

## Sistemas cognitivos para la predicción de cáncer de mama

**AUTORES:** Jonathan Stalin Delgado Guerrero<sup>1</sup>

Jhon Rigner Meléndez Meléndez<sup>2</sup>

Raúl Huillca Huallparimachi<sup>3</sup>

**DIRECCIÓN PARA CORRESPONDENCIA:** [jonathan.delgado1@unmsm.edu.pe](mailto:jonathan.delgado1@unmsm.edu.pe)

**Fecha de recepción:** 2022-12-17

**Fecha de aceptación:** 2023-02-10

### RESUMEN

El objeto del presente artículo es la comparación de cuatro algoritmos de clasificación existentes con aprendizaje no supervisado: red neuronal, k-means, fuzzy c-means y hierarchical clustering y una alternativa de solución para predecir una eventual enfermedad cancerígena en las glándulas mamarias de una mujer. Para realizar el pronóstico presente, se ha hecho uso de un conjunto de datos que comprende un total de 589 registros; de las cuales podemos citar algunas variables, las según su importancia como son: media de distancias, desviación estándar, perímetro de masa, área de masa y suavidad. Se ha demostrado con los valores de las métricas bases y derivadas usadas para la evaluación de algoritmos de clasificación, que el algoritmo red neuronal presenta mejores resultados de predicción para la configuración del experimento seleccionado. Este modelo entrenado es almacenado en la base de datos a la cual accede el servicio Azure Machine Learning para construir el servicio web. Este servicio web construido, encapsula el proyecto de implementación del algoritmo en una función para poder implementar el resto de la arquitectura REST con un verbo de ejecución disponibles: POST. El formato de intercambio de datos usados es JSON tanto para la solicitud como para la respuesta del servicio.

**PALABRAS CLAVE:** predicción del cáncer de mama; cáncer de mama; comparación de aprendizaje automático.

### Cognitive systems for the prediction of breast cancer

#### ABSTRACT

The object of this article is the comparison of four existing classification algorithms with unsupervised learning: neural network, k-means, fuzzy c-means and hierarchical clustering and an alternative solution to predict a possible cancer disease in the mammary glands of a woman. To carry out the present forecast, use has been made of a data set comprising a total of 589 records; of which we can cite some variables, according to their importance such as: mean of

---

<sup>1</sup> Ingeniero en Sistemas Computacionales, Magíster en Educación Informática, Magíster en Sistemas de Información Gerencial, Guayaquil, Ecuador. E-mail: [jonathan.delgado1@unmsm.edu.pe](mailto:jonathan.delgado1@unmsm.edu.pe) Código ORCID <https://orcid.org/0000-0003-0724-3358>

<sup>2</sup> Ingeniero Industrial, Magíster en Gerencia de Proyectos de Ingeniería, Universidad Nacional Federico Villareal. Perú. E-mail: [jhon.melendez@unmsm.edu.pe](mailto:jhon.melendez@unmsm.edu.pe) Código ORCID <https://orcid.org/0000-0003-3613-2569>

<sup>3</sup> Bach. Ing. Sistemas, Ing. Sistemas, Mtro. Ciencias mención Informática, Cuzco, Perú. E-mail: [raul.huillca@unmsm.edu.pe](mailto:raul.huillca@unmsm.edu.pe) Código ORCID <https://orcid.org/0009-0003-1082-5153>

distances, standard deviation, perimeter of mass, area of mass and smoothness. It has been shown with the values of the base and derived metrics used for the evaluation of classification algorithms, that the neural network algorithm presents better prediction results for the configuration of the selected experiment. This trained model is stored in the database accessed by the Azure Machine Learning service to build the web service. This built web service encapsulates the algorithm implementation project in a function to be able to implement the rest of the REST architecture with an available execution verb: POST. The data exchange format used is JSON for both the request and the response of the service.

**KEYWORDS:** breast cancer prediction; breast cancer; machine learning compara.

## INTRODUCCIÓN

La presente investigación se refiere al tema de la detección de enfermedades cancerígenas en las glándulas mamarias. La característica principal de este tipo de enfermedad es el perímetro de la masa maestra. Para analizar esta problemática es necesario mencionar sus causas. Y dentro de ellas podemos citar la edad, la historia personal de cáncer de mama y los antecedentes familiares de cáncer de mama. La investigación de esta problemática es por el interés académico en donde nos interesamos por aportar herramientas recientes para la resolución de la problemática expuesta.

El conjunto de datos usado como modelo para la presente investigación está compuesto por seis columnas, cinco contienen datos cuantitativos y una cualitativa, y 589 registros; de estos un 85% de los mismo se usaron para el entrenamiento y un 15% para la prueba u validación. En efecto, se hizo un análisis de la data con los cuatro algoritmos de clasificación existentes de aprendizaje no supervisado: red neuronal, k-means, fuzzy c-means, hierarchical clustering y la arquitectura propuesta para después diferenciar cuál de todos tiene un buen grado de precisión.

