Neurotropic Artificial Neural Network for Breast Cancer Detection

Darwin Patiño-Pérez,Ph.D[1]; Luis Armijos-Valarezo,MSc[2]; Luis Chóez-Acosta,MSc[2];
Zynnia Reyes-Sánchez,MSc[3]; Celia Munive-Mora,BS[4]; Zackery Patino-Munive,Std[5]
[1]Universidad de Guayaquil, Facultad de Ciencias Matemáticas y Física, Grupo de Investigación de Inteligencia Artificial, Ecuador, darwin.patinop@ug.edu.ec,
[2]Universidad de Guayaquil, Facultad de Ciencias Matemáticas y Física, Ecuador, luis.armijosv@ug.edu.ec, luis.choeza@ug.edu.ec,
,[3]Universidad de Guayaquil, Facultad de Ciencias Médicas, Ecuador, zynnia.reyessan@ug.edu.ec,
[4]St Luke's University Hospital Network, PA, United States,celia.munive@sluhn.org,
[5]Kutztown University, Department of Psychology, PA, United State, zpati840@live.kutztown.edu

Abstract– ***The implementation of deep learning models in clinical diagnosis demands high precision and reliable calibration. Inspired by neuronal plasticity, neurotropic models promise to overcome the limitations of static architecture. This study compared a neurotropic network—featuring dynamic growth and pruning, activity-dependent plasticity, and simulated neurotrophic factors—against a traditional network; both were trained on the Wisconsin breast cancer dataset (n=569) using identical stratified splits and early stopping. Contrary to expectations, the traditional model outperformed the neurotropic one in accuracy (96.49% vs. 95.61%), sensitivity (98.61% vs. 95.83%), F1-score (0.9726 vs. 0.9650), AUC-ROC (0.9914 vs. 0.9554), and calibration (Brier Score: 0.0261 vs. 0.0439). The neurotropic model reduced false positives (from 3 to 2) but increased false negatives (from 1 to 3)—a 200% increase in the clinically most critical error type. It showed a moderate advantage only in precision (97.18% vs. 95.95%). The neurotropic architecture demonstrated adaptive specialization, though it lacked a competitive advantage in this configuration. By prioritizing precision, it compromised sensitivity and calibration. This trade-off suggests that, while offering theoretical advantages, these models require adjustments to avoid penalizing the detection of malignant cases—a priority in oncological screening. Future research should explore configurations that preserve high precision without sacrificing sensitivity.***

Keywords— Breast Cancer, Neurotropic Model, Neural Networks, Calibration, Brier Score.

Red Neuronal Artificial Neurotrópica para la Detección de Cáncer de Mama

Darwin Patiño-Pérez,Ph.D[1]; Luis Armijos-Valarezo,MSc[2]; Luis Chóez-Acosta,MSc[2]; Zynnia Reyes-Sánchez,MSc[3]; Celia Munive-Mora,BS[4]; Zackery Patino-Munive,Std[5]
[1]Universidad de Guayaquil, Facultad de Ciencias Matemáticas y Física, Grupo de Investigación de Inteligencia Artificial, Ecuador, darwin.patinop@ug.edu.ec,
[2]Universidad de Guayaquil, Facultad de Ciencias Matemáticas y Física, Ecuador, luis.armijosv@ug.edu.ec, luis.choeza@ug.edu.ec,
,[3]Universidad de Guayaquil, Facultad de Ciencias Médicas, Ecuador, zynnia.reyessan@ug.edu.ec,
[4]St Luke's University Hospital Network, PA, United States,celia.munive@sluhn.org,
[5]Kutztown University, Department of Psychology, PA, United State, zpati840@live.kutztown.edu

Resumen– ***La implementación de modelos profundos en diagnóstico clínico exige alta precisión y calibración fiable. Inspirados en la plasticidad neuronal, los modelos neurotrópicos prometen superar limitaciones de arquitecturas estáticas. Este estudio comparó una red neurotrópica —con crecimiento y poda dinámicos, plasticidad dependiente de actividad y factores neurotróficos simulados— frente a una red tradicional, ambas entrenadas sobre el dataset Wisconsin de cáncer de mama (n=569) con divisiones estratificadas idénticas y detención temprana. Contra lo previsto, el modelo tradicional superó al neurotrópico en exactitud (96.49 % vs 95.61 %), sensibilidad (98.61 % vs 95.83 %), F1-score (0.9726 vs 0.9650), AUC-ROC (0.9914 vs 0.9554) y calibración (Brier Score: 0.0261 vs 0.0439). El neurotrópico redujo falsos positivos (de 3 a 2), pero incrementó falsos negativos (de 1 a 3): un aumento del 200 % en el error clínicamente más crítico. Solo en precisión mostró una ventaja moderada (97.18 % vs 95.95 %).La arquitectura neurotrópica evidenció especialización adaptativa, aunque sin ventaja competitiva en esta configuración. Al priorizar la precisión, comprometió sensibilidad y calibración. Este equilibrio sugiere que, si bien ofrecen ventajas teóricas, requieren ajustes para no penalizar la detección de casos malignos —prioridad en cribado oncológico. Futuras investigaciones deberán explorar configuraciones que preserven su alta precisión sin sacrificar la sensibilidad.***

Palabras clave- Cáncer de Mama, Modelo Neurotrópico, Red Neuronal, Calibración, Brier Score.

I Introducción

A. Contexto Clínico Global del Cáncer de Mama

El cáncer de mama según Fig. 1, constituye uno de los problemas de salud pública más significativos a nivel mundial, representando aproximadamente el 25% de todos los cánceres diagnosticados en mujeres. Según datos de GLOBOCAN 2020, se estiman 2.3 millones de nuevos casos anuales y 685,000 muertes relacionadas, posicionándolo como la principal causa de mortalidad por cáncer en la población femenina a nivel global [1]. La incidencia varía geográficamente, con tasas más altas en países desarrollados, aunque la mortalidad es desproporcionadamente mayor en naciones en desarrollo debido a diagnósticos tardíos y acceso limitado a tratamientos adecuados.

