- process of Swiss-type cheese using spectral profiles», *Journal of Food Engineering*, vol. 219, pp. 8-15, feb. 2018, doi: 10.1016/j.jfoodeng.2017.09.008.
- [2] P. Joolaei Ahranjani, P. Joolaei Ahranjani, K. Dehghan, Z. Esfandiari, y G. Ferrentino, «Advances in spectroscopic techniques for the detection of cheese adulteration: A systematic review», *Food Chemistry: X*, vol. 29, p. 102685, jul. 2025, doi: 10.1016/j.fochx.2025.102685.
 - [3] P. Joolaei Ahranjani, K. Dehghan, Z. Esfandiari, y P. Joolaei Ahranjani, «A Systematic Review of Spectroscopic Techniques for Detecting Milk Adulteration», *Critical Reviews in Analytical Chemistry*, vol. 0, n.o 0, pp. 1-32, abr. 2025, doi: 10.1080/10408347.2025.2477535.
 - [4] J. Oblitas-Cruz, A. C. Miano, y G. Terrones, «Terahertz time-domain spectroscopy for the classification of mature cheese», en *Proc. LACCEI int. multi-conf. eng. educ. technol., Latin American and Caribbean Consortium of Engineering Institutions*, 2021. doi: 10.18687/LACCEI2021.1.1.59.
 - [5] H. Patel, V. Aru, K. M. Sørensen, y S. B. Engelsens, «Towards on-line cheese monitoring: Exploration of semi-hard cheeses using NIR and ¹H NMR spectroscopy», *Food Chemistry*, vol. 454, p. 139786, oct. 2024, doi: 10.1016/j.foodchem.2024.139786.
 - [6] W. Castro, J. Oblitas, L. Nuñez, I. Yoplac, H. Avila-George, y M. De-la-Torre, «Adulterant estimation in paprika powder using deep learning and chemometrics through near-infrared spectroscopy», *Neural Comput & Applic*, vol. 36, n.o 23, pp. 14263-14273, ago. 2024, doi: 10.1007/s00521-024-09830-8.
 - [7] L. G. Visconti, P. H. G. D. Diniz, D. D. de Sousa Fernandes, M. S. Rodriguez, y C. V. Di Anibal, «Authentication of grated hard cheeses and quantification of adulteration by FT-NIR spectroscopy and multivariate analysis», *International Dairy Journal*, vol. 158, p. 106035, nov. 2024, doi: 10.1016/j.idairyj.2024.106035.
 - [8] L. G. Visconti, M. S. Rodríguez, y C. V. Di Anibal, «Determination of grated hard cheeses adulteration by near infrared spectroscopy (NIR) and multivariate analysis», *International Dairy Journal*, vol. 104, p. 104647, may 2020, doi: 10.1016/j.idairyj.2020.104647.
 - [9] P. G. Silveira et al., «Detection and quantification of water adulteration in goat milk using FT-NIR spectroscopy combined with partial least squares and logistic regression», *Dairy Sci. Manag.*, vol. 2, n.o 1, p. 10, sep. 2025, doi: 10.1186/s44363-025-00008-8.
 - [10] A. Mamgain, V. Kumar, y S. Dash, «Image-Based Detection of Adulterants in Milk Using Convolutional Neural Network», *ACS Omega*, vol. 9, n.o 25, pp. 27158-27168, jun. 2024, doi: 10.1021/acsomega.4c01274.
 - [11] A. Barreto, J. P. Cruz-Tirado, R. Siche, y R. Quevedo, «Determination of starch content in adulterated fresh cheese using hyperspectral imaging», *Food Bioscience*, vol. 21, pp. 14-19, feb. 2018, doi: 10.1016/j.fbio.2017.10.009.
 - [12] M. L. da S. Medeiros, A. Freitas Lima, M. Correia Gonçalves, H. Teixeira Godoy, y D. Fernandes Barbin, «Portable near-infrared (NIR) spectrometer and chemometrics for rapid identification of butter cheese adulteration», *Food Chemistry*, vol. 425, p. 136461, nov. 2023, doi: 10.1016/j.foodchem.2023.136461.
 - [13] J. Oblitas-Cruz, Y. Cieza-Rimarachin, y J. Cruz-Oblitas, «Evaluation of the maturation of Swiss-type cheese in a prototype system of NIR spectroscopy and artificial intelligence», en *Proc. LACCEI int. multi-conf. eng. educ. technol., Latin American and Caribbean Consortium of Engineering Institutions*, 2025. doi: 10.18687/LACCEI2025.1.1.1025.
 - [14] P. Joolaei Ahranjani, P. Joolaei Ahranjani, K. Dehghan, Z. Esfandiari, y G. Ferrentino, «Advances in spectroscopic techniques for the detection of cheese adulteration: A systematic review», *Food Chemistry: X*, vol. 29, p. 102685, jul. 2025, doi: 10.1016/j.fochx.2025.102685.
 - [15] J. F. Oblitas-Cruz y J. A. Sánchez-González, «Application of Weibull analysis and artificial neural networks to predict the useful life of the vacuum packed soft cheese», *Revista Facultad de Ingeniería Universidad de Antioquia*, n.o 82, pp. 53-59, mar. 2017, doi: 10.17533/udea.redin.n82a07.