Por otra parte, la investigación esta estructura en siete capítulos. En el primer capítulo se expone el estado del arte y la revisión literaria de los últimos cinco años. En el segundo capítulo se considera los procedimientos para la evaluación de los algoritmos de clasificación, dentro del mismo se identifican las variables de predicción, las características de valor real para cada núcleo, el conjunto de datos usados y la selección de datos. En el Capítulo tres realizamos la ejecución y evaluación de los algoritmos obteniendo de estos las matrices de confusión y los diagramas de puntos correspondientes. En el capítulo cuatro realizamos el análisis de los resultados en donde evaluamos la exactitud, precisión, sensibilidad y especificidad de los diferentes algoritmos. En el capítulo quinto determinamos la variable más importante. En el capítulo seis se expone la aplicación de los sistemas cognitivos de inteligencia artificial y la arquitectura propuesta para finalmente concluir con el capítulo siete donde se precisan las conclusiones.

## DESARROLLO

Para realizar la revisión de literatura se utilizaron los siguientes criterios de inclusión:

- Año de publicación: 5 años atrás.
- Idioma: inglés.
- Base de datos: Scopus.
- Área: Computación.

- Documento: Artículos.
- Palabras claves: breast cancer prediction, breast cancer machine learning compara\*.
- Prioridad de revisión: título, resumen, conclusiones.

Con base en la aplicación de los criterios de inclusión se determinaron 25 potenciales artículos como resultados. Dentro de estos se analizó el resumen para conocer si se usaron más de un algoritmo de clasificación, filtrándose a siete artículos los cuales fueron seleccionados.

Estos artículos utilizaron algoritmos de clasificación como: K-vecino más cercano (Algoritmo k-Nearest Neighbor - KNN), Máquinas de vectores de soporte (Support vector machine - SVM), Árbol de decisión (Decision tree - DT), Regresión logística (Logistic regression - LR), Red neuronal artificial (Artificial neural network - ANN), Naives Bayes (NB), Bosques Aleatorios (Random forest, RF).

Se evidencia que el árbol de decisión fue usado en todos los artículos, aunque no siempre obtuvo los mejores resultados.

En el trabajo (Nanglia et al., 2022) se propone un enfoque basado en el proceso CRISP-DM y utiliza el apilamiento para construir el modelo de conjunto utilizando tres algoritmos diferentes: K-Vecinos más cercanos (KNN), Máquina de vectores de soporte (SVM) y Árbol de decisión (DT). Esto como la primera parte de su modelo ya que fue usado como salida para las siguientes fases propuestas. El árbol de decisión fue el que presentó mejores resultados, para comprobarlos se utilizan las métricas: Accuracy, AUC, ROC Curve, f1-score, precision, recall, log loss, y specificity.

En el artículo (Zain et al., 2020) los resultados mostraron que el algoritmo KNN, produjo una mejor predicción sin usar principal component analysis (PCA) (medida F = 72,1%), mientras que las otras dos técnicas: NB y REPTree, mejoraron cuando se usaron con PCA (medida F = 76,1% y 72,8%, respectivamente), según la comparación del F-score.

La investigación de (Talha et al., 2020) mostró que la regresión logística se puede aplicar para una predicción temprana precisa y precisa de la detección de cáncer. Se utilizaron los algoritmos KNN, DT y LR evaluados a través de la medida Accuracy. Este trabajo particularmente no solo hizo el estudio de la predicción del cáncer temprano de mama sino también del grado de gravedad de la enfermedad.

En (Alghunaim & Al-Baity, 2019) se usaron datasets provenientes de Wisconsin Breast Cancer y se evaluaron la exactitud, precisión de los algoritmos de árbol de decisión y vecino más cercano evidenciando que el primero obtuvo los mejores resultados. Similar resultado fue encontrado en los trabajos de (Juneja & Rana, 2018) y (Alghunaim & Al-Baity, 2019), donde se evaluaron algoritmos de machine learning. En el mismo se compararon como predictores algoritmos de clasificación y clusterización, obteniendo como resultado que el árbol de decisión con un vector de características tuvo un mejor performance que la red neuronal y el k-vecino cercano usados como contraste.

Por otra parte, en (Franco, L. A. R, 2014) expone que las ideas fundamentales de las ciencias cognitivas estaban inmanentes en los primeros trabajos de McCulloch, Turing, von Neumann, Wiener y Pitts. En este sentido, podemos citar la prueba de la máquina de Turing, según la cual se podría atribuir a la máquina la facultad del pensamiento siempre y cuando un observador no fuese capaz de distinguir su comportamiento del de un ser humano. Por el lado de la cibernética y los

estudios en la inteligencia artificial, es innegable que el boom de silicio exaltó las emociones científicas al punto de augurar el surgir de una inteligencia prácticamente robótica que sería análoga a la del ser humano, e incluso que sería capaz de superarla. Gracias a sus esfuerzos, hoy en día se construyen grandes ejecutores de programas que facilitan la vida ordinaria y la investigación científica.