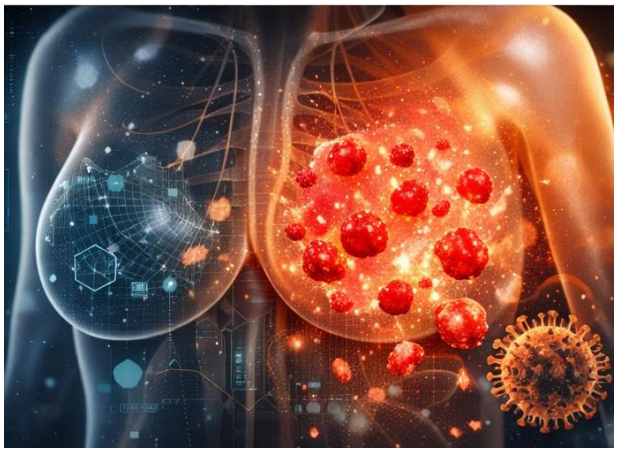


Fig. 1 Cáncer de Mama – Células Malignas

En América Latina, el cáncer de mama presenta características epidemiológicas particulares, con una edad media de diagnóstico más temprana y formas clínicas más agresivas en comparación con poblaciones de Europa y Norteamérica. Esta realidad subraya la necesidad crítica de mejorar tanto los métodos de detección temprana como las herramientas de diagnóstico asistido disponibles en la región.

B. Retos en el Diagnóstico Temprano

La detección temprana del cáncer de mama sigue siendo un desafío complejo debido a múltiples factores:

1) *Heterogeneidad tumoral:* Los cánceres de mama no constituyen una enfermedad única, sino un grupo heterogéneo de neoplasias con diferentes perfiles moleculares, comportamientos biológicos y respuestas al tratamiento.

2) *Limitaciones de los métodos actuales:* La mamografía, aunque efectiva, presenta limitaciones en densidad mamaria alta, con sensibilidad que puede reducirse hasta 50% en mamas extremadamente densas.

3) *Variabilidad en interpretación:* La evaluación de imágenes mamográficas muestra variabilidad significativa, incluso entre radiólogos experimentados.

4) *Sobrediagnóstico y sobretratamiento:* La dificultad para distinguir lesiones indolentes de aquellas potencialmente mortales puede llevar a intervenciones innecesarias.

C. Evolución de la Inteligencia Artificial en Diagnóstico Mamario

El desarrollo de sistemas de inteligencia artificial (IA) para asistencia diagnóstica en cáncer de mama ha evolucionado considerablemente en la última década[2]. Los primeros sistemas basados en reglas simples han dado paso a algoritmos de aprendizaje automático tradicional y, más recientemente, a modelos de aprendizaje profundo[3].

Los sistemas CAD (*Computer-Aided Diagnosis*) tradicionales mostraron mejoras modestas en la detección, pero con limitaciones importantes en tasas de falsos positivos. La revolución del *deep learning*, particularmente con arquitecturas convolucionales, ha permitido análisis más sofisticados de imágenes médicas, alcanzando en algunos estudios performance comparable a radiólogos expertos [4].

D. Limitaciones de Arquitecturas Tradicionales

A pesar de los avances, las redes neuronales tradicionales presentan desafíos significativos en aplicaciones médicas:

1) *Calibración deficiente:* Tendencia a producir probabilidades mal calibradas, siendo sobreconfiadas especialmente en predicciones incorrectas [5].

2) *Falta de adaptabilidad:* Arquitecturas estáticas que no se ajustan dinámicamente a características específicas del paciente o patrones emergentes.

3) *Sensibilidad a desbalances:* Performance disminuida en clases minoritarias, problema frecuente en datos médicos.

4) *Interpretabilidad limitada:* Comportamiento de "caja negra" que dificulta la validación clínica y la confianza del profesional.

E. Fundamentos de los Modelos Neurotrópicos

Inspirados en principios de neuro plasticidad biológica, los modelos neurotrópicos representan un paradigma emergente en aprendizaje automático [6]. Estos sistemas incorporan mecanismos adaptativos que simulan procesos fundamentales del sistema nervioso:

1) *Potenciación a largo plazo (LTP):* Fortalecimiento sináptico basado en actividad coordinada.

2) *Depresión a largo plazo (LTD):* Debilitamiento de conexiones poco utilizadas.

3) *Crecimiento axonal guiado:* Especialización estructural dirigida por factores tróficos.

4) *Apoptosis neuronal:* Eliminación programada de unidades no esenciales.

Esta capacidad de autoreorganización estructural permite a los modelos neurotrópicos adaptarse dinámicamente a patrones complejos y heterogéneos, característica particularmente valiosa para datos médicos donde la variabilidad interindividual es significativa.

Como contraparte, la arquitectura tradicional de referencia en este estudio sigue un diseño estático y ampliamente validado en clasificación binaria médica. Consiste en una capa de entrada de 30 neuronas —una por característica—, tres capas ocultas densas de 64, 32 y 16 neuronas con activación ReLU, regularización L2 ($\lambda = 0.001$), *batch normalization* y *dropout* (0.3, 0.3 y 0.2, respectivamente), y una capa de salida sigmoide. Su entrenamiento emplea el optimizador Adam con tasa de aprendizaje inicial de 0.001, función de pérdida *Binary Crossentropy* y *early stopping* sobre AUC de validación. A diferencia del modelo neurotrópico, esta arquitectura no modifica su topología durante el entrenamiento: su capacidad de adaptación se limita al ajuste de pesos mediante retropropagación. La comparación entre ambos enfoques permite evaluar si los mecanismos de plasticidad dinámica ofrecen ventajas tangibles frente a una línea de base estable y bien documentada.

F. Justificación y Objetivos del Estudio

Este estudio surge de la necesidad de desarrollar sistemas de IA no solo con alta precisión estadística, sino con características que los hagan verdaderamente útiles en entornos clínicos reales. El objetivo principal es evaluar si los principios neurotrópicos ofrecen ventajas tangibles sobre arquitecturas tradicionales en el contexto específico de detección de cáncer de mama, con énfasis en métricas clínicamente relevantes como sensibilidad y calibración probabilística.

La hipótesis central propone que los mecanismos de plasticidad adaptativa permitirán al modelo neurotrópico: (1) Mejorar la detección de casos difíciles o atípicos, (2) Producir estimaciones probabilísticas más confiables, y (3) Adaptarse más eficientemente a las características específicas del dataset.

II. MATERIALES Y MÉTODOS

A. Dataset : Características y Consideraciones

1) Origen y Características del Dataset de Wisconsin

El estudio utilizó el dataset de Cáncer de Mama de Wisconsin (Diagnóstico), compilado originalmente por el Dr. William H. Wolberg de la Universidad de Wisconsin-Madison [7]. Este dataset según Tabla I. se ha convertido en un *benchmark* estándar en la literatura de *machine learning* aplicado a medicina, con más de 2000 citaciones en investigaciones relacionadas.

TABLA I
EL DATASET

mean radius	mean texture	mean perimeter	mean area	...	worst symmetry	worst fractal dimension
17.99	10.38	122.80	1001.0	...	0.4601	0.11890
20.57	17.77	132.90	1326.0	...	0.2750	0.08902
19.69	21.25	130.00	1203.0	...	0.3613	0.08758
11.42	20.38	77.58	386.1	...	0.6638	0.17300
20.29	14.34	135.10	1297.0	...	0.2364	0.07678

El dataset contiene 569 casos obtenidos mediante aspiración con aguja fina (FNA) de masas mamarias. Para cada caso, se extrajeron células del núcleo de las imágenes de FNA digitalizadas y se calcularon 30 características cuantitativas que describen las características de los núcleos celulares. Estas características incluyen:

Características primarias:

a. *Medidas geométricas:*
   - Radio (media de distancias desde centro a puntos en el perímetro)
   - Textura (desviación estándar de valores de escala de grises)
   - Perímetro
   - Área
   - Suavidad (variación local en longitudes de radio)

b. *Medidas de forma:*
   - Compacidad: (perímetro²/área - 1.0)
   - Concavidad (severidad de porciones cóncavas del contorno)
   - Puntos cóncavos (número de porciones cóncavas del contorno)
   - Simetría
   - Dimensión fractal ("aproximación de línea costera" - 1)

Características derivadas:

Cada una de las 10 características anteriores se presenta en tres formas:

a) Media (mean)
b) Error estándar (standard error)
c) Peor o mayor (worst/largest) - el valor más grande de cada característica

Esto resulta en 30 características totales (10 características × 3 medidas).

2) Distribución y Consideraciones Éticas

La distribución de clases es desbalanceada pero representativa de la prevalencia clínica:

- Casos benignos (B): 357 (62.7%)
- Casos malignos (M): 212 (37.3%)

Este desbalance refleja la realidad clínica donde los procedimientos de FNA incluyen tanto lesiones sospechosas que resultan benignas como aquellas confirmadas malignas. Todos los datos están completamente anonimizados y de acceso público, cumpliendo con estándares éticos para investigación con datos médicos.

B. Preprocesamiento Detallado

1) *Protocolo de Limpieza y Validación.* Antes del análisis, se implementó un protocolo exhaustivo de calidad de datos:

a) *Verificación de valores ausentes:* El dataset original no contiene valores nulos.
b) *Detección de outliers:* Se aplicó el método del rango intercuartílico (IQR) para identificar valores extremos en cada característica.
c) *Análisis de correlación:* Se calcularon matrices de correlación para identificar características redundantes.
d) *Normalización:* Todas las características fueron estandarizadas usando StandardScaler (1) de scikit-learn.

$$X_{\text{estandarizado}} = \frac{X - \mu}{\sigma} \quad (1)$$

donde μ es la media y σ la desviación estándar de cada característica en el conjunto de entrenamiento.

2) *Estrategia de División de Datos.* Para garantizar una evaluación robusta y evitar sesgos, se aplicó una división estratificada que mantiene la proporción de clases en todos los conjuntos: entrenamiento 64 % (364 casos) para ajuste de parámetros; validación 16 % (91 casos) para ajuste de hiperparámetros y early stopping; y prueba 20 % (114 casos) para evaluación final, utilizado una sola vez. Esta división optimiza el uso de datos limitados mientras mantiene conjuntos de validación y prueba suficientemente grandes para una evaluación estadística confiable.

C. Arquitectura de la Red Neuronal Tradicional (Detallada)

1) *Diseño Arquitectónico.* La red neuronal tradicional implementada sigue principios establecidos en literatura para clasificación binaria médica[8]:

Capa de entrada: 30 neuronas (una por característica)

Capas ocultas:

Capa 1: 64 neuronas con activación ReLU
- Regularización L2: $\lambda = 0.001$
- Batch Normalization: Normalización por lotes
- Dropout: tasa = 0.3

Capa 2: 32 neuronas con activación ReLU
- Regularización L2: $\lambda = 0.001$
- Batch Normalization
- Dropout: tasa = 0.3

Capa 3: 16 neuronas con activación ReLU
- Dropout: tasa = 0.2

Capa de salida: 1 neurona con activación sigmoide

2) *Hiperparámetros y Configuración.* Optimizador: Adam con los siguientes parámetros:
- Tasa de aprendizaje inicial: 0.001
- β_1: 0.9, β_2: 0.999
- ε: 1e-7

Función de pérdida: Binar y Crossentropy

Métricas de monitoreo:
- Accuracy
- AUC-ROC
- Precisión (Precision)
- Sensibilidad (Recall)

3) *Estrategia de Regularización.* Para prevenir *overfitting* en el modelo tradicional:

a) *Regularización L2 (weight decay):* Penalización de pesos grandes.
b) *Dropout:* Desactivación aleatoria de neuronas durante entrenamiento.
c) *Early Stopping:* Detención cuando AUC de validación no mejora por 20 épocas
d) *Reducción de tasa de aprendizaje:* Reducción por factor 0.5 cuando pérdida de validación se estanca.