Citamos como proveedores de servicios en la nube a: Microsoft, Google y Amazon, Rackspace, IBM, Oracle, Salesforce, etc. Todas ellas tienen las características de calidad relevantes, unas de otras, dentro de las plataformas cloud.

En (Álvarez Vañó, J. M, 2018) que la sencilla interfaz de servicios web de Amazon EC2 permite obtener y configurar su capacidad fácilmente. Proporciona un control completo sobre sus recursos informáticos y permite ejecutarse en el entorno informático acreditado de Amazon. Amazon EC2 reduce el tiempo necesario para obtener y arrancar nuevas instancias de servidor en minutos, lo que permite escalar rápidamente la capacidad.

#### *Factores de riesgo del cáncer de mama*

Los siguientes son algunos de los factores de riesgo conocidos del cáncer de mama. Sin embargo, la mayoría de los casos de cáncer de mama no se pueden relacionar con una causa específica (Jerónimo et al., 2017).

- La edad: la probabilidad de contraer cáncer de mama aumenta a medida que las mujeres envejecen. Casi el 80 por ciento de los cánceres de mama se encuentran en mujeres mayores de 50 años (González Ortega et al., 2011).
- Historia personal de cáncer de mama: una mujer que ha tenido cáncer de mama en una mama tiene un mayor riesgo de desarrollar cáncer en la otra mama.
- Antecedentes familiares de cáncer de mama: una mujer tiene un mayor riesgo de cáncer de mama si su madre, hermana o hija tuvo cáncer de mama, especialmente a una edad temprana (antes de los 40). Tener otros parientes con cáncer de mama también puede aumentar el riesgo.

#### *Procedimiento para la evaluación de los algoritmos de clasificación*

##### *Variables de predicción*

El conjunto de datos fue creado por el Dr. William H. Wolberg, médico del Hospital de la Universidad de Wisconsin en Madison, Wisconsin, EE. UU. Para crear el conjunto de datos, el Dr. Wolberg utilizó muestras de fluidos, tomadas de pacientes con masas mamarias sólidas y un programa informático gráfico fácil de usar llamado Xcyt, que es capaz de realizar el análisis de características citológicas basadas en un escaneo digital. El programa utiliza un algoritmo de ajuste de curvas, para calcular diez características de cada una de las células de la muestra, luego calcula el valor medio, el valor extremo y el error estándar de cada característica, devolviendo un vector con valor real de 30.

##### *Características de valor real para cada núcleo celular.*

- radio (media de las distancias desde el centro hasta los puntos del perímetro)
- textura (desviación estándar de los valores de escala de grises)
- perímetro

- zona
- suavidad (variación local en longitudes de radio)
- compacidad ( $\text{perímetro}^2 / \text{área} - 1.0$ )
- concavidad (gravedad de las porciones cóncavas del contorno)
- puntos cóncavos (número de porciones cóncavas del contorno)
- simetría
- dimensión fractal ("aproximación de la línea de costa" - 1)

### *Conjunto de datos usado*

Para la predicción se ha utilizado el dataset con las siguientes características presentadas en la Tabla 1.

Tabla 1. Variables del conjunto de datos

Variable	Tipo	Medida	Descripción
<b>mean_radius</b>	Cuantitativa	Razón	(media de las distancias desde el centro hasta los puntos del perímetro)
<b>mean_texture</b>	Cuantitativa	Razón	(desviación estándar de los valores de escala de grises)
<b>mean_perimeter</b>	Cuantitativa	Razón	Perímetro de la masa muestral
<b>mean_area</b>	Cuantitativa	Razón	Área de la masa muestral
<b>mean_smoothness</b>	Cuantitativa	Razón	(suavidad: variación local en longitudes de radio)
<b>diagnosis</b>	Cualitativa	Clase	Diagnóstico médico

Estos son evaluados a través de métricas tales como: exactitud, precisión, sensibilidad y especificidad (Bilal et al., 2016). Estas métricas derivadas toman sus valores bases de una matriz de confusión. Esta es una herramienta visual que representa el desempeño de un algoritmo de aprendizaje supervisado.

Sus columnas son el número de predicciones de cada clase del conjunto de datos y sus filas representan las instancias de la clase con los valores reales. Esta matriz principalmente responde a la pregunta: ¿qué tipos de aciertos y errores está teniendo nuestro modelo? (Caelen, 2017).

El conjunto de datos está compuesto por seis columnas (cinco contienen datos cuantitativos y 1 cualitativa) y 589 registros. La columna cualitativa almacena el diagnóstico médico avalado por el Hospital y es la columna que se usará como clase para los algoritmos a evaluar.

### *Selección de datos*

Para realizar la selección de los datos con los que se evaluarán los algoritmos de clasificación recurrimos a visualización de sus respectivos histogramas para conocer qué tan balanceados están. En la Figura 1 podemos observar que las variables mean\_texture, mean\_radius y mean\_smoothness presentan una gráfica aproximada a una distribución normal, lo que podría ser usado como un criterio de selección.

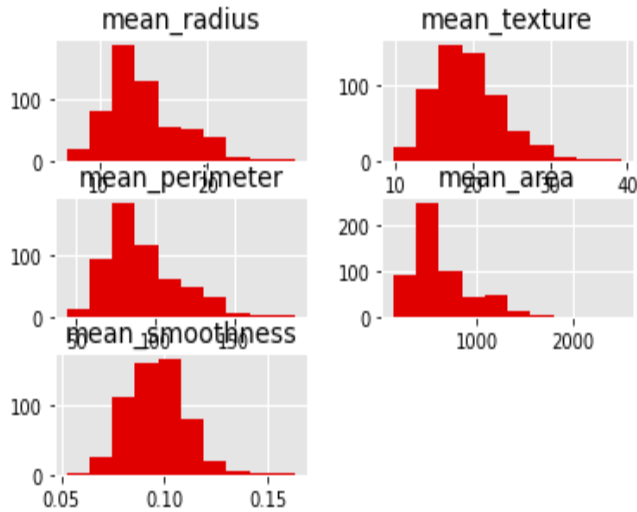


Fig. 1. Histograma de las variables del conjunto de datos.

Para continuar con el proceso de selección se analizarán estas tres variables para a través de los gráficos de dispersión conocer si hay alguno que nos permita conocer la clasificación con respecto a la columna diagnosis que almacena el diagnóstico médico.

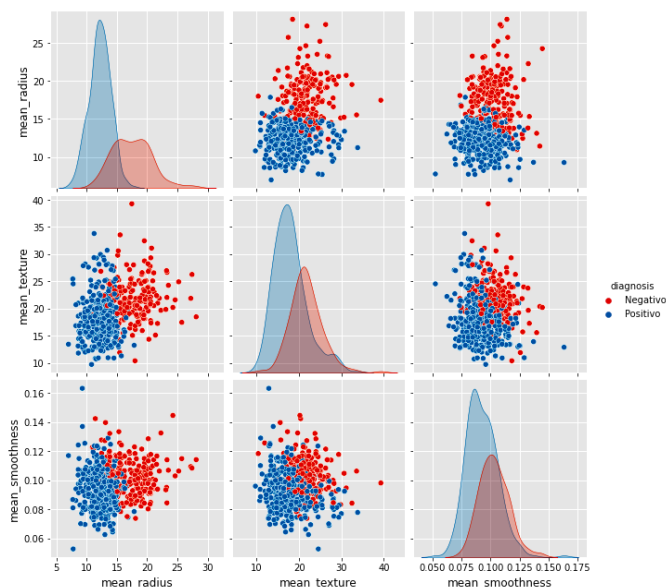


Fig. 2. Histograma de las variables: mean\_radius, mean\_texture y mean\_smoothness.

En la Figura 2 se evidencia que con respecto a las variables diagnosis, las variables mean\_texture y mean\_smoothness presentan delimitaciones marcadas que agrupan los valores del diagnóstico de manera adecuada. A partir de estos dos criterios se seleccionarán las variables: mean\_texture y mean\_smoothness para su uso como entradas de los algoritmos de clasificación a evaluar.

### *Ejecución y evaluación de los algoritmos*

#### *Configuración de los algoritmos*

Los algoritmos de clasificación para aprendizaje no supervisado a evaluar son: red neuronal, k-means, fuzzy c-means, y hierarchical clustering. Con lo que respecta a la red neuronal, su modelo

será de perceptron multicapa y su configuración será a partir de la arquitectura BackPropagation (Contreras Urgilés et al., 2019) que recomienda utilizar más de dos capas ocultas, siendo la siguiente: 100 iteraciones o épocas, 2 capas ocultas y 10 neuronas por cada una. Esta configuración fue la que mejor resultado obtuvimos en cuanto al entrenamiento de la red neuronal, donde se evaluaron desde una capa oculta (Boltzman Machine) y hasta tres capas (BackPropagation) (BODERO et al., 2020).

En cuanto al número de grupos de clusterización será conformado por dos para equiparlos a los posibles resultados de la columna diagnosis.

Para la inicialización del centroide del algoritmo k-means se utilizará las primeras k filas con un máximo de 99 iteraciones.

Con lo que respecta al algoritmo fuzzy c-means se utilizarán las mismas iteraciones del algoritmo k-means y un valor de dos para el difusor del algoritmo. Esta configuración permitirá acercar los valores de clusterización con límites definidos.

Finalmente, el algoritmo hierarchical clustering utilizará una vinculación completa y Manhattan como función de distancia. Esta fue seleccionada porque presentó grupos con límites más marcados que la función de distancia Euclideana.

#### *Ejecución de los algoritmos*

En las Tablas 2, 3, 4 y 5 se presentan las matrices de confusión de cada algoritmo ejecutado.

Tabla 2. Matriz de confusión de la red neuronal.