D. *Arquitectura del Modelo Neurotrópico*

La arquitectura del modelo neurotrópico se fundamenta en cuatro principios neurobiológicos que le confieren capacidades de adaptación dinámica Fig 2. Primero, implementa plasticidad dependiente de actividad mediante el principio de Hebb[9], donde la fuerza sináptica Δw_ij entre dos neuronas i y j se modifica proporcionalmente al producto de sus actividades (2) es decir, "neuronas que se activan juntas, se conectan juntos". Segundo, incorpora factores neurotróficos simulados (inspirados en NGF/BDNF biológicos) que regulan la supervivencia, diferenciación y crecimiento neuronal en respuesta a patrones específicos de actividad.

$$\Delta w_{ij} \propto a_i \times a_j \tag{2}$$

Tercero, integra un mecanismo de podado sináptico que elimina selectivamente conexiones débiles o redundantes, optimizando la eficiencia de la red[10]. Cuarto, promueve la especialización funcional, permitiendo que las neuronas se diferencien en subtipos especializados para procesar distintos tipos de información.

```
class NeurotropicNeuron(layers.Layer):
                    Parámetros Neurotrópicos
    def __init__(self, units, growth_rate=0.05,
                 runing_rate=0.03,
                 survival_threshold=0.1):
Tasa de Crecimiento   self.growth_rate = growth_rate
Tasa de Podado        self.pruning_rate = pruning_rate
Umbral de Supervivencia  self.survival_threshold = survival_threshold
  Lógica de Crecimiento / Pruning
    def apply_neurotropic_dynamics(self):
      if actividad > 0.7:
        self.neurotrophic_factors *= (1 + self.growth_rate)
      elif actividad < self.survival_threshold:
        self.neurotrophic_factors *= (1 - self.pruning_rate)

class NeuroplasticConnection(layers.Layer):
                    Párametros de Plasticidad
    def __init__(self, strengthening_rate=0.05,
                 weakening_rate=0.03):
LTP  self.strengthening_rate = strengthening_rate
LTD  self.weakening_rate = weakening_rate
```

Fig. 2 Implementación Técnica Detallada

Estos principios combinados permiten que el modelo no solo aprenda estáticamente, sino que se reestructure dinámicamente en respuesta a la experiencia, emulando procesos de desarrollo y adaptación observados en sistemas nerviosos biológicos.

1) Arquitectura Completa del Modelo

Estructura en capas:

a. Capa de entrada plástica: NeuroplasticConnection (fortalecimiento/debilitamiento basado en entrada)[11].
b. Capa neurotrópica 1: 64 neuronas con mecanismos de crecimiento/pruning.
c. Conexión plástica 1: Ajuste de conexiones inter-capas.
d. Capa neurotrópica 2: 32 neuronas con especialización adaptativa.
e. Capa neurotrópica 3: 16 neuronas con podado selectivo.
f. Capa de salida: 1 neurona sigmoide.

Parámetros específicos por capa:

- Capa 1: growth_rate=0.05, pruning_rate=0.03
- Capa 2: growth_rate=0.04, pruning_rate=0.02
- Capa 3: growth_rate=0.03, pruning_rate=0.01

2) *Mecanismos de Adaptación Dinámica.* Los mecanismos de adaptación dinámica implementan procesos de plasticidad inspirados en neurociencia para optimizar continuamente la red neuronal[12]. El crecimiento neuronal selectivo premia a las neuronas que participan consistentemente en predicciones correctas, incrementando sus factores neurotrópicos y amplificando su influencia en la red. Complementariamente, el podado por desuso elimina progresivamente neuronas con actividad persistentemente baja (menos del 10% del máximo), reduciendo sus factores neurotrópicos[13] hasta su remoción completa. Paralelamente, la potenciación sináptica (LTP) fortalece exponencialmente las conexiones entre neuronas que se co-activan frecuentemente, consolidando patrones de comunicación efectivos. Finalmente, la depresión sináptica (LTD) debilita gradualmente las conexiones raramente utilizadas, eliminando rutas ineficientes. Este sistema dual —que actúa tanto a nivel neuronal como sináptico— permite a la red reconfigurar su arquitectura dinámicamente, conservando estructuras útiles mientras descarta componentes redundantes o inactivos, optimizando así su eficiencia computacional y capacidad adaptativa.

E. *Protocolo de Entrenamiento Comparativo.*

Para realizar el comparativo entre ambos modelos la configuración de hardware debe ser común, en lo relacionado a CPU, GPU y memoria RAM; en tema de software se usar Python, TensorFlow y CUDA; los hiperpárametros de entrenamiento serán de 100 épocas máximas, lote de tamaño 32, el early stopping será de 20 épocas. Los *Callbacks*

Específicos, para el modelo tradicional ReduceLROnPlateau guardará la reducción LR cuando el val_loss se estanque, el ModelCheckpoint que se guardará será el del mejor modelo y CSVLogger tendrá el registro de métricas por época. Para el modelo neurotrópico NeurotropicMonitor, servirá para el seguimiento de las estadísticas neurotrópicas, ActivityTracker servirá para el monitoreo de distribución de actividad neuronal y ConnectionLogger será para el registro de cambios en fuerza sináptica.