<i>diagnosis</i>	<i>Negativo</i>	<i>Positivo</i>
<i>Negativo</i>	56 (VP)	16 (FP)
<i>Positivo</i>	7 (FN)	92 (VN)

Tabla 3. Matriz de confusión de k-means.

<i>diagnosis</i>	<i>Negativo</i>	<i>Positivo</i>
<i>Negativo</i>	71 (VP)	141 (FP)
<i>Positivo</i>	276 (FN)	81 (VN)

Tabla 4. Matriz de confusión de fuzzy c-means.

<i>diagnosis</i>	<i>Negativo</i>	<i>Positivo</i>
<i>Negativo</i>	67 (VP)	145 (FP)
<i>Positivo</i>	272 (FN)	85 (VN)

Tabla 5. Matriz de confusión de hierarchical clustering.

<i>diagnosis</i>	<i>Negativo</i>	<i>Positivo</i>
<i>Negativo</i>	64 (VP)	148 (FP)
<i>Positivo</i>	268 (FN)	89 (VN)

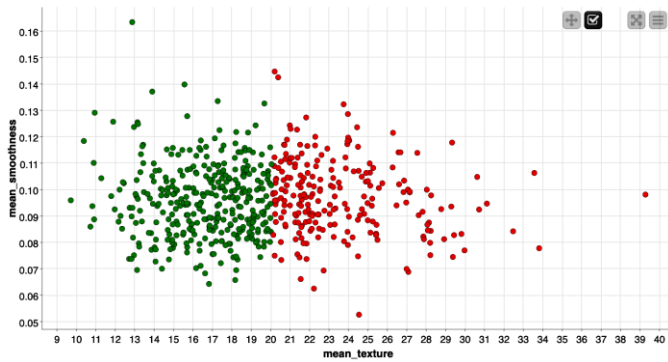


Fig. 3. Diagrama de puntos del algoritmo k-means.

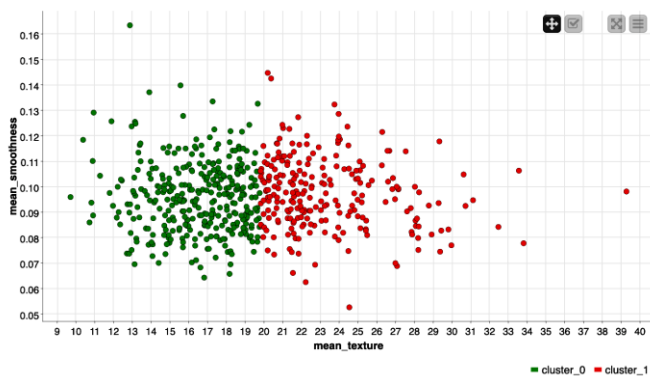


Fig. 4. Diagrama de puntos del algoritmo hierarchical clustering.

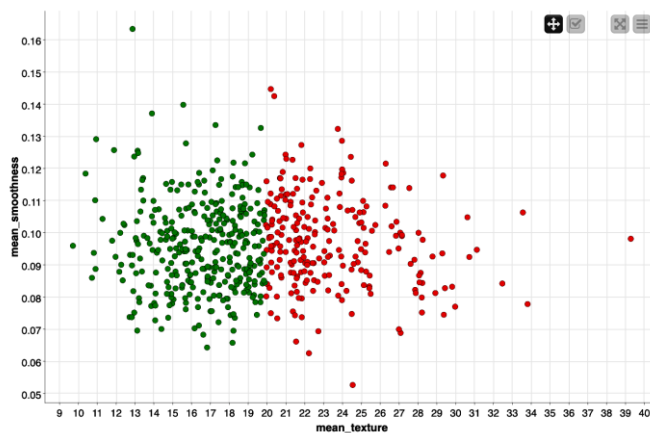


Fig. 5. Diagrama de puntos del algoritmo fuzzy c-means.

### *Análisis de los resultados*

Se puede visualizar en las Figuras 3 a 5 cómo los algoritmos de clasificación: k-means, fuzzy c-means y hierarchical clustering crearon los grupos a partir de las configuraciones dadas. Las imágenes cuentan con coherencia entre los algoritmos usados y su determinación de evaluación se deberá realizar tomando los resultados de sus métricas en la Tabla 6.

Con base en los resultados de cada uno de los algoritmos y tomando como métrica inicial su exactitud, la red neuronal presenta una diferencia porcentual significativa con los valores



obtenidos de los restantes tres algoritmos. Esto nos indica que el primero clasificó de mejor forma (más aciertos) a los elementos en comparación con su columna clase.

Si evaluamos esta medida conjuntamente con la exactitud se obtiene nuevamente una diferencia favorable para la red neuronal sobre los otros algoritmos evaluados.

Tabla 6. Valores de evaluación obtenidos en los algoritmos

	<i>Red neuronal</i>	<i>K-means</i>	<i>Fuzzy c-means</i>	<i>Hierarchical clustering</i>
<i>Exactitud</i>	86,54%	71,14%	67,14%	64,15%
<i>Precisión</i>	77,77%	0,33%	31%	30%
<i>Sensibilidad</i>	88,88%	20,46%	19,76%	18%
<i>Especificidad</i>	85,18%	36,48%	36,95%	37%

### *Determinación de la variable más importante*

Se ha utilizado un sistema basado en inteligencia artificial capaz de responder a preguntas formuladas en lenguaje natural el cual ha permitido predecir que la variable más importante para determinar el cáncer de mama es *mean\_perimeter*, conforme al reporte que se muestra a continuación.

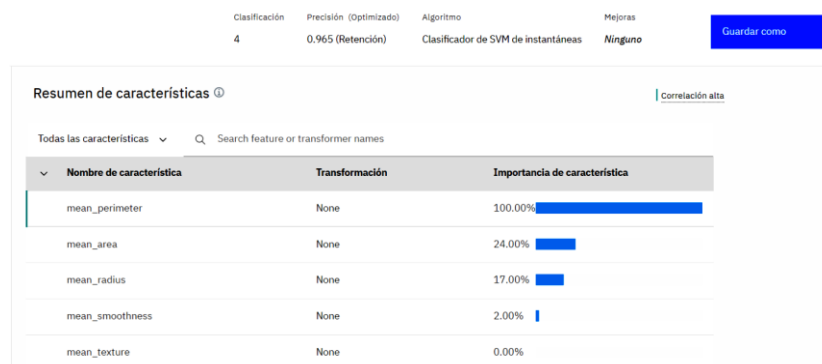


Fig. 6. Resumen de características de predicción.

### *Aplicación de sistemas cognitivos*

Los sistemas cognitivos son sistemas basados en tecnologías de Inteligencia Artificial que los dotan de una capacidad de aprendizaje y entendimiento simulando el comportamiento del cerebro humano.

IBM define la computación cognitiva como un sistema avanzado que aprende a escala, razona con un propósito e interactúa con los humanos de forma natural.

Para poder probar el modelo seleccionado, se creó una aplicación demostrativa usando los servicios dispuestos en la nube referente a los sistemas cognitivos. El proveedor usado será Microsoft Azure.

Desde su cuenta estudiantil que permite hacer uso de las herramientas del esquema a proponer.

Los servicios a utilizar serán: machine learning services, azure cosmos db, machine learning web services, cognitive services y api connections.

Estos se agruparán en dos secciones con objetivos definidos. La primera se denominará entrenamiento y estará destinada a la lectura del archivo, la aplicación del algoritmo de aprendizaje y la escritura del modelo entrenado dentro de una base de datos. La segunda será llamada deploy y tendrá como objetivo disponibilizar el modelo entrenado en un servicio web con acceso a cognitive services y api connections.

Para ejemplificar el acceso al servicio web se construyó una single page app con despliegue en la nube. En la Figura 9 se representa la arquitectura planteada y el flujo que esta sigue hasta su fase de despliegue.

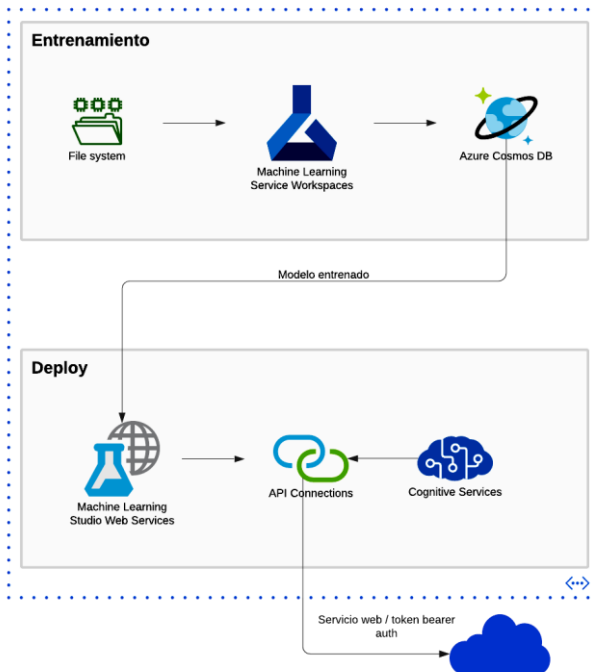


Fig. 8. Arquitectura propuesta de la aplicación de demostración.

### *Demostración de la solución propuesta*

Una vez ejecutado la fase de entrenamiento de la arquitectura propuesta, el modelo entrenado es almacenado en la base de datos Azure Cosmos. Para el entrenamiento del modelo se implementó la red neuronal de dos clases con la configuración propuesta en la evaluación.

Este modelo entrenado es almacenado en la base de datos a la cual accede el servicio Azure Machine Learning para construir el servicio web. Este servicio web construido encapsula el proyecto de implementación del algoritmo en una función para poder implementar el resto de la arquitectura REST con un verbo de ejecución disponibles: POST. El formato de intercambio de datos usados es JSON tanto para la solicitud como para la respuesta del servicio.