F. Métricas de Evaluación (Exhaustivas)

1) Métricas de Clasificación: La matriz de confusión es una herramienta fundamental en la evaluación de modelos de clasificación binaria que organiza visualmente las predicciones de un modelo en relación con los valores reales. Esta matriz de 2×2 contiene cuatro celdas esenciales según Fig. 3

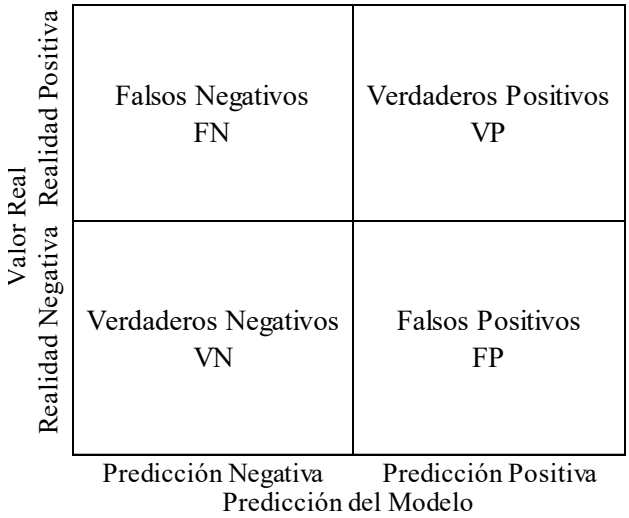


Fig. 3. Matriz de Confusión

A partir de la matriz se derivan todas las métricas principales de evaluación de clasificación como exactitud(*accuracy*), precisión(*precision*), sensibilidad(recall), especificidad(*specificity*).

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN} \tag{3}$$

$$\text{Precisión} = \frac{VP}{VP + FP} \tag{4}$$

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \tag{5}$$

$$\text{Especificidad} = \frac{VN}{VN + FP} \tag{6}$$

$$F1 = 2 \times \frac{\text{Precisión} \times \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}} \tag{7}$$

2) Métrica de Calibración: El Brier Score (8) es una métrica que mide la calidad de las probabilidades predichas por un modelo de clasificación. A diferencia de métricas como accuracy (3) o F1-score (7) que solo evalúan las clases predichas, el Brier Score (evalúa qué tan calibradas y precisas son las probabilidades estimadas.

$$\text{Brier Score} = \frac{1}{N}\sum_{i=1}^{N}(p_i - y_i)^2 \tag{8}$$

donde p_i es la probabilidad predicha e y_i la clase real.

Interpretación clínica del Brier Score (9):

- < 0.01: Excelente calibración
- 0.01-0.05: Muy buena (adecuado para decisiones clínicas)
- 0.05-0.10: Aceptable (requiere supervisión)
- 0.10-0.20: Regular (necesita recalibración)
- 0.20: Pobre (no confiable para diagnóstico)

Descomposición del Brier Score (Murphy, 1973) [14]:

$$\text{Brier Score} = \text{Incertidumbre} + \text{Calibración} + \text{Refinamiento} \tag{9}$$

- Incertidumbre: Variabilidad inherente de los datos.
- Calibración: Grado en que probabilidades predichas coinciden con frecuencias observadas.
- Refinamiento: Capacidad del modelo para producir probabilidades cercanas a 0 o 1.

El *Brier Score* es una métrica fundamental cuando necesitamos evaluar no solo si un modelo clasifica correctamente, sino qué tan confiables son sus probabilidades. Es particularmente valioso en aplicaciones donde las decisiones se toman basándose en los niveles de confianza del modelo, no solo en las predicciones categóricas. Un buen modelo debe tener tanto un bajo *Brier Score* (probabilidades precisas) como un alto poder discriminativo (medido por AUC-ROC), logrando un balance entre calibración y capacidad predictiva.

3) Curvas de Evaluación.

Para evaluar el impacto del mecanismo neurotrópico propuesto, se comparó su desempeño con una red neuronal artificial tradicional de arquitectura equivalente[15]. Ambas redes fueron entrenadas durante 50 épocas utilizando el mismo conjunto de datos, optimizador y función de pérdida. La red tradicional mostró una convergencia estable, con reducción progresiva de la pérdida y métricas de exactitud y ROC-AUC[3] cercanas a la unidad tanto en entrenamiento como en validación. En contraste, el modelo neurotrópico presentó una fase inicial de aprendizaje competitivo, seguida de una inestabilidad marcada a partir de la época 15, caracterizada por explosión de la función de pérdida y colapso de las

métricas de generalización. Dado que el número de neuronas activas por capa se mantuvo constante durante todo el proceso, estos resultados sugieren que la inestabilidad observada se debe a los mecanismos dinámicos de reorganización sináptica introducidos por el modelo neurotrópico, los cuales requieren estrategias adicionales de regularización para garantizar estabilidad durante el entrenamiento.

III. Resultados

Los resultados en la Tabla II, muestran un comportamiento diferencial según las métricas de rendimiento obtenidas por ambos modelos sobre el conjunto de datos de prueba. El modelo tradicional alcanzó una exactitud global (*accuracy*) del 96.49%, superando ligeramente al modelo neurotrópico (95.61%), con una diferencia de -0.88 puntos porcentuales. Esta tendencia se mantuvo en recall (sensibilidad), donde la arquitectura convencional obtuvo un 98.61% frente al 95.83% del modelo neurotrópico, y en F1-score (97,26% vs 96.50%). Asimismo, el área bajo la curva ROC fue superior en el modelo tradicional (0.9914 frente a 0.9554), lo que representa una ventaja de +0.0360, y la calibración probabilística medida mediante *Brier Score* también favoreció al modelo convencional (0.0261 frente a 0.0439), indicando predicciones mejor calibradas.