Demo for Breast Cancer Prediction [Predictive Exp.]

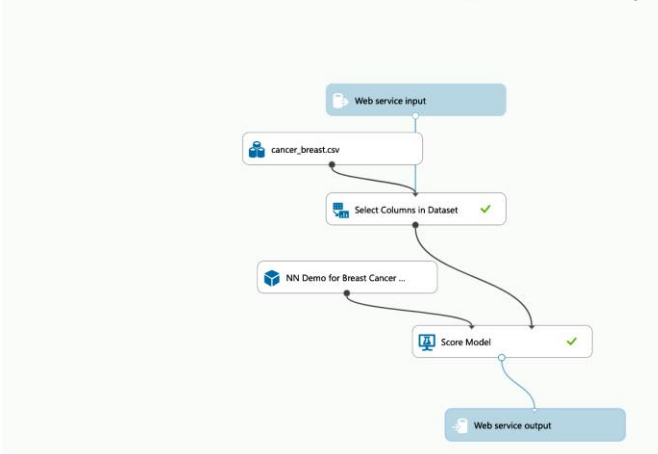


Fig. 7. Esquema del servicio web de predicción.

Este servicio desplegado puede ser accedido por cualquier aplicación que sea capaz de llamar servicios REST API y enviar una solicitud en formato JSON con los datos de entradas necesarios para el algoritmo.

**Formatted JSON Data**

```
{
  "Inputs": {
    "input1": [
      {
        "mean_radius": 17.99,
        "mean_texture": 10.38,
        "mean_perimeter": 122.8,
        "mean_area": 1001,
        "mean_smoothness": 0.1184
      }
    ]
  }
}
```

Fig. 9. Esquema del servicio web de predicción.

La arquitectura propuesta también provee de una herramienta de ejecución que permite evaluar la solución construida. Esta puede ser accedida de manera pública a través de un navegador web.

input1

mean_radius	17.99
mean_texture	10.38
mean_perimeter	122.8
mean_area	1001
mean_smoothness	0.1184
diagnosis	Negativo

Test Request-Response

output1

mean_radius	17.99
mean_texture	10.38
mean_perimeter	122.8
mean_area	1001
mean_smoothness	0.1184
Scored Labels	Negativo
Scored Probabilities	0.0317362397909164

Fig. 10. Esquema del servicio web de predicción.

### Arquitectura futura

Como parte de mejora del presente modelo evaluado se propone anexar a la arquitectura un módulo que además de permitir importar datos desde un archivo, se posibilite el agregar datos desde otra fuente. Esta a partir de la imagen de la biopsia y utilizando los servicios cognitivos de reconocimiento de imágenes podría brindar datos de sus dimensiones y textura.

El servicio cognitivo de visión computarizada permite reconocer caracteres OCR, análisis de imágenes y análisis espacial. Con el uso del sistema de detección de caracteres se puede detectar el texto de las dimensiones que son provistas a través de las placas de análisis que se la hacen a la biopsia.

También se puede determinar datos del mismo a través del reconocimiento espacial y conocer sus dimensiones con respecto a una marca guía. Con este servicio se podrían determinar las medidas: mean\_texture, mean\_radius, mean\_perimeter y mean\_area. Para obtener la medida de la variable mean\_smoothness se propone incorporar tecnología proveniente de internet de las cosas a través del uso de un sensor sobre una placa de arduino.

El procedimiento para obtener la medición de la suavidad de la muestra se basa en la proposición del uso de sensores ultrasónicos que permiten evaluar la reflectancia y la dispersión ultrasónica en muestras de metal. Se utilizó la amplitud del eco de los pulsos cortos incidentes de ultrasonido en el rango de frecuencia de 1 a 30 MHz, a modo de comparador, para medir las rugosidades relativas con un enfoque promediado de área definido por el tamaño del punto del haz ultrasónico.

La variación de las longitudes de onda ultrasónicas son las que permiten conocer la suavidad. Si

se encuentra en una variación de 50 a 300  $\mu\text{m}$  se lo puede relacionar una superficie suave, si supera ese rango se considera una superficie rugosa.

En la Figura 11 se esquematiza la arquitectura futura propuesta.

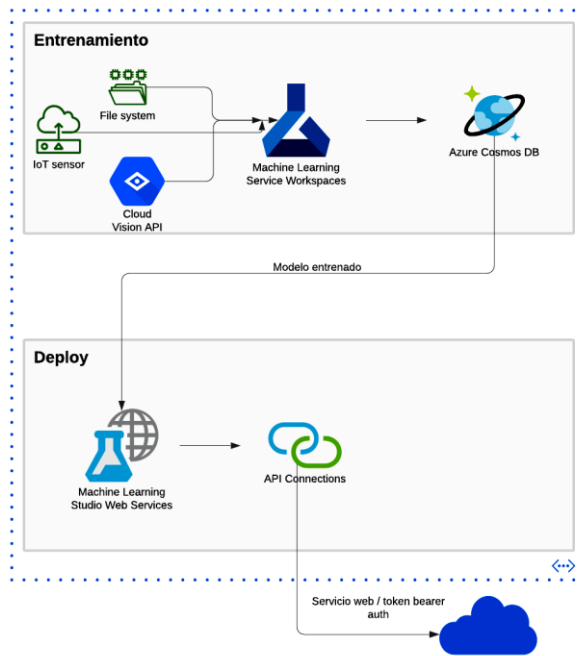


Fig. 11. Arquitectura futura propuesta.