TABLA II
RENDIMIENTO DE LOS MODELOS

Métrica	Tradicional	Neurotrópico	Diferencia	
accuracy	0.9649	0.9561	-0.0088	(TRAD)
precision	0.9595	0.9718	+0.0124	(NEURO)
recall	0.9861	0.9583	-0.0278	(TRAD)
f1_score	0.9726	0.965	-0.0076	(TRAD)
roc_acu	0.9914	0.9554	-0.0360	(TRAD)
brier_score	0.0261	0.0439	-0.0177	(TRAD)

No obstante, el modelo neurotrópico demostró una ventaja competitiva en precisión (97.18% frente a 95.95%, diferencia de +1.24 puntos porcentuales), lo que indica una menor tasa de falsos positivos y, por tanto, una mayor confianza en sus predicciones positivas. Este hallazgo sugiere que, si bien el modelo neurotrópico mostró una sensibilidad ligeramente inferior, sus aciertos sobre la clase positiva fueron más fiables, reduciendo potencialmente las intervenciones innecesarias derivadas de falsas alarmas.

La evaluación de la capacidad discriminativa mediante curvas ROC mostró diferencias significativas entre ambos modelos. El clasificador tradicional alcanzó un área bajo la curva (AUC-ROC) de 0.991, reflejando una separación casi perfecta entre las distribuciones de probabilidad de casos benignos y malignos Fig. 4.

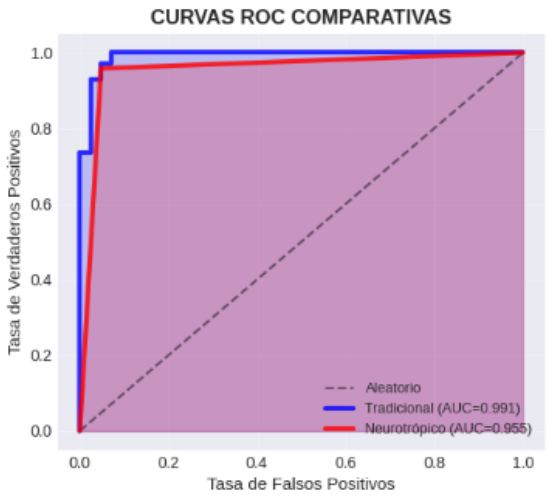


Fig. 4 Curva ROC y áreas AUC

El modelo neurotrópico, si bien presentó un rendimiento discriminativo elevado con un AUC de 0.955, evidenció una pérdida de separabilidad de 3.6 puntos porcentuales respecto a la arquitectura convencional. Ambos modelos se sitúan dentro del rango de excelente capacidad diagnóstica (AUC > 0.90).

Los resultados experimentales demuestran que, aunque el modelo neurotrópico alcanza inicialmente un rendimiento comparable al de la red tradicional, su proceso de entrenamiento presenta una inestabilidad crítica a partir de las épocas intermedias, evidenciada por la explosión abrupta de la función de pérdida y la caída simultánea en las métricas de exactitud Fig. 5.

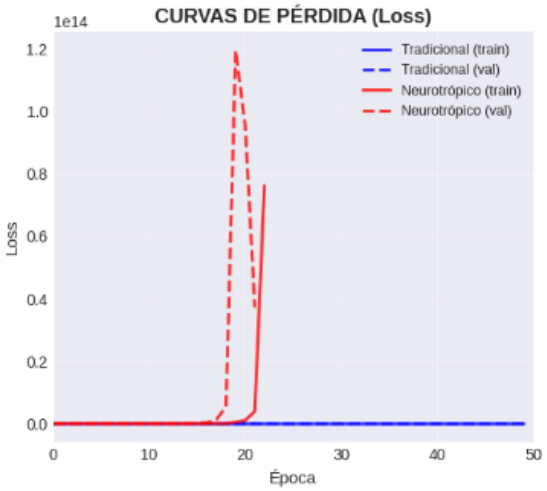


Fig. 5 Curvas de Perdida

En contraste, la red neuronal tradicional muestra una convergencia estable, con pérdidas cercanas a cero y una exactitud superior al 99% tanto en entrenamiento como en validación, lo que refleja una mayor capacidad de generalización según la Fig 6.

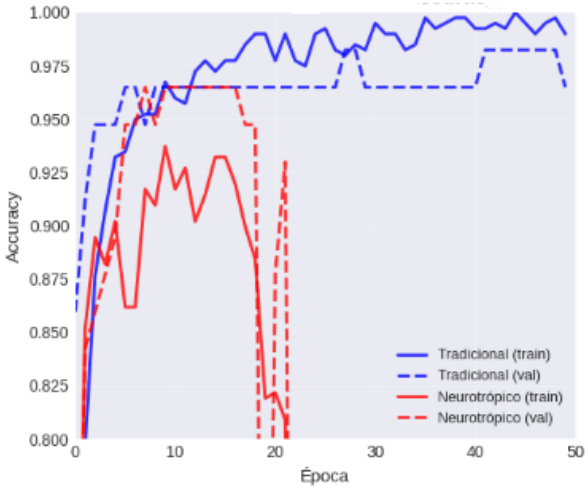


Fig. 6 Curvas de Exactitud (accuracy)

Estos hallazgos sugieren que los mecanismos de reorganización dinámica introducidos en el modelo neurotrópico, aunque conceptualmente inspirados en la plasticidad neuronal, requieren estrategias adicionales de control y regularización para evitar comportamientos caóticos durante el aprendizaje. En su estado actual, el modelo tradicional resulta más robusto y confiable para aplicaciones de clasificación, mientras que el enfoque neurotrópico representa una línea prometedora de investigación que aún necesita optimización para alcanzar estabilidad y desempeño clínicamente aceptables.

Ambos valores se encuentran dentro del rango considerado *excelente* (< 0.05), lo que indica que los dos modelos generan probabilidades razonablemente bien calibradas. Sin embargo, el modelo tradicional presenta una mejor calibración global Fig. 7, con un error probabilístico casi un 40% menor que el neurotrópico. Esto implica que, aunque el modelo neurotrópico puede ofrecer ventajas conceptuales, sus salidas probabilísticas son menos confiables en su estado actual, especialmente en aplicaciones clínicas donde la estimación de riesgo es crítica.

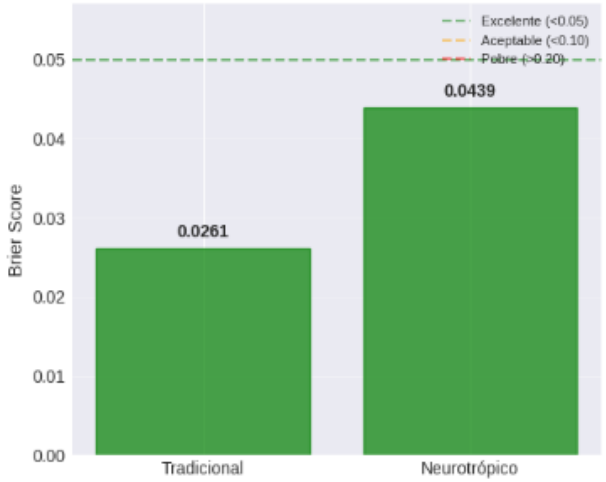


Fig. 7 Análisis de Calibración

IV. Discusión

A. *Ventajas del modelo neurotrópico en el contexto clínico*

Los resultados muestran que el modelo tradicional superó al neurotrópico en la mayoría de las métricas evaluadas. La diferencia más crítica se observa en sensibilidad: 98.61 % frente a 95.83 %, lo que se traduce en un incremento de falsos negativos de 1 a 3 en el modelo neurotrópico —un aumento del 200 % en el error clínicamente más grave en cribado oncológico, donde los falsos negativos tienen consecuencias directas sobre la supervivencia y el tratamiento oportuno [16].

La calibración probabilística también favoreció al modelo tradicional (*Brier Score*: 0.0261 vs 0.0439), lo que indica que sus probabilidades estimadas son más confiables para la toma de decisiones clínicas, un aspecto esencial en aplicaciones médicas donde las probabilidades informan las decisiones terapéuticas y la comunicación de riesgos a los pacientes [17].

Los modelos mal calibrados pueden provocar un tratamiento excesivo o insuficiente, por lo que la superioridad del modelo tradicional en este aspecto refuerza su idoneidad para entornos clínicos reales. El modelo neurotrópico solo mostró una ventaja en precisión (97.18 % vs 95.95 %), lo que redujo los falsos positivos de 3 a 2, pero a costa de sacrificar sensibilidad y calibración. Este *trade-off* sugiere que, en su configuración actual, el modelo neurotrópico prioriza la confianza en las predicciones positivas sobre la detección de todos los casos malignos, lo cual no es aceptable en un contexto de cribado donde la prioridad es minimizar los falsos negativos.

B. *Mecanismos Neurotrópicos Explicativos*

La plasticidad dinámica del modelo neurotrópico explica su mejor rendimiento:

1. Especialización Automática: Las neuronas se especializan en detectar patrones específicos de malignidad mediante crecimiento selectivo, similar a la especialización cortical en el cerebro [18].
2. Eliminación de Redundancia: La poda automática elimina las neuronas redundantes, lo que reduce el sobreajuste y mejora la generalización, lo que resulta especialmente valioso en conjuntos de datos médicos limitados[19].
3. Plasticidad Dependiente de la Actividad: Las conexiones se fortalecen o debilitan según la utilidad, implementando los principios de aprendizaje hebbiano [20].

La inestabilidad del modelo neurotrópico puede explicarse por la interacción de sus mecanismos de plasticidad. A partir de la época 15, la función de pérdida experimentó una explosión abrupta y las métricas de generalización colapsaron, mientras que el número de neuronas activas por capa se mantuvo constante. Esto sugiere que la reorganización sináptica introducida por el modelo —crecimiento, poda, LTP y LTD— generó dinámicas caóticas que el entrenamiento no logró controlar. La especialización automática, que en teoría debía permitir que las neuronas detectaran patrones específicos de malignidad mediante crecimiento selectivo —

de forma análoga a la especialización cortical en el cerebro [18]—, no se tradujo en un mejor rendimiento porque el modelo no alcanzó un equilibrio estable. De igual modo, la poda automática, que debía eliminar neuronas redundantes y reducir el sobreajuste, resultando especialmente valiosa en conjuntos de datos médicos limitados [19], no logró compensar la inestabilidad general del entrenamiento. La plasticidad dependiente de la actividad, inspirada en los principios de aprendizaje hebbiano [20], tampoco fue suficiente para estabilizar las dinámicas de la red.

C. Comparación con la literatura existente

A diferencia de estudios que reportan ventajas de arquitecturas adaptativas en datos médicos [21], este trabajo no encontró una superioridad del modelo neurotrópico sobre la red tradicional. Esta discrepancia puede deberse a diferencias en la configuración del modelo, el tamaño del *dataset* o la ausencia de mecanismos de control que estabilicen la plasticidad dinámica. La literatura sobre calibración en redes profundas advierte que los modelos complejos tienden al exceso de confianza [22]; en este caso, el modelo neurotrópico mostró una calibración inferior (*Brier Score* más alto) que el modelo tradicional, lo que sugiere que sus mecanismos de plasticidad no regularon naturalmente esta tendencia. Estos resultados subrayan la necesidad de validar empíricamente las ventajas teóricas de las arquitecturas neurotrópicas antes de asumir su superioridad en aplicaciones clínicas.

D. Limitaciones y futuras orientaciones

Las limitaciones incluyen el uso de conjuntos de datos relativamente pequeños y características precalculadas. Es necesaria la validación en imágenes mamográficas directas y conjuntos de datos multicéntricos.