## CONCLUSIONES

Se ha demostrado con los valores de las métricas bases y derivadas usadas para la evaluación de algoritmos de clasificación, que el algoritmo red neuronal presenta mejores resultados de predicción para la configuración del experimento seleccionado. Esto nos puede brindar una ayuda para la selección de un algoritmo de clasificación si tenemos como opciones al algoritmo k-means, fuzzy c-means y hierarchical clustering conjuntamente con una red neuronal.

La red neuronal que mejor resultado se encontró fue con la configuración backpropagation, utilizando dos capas ocultas, 100 iteraciones y diez neuronas por cada capa. Se evaluaron otras arquitecturas como la Boltzman Machine, pero no se encontraron buenos resultados.

Se ha demostrado que la variable más importante para la predicción de cáncer de mama es mean\_perimeter, con una certeza del 100% conforme a la predicción realizada por la plataforma de inteligencia artificial IBM WATSON.

La lectura cognitiva de imágenes, gracias al uso de la inteligencia artificial, para la captura de datos no solo nos permite identificar emociones, sino que también objetos y rostros en videos.

## REFERENCIAS BIBLIOGRÁFICAS

- Alghunaim, S., & Al-Baity, H. H. (2019). On the scalability of machine-learning algorithms for breast cancer prediction in big data context. *IEEE Access*, 7, 91535-91546.
- Álvarez Vañó, J. M. (2018). *Modelo Comparativo de Plataformas Cloud y Evaluación de Microsoft Azure, Google App Engine y AmazonEC2* (Doctoral dissertation, Universitat Politècnica de València).
- Bilal, M., Israr, H., Shahid, M., & Khan, A. (2016). Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques. *Journal of King Saud University-Computer and Information Sciences*, 28(3), 330-344.
- BODERO, E. M., LOPEZ, M. P., CONGACHA, A. E., CAJAMARCA, E. E., & MORALES, C. H. (2020). Google Colaboratory como alternativa para el procesamiento de una red neuronal convolucional. *Revista ESPACIOS*, 41(07).
- Caelen, O. (2017). A Bayesian interpretation of the confusion matrix. *Annals of Mathematics and Artificial Intelligence*, 81(3), 429-450.
- Contreras Urgilés, W., Maldonado Ortega, J., & León Japa, R. (2019). Aplicación de una red neuronal feed-forward backpropagation para el diagnóstico de fallas mecánicas en motores de encendido provocado. *Ingenius. Revista de Ciencia y Tecnología*, 21, 32-40.
- Franco, L. A. R. (2014). Psicología cognitiva e inteligencia artificial: mitos y verdades. *Avances en Psicología*, 22(1), 21-27.
- González Ortega, J. M., Morales Wong, M. M., López Cuevas, Z., & Díaz Valdéz, M. (2011). Factores pronósticos del cáncer de mama. *Revista Cubana de Cirugía*, 50(1), 130-138.
- Jerônimo, A. F. de A., Freitas, Â. G. Q., & Weller, M. (2017). Risk factors of breast cancer and knowledge about the disease: An integrative revision of Latin American studies. *Ciencia & saude coletiva*, 22, 135-149.
- Juneja, K., & Rana, C. (2018). An improved weighted decision tree approach for breast cancer prediction. *International Journal of Information Technology*, 1-8.
- Nanglia, S., Ahmad, M., Khan, F. A., & Jhanjhi, N. (2022). An enhanced Predictive heterogeneous ensemble model for breast cancer prediction. *Biomedical Signal Processing and Control*, 72, 103279.
- Talha, A. K., Kushsairy, A. K., Shahzad, N., Muhammad, A., Zeeshan, S., & Mazliham, M. (2020). Proficiency Assessment of Machine Learning Classifiers: An Implementation for the Prognosis of Breast Tumor and Heart Disease Classification.
- Zain, Z. M., Alshenaifi, M., Aljaloud, A., Albednah, T., Alghanim, R., Alqifari, A., & Alqahtani, A. (2020). Predicting breast cancer recurrence using principal component analysis as feature extraction: An unbiased comparative analysis. *International Journal of Advances in Intelligent Informatics*, 6(3), 313-327.
- Zhang, Y., Lu, S., Zhou, X., Yang, M., Wu, L., Liu, B., Phillips, P., & Wang, S. (2016). Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: Decision tree, k-nearest neighbors, and support vector machine. *Simulation*, 92(9), 861-871.