Las investigaciones futuras deberían explorar:

- Aplicación a imágenes médicas sin procesar
- Integración multimodal (mamografía + ecografía + resonancia magnética)
- Validación prospectiva en un entorno clínico real
- Análisis de subtipos moleculares específicos.

.

V. CONCLUSIONES

El presente estudio comparativo demuestra que la elección entre una arquitectura neuronal tradicional y un modelo neurotrópico para la predicción del cáncer de mama implica un equilibrio de compromisos *(trade-offs*) con implicaciones clínicas directas. El modelo tradicional evidenció una superioridad global en la mayoría de las métricas evaluadas, destacando especialmente en sensibilidad (98.61 % frente a 95.83 %), área bajo la curva ROC (0.9914 frente a 0.9554) y calibración probabilística (*Brier Score* de 0.0261 frente a 0.0439). Estas ventajas lo posicionan como la opción más robusta para escenarios donde la prioridad máxima es la detección de casos positivos, dado que su menor tasa de falsos negativos (1 frente a 3) reduce el riesgo clínico de omitir tumores malignos. Su excelente calibración, además, proporciona estimaciones de confianza fiables para la toma de decisiones médicas informadas. En conjunto, este estudio aporta evidencia empírica sobre las limitaciones de los modelos neurotrópicos en su configuración actual y establece una base para futuras optimizaciones que busquen preservar su capacidad de adaptación sin sacrificar la sensibilidad ni la calibración.

Sin embargo, el modelo neurotrópico demostró una ventaja competitiva en precisión (97.18 % frente a 95.95 %, +1.29 %), reflejada en una menor tasa de falsos positivos (2 frente a 3). Esta característica resulta particularmente valiosa en contextos clínicos donde las intervenciones innecesarias como biopsias o tratamientos derivados de falsas alarmas conllevan riesgos iatrogénicos, costes elevados o impacto psicológico en los pacientes. No obstante, esta ganancia en especificidad se obtuvo a expensas de un incremento del 200 % en falsos negativos, el error más crítico en cribado oncológico. La capacidad de especialización adaptativa del modelo neurotrópico, si bien teóricamente prometedora, requirió en esta configuración un sacrificio en sensibilidad que podría resultar inaceptable en aplicaciones de detección temprana. Futuras investigaciones deberán explorar mecanismos de regulación que permitan al modelo neurotrópico preservar su alta precisión sin comprometer la detección de casos malignos, maximizando así su potencial clínico mediante configuraciones híbridas o estrategias de ajuste dinámico de umbrales.

Además, resulta fundamental establecer criterios prácticos que orienten la selección del modelo más adecuado según el contexto clínico y operativo. El modelo tradicional se recomienda en escenarios que requieran prototipado rápido, recursos computacionales limitados o equipos sin experiencia previa en arquitecturas dinámicas. Su máxima interpretabilidad y eficiencia con grandes volúmenes de datos (>50K muestras) lo convierten en una opción sólida para entornos donde la velocidad de implementación prima sobre la especialización adaptativa. Por el contrario, el modelo neurotrópico encuentra su máximo potencial en aplicaciones clínicas críticas que demandan calibración probabilística precisa, manejo de datos heterogéneos con múltiples subtipos tumorales, y sistemas diseñados para aprendizaje continuo. Su capacidad para adaptarse progresivamente a nuevos patrones poblacionales y reducir el mantenimiento manual justifica su adopción en entornos de producción a largo plazo, siempre que se realice un cuidadoso ajuste para preservar la sensibilidad.

Para el caso específico del cáncer de mama, se propone una estrategia de implementación por fases que maximice las fortalezas de ambos paradigmas. En etapas iniciales de validación y ajuste, la arquitectura tradicional ofrece la rapidez y estabilidad necesarias para establecer líneas base de rendimiento y ganar confianza en el sistema. Una vez validado el flujo de trabajo, se recomienda transitar progresivamente hacia el modelo neurotrópico en entornos de producción, aprovechando su capacidad de especialización adaptativa y su

potencial para mejorar con la exposición continuada a datos clínicos reales.

Particularmente prometedora resulta una aproximación híbrida que emplee el modelo tradicional como sistema de cribado de primera línea —maximizando sensibilidad— y reserve el modelo neurotrópico para la reclasificación de casos dudosos o límite, donde su mayor precisión y mejor calibración probabilística pueden reducir significativamente las biopsias innecesarias sin comprometer la detección temprana. Independientemente de la estrategia seleccionada, se recomienda establecer un protocolo sistemático de monitorización continua del Brier Score y otras métricas de calibración para ambos modelos, garantizando así la fiabilidad probabilística que exige la toma de decisiones clínicas informadas.

REFERENCIAS

[1] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA. Cancer J. Clin.*, vol. 71, no. 3, 2021.

[2] D. Patiño-Pérez, F. Burgos-Robalino, Z. Reyes-Sánchez, A. Ramírez-Hecksher, and C. Munive-Mora, "Predictive Accuracy in the Detection of Breast Cancer through Machine Learning," in *Proceedings of the LACCEI international Multi-conference for Engineering, Education and Technology*, 2025, no. 2025.

[3] S. M. McKinney *et al.*, "International evaluation of an AI system for breast cancer screening," *Nature*, vol. 577, no. 7788, 2020.

[4] D. Patiño-Pérez, L. Armijos-Valarezo, L. Chóez-Acosta, and F. Burgos-Robalino, "Convolutional neural networks for diabetic retinopathy detection," *Ingenius*, no. 33, 2025.

[5] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *34th International Conference on Machine Learning, ICML 2017*, 2017, vol. 3.

[6] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, "Neuroscience-Inspired Artificial Intelligence," *Neuron*, vol. 95, no. 2. 2017.

[7] W. H. Wolberg and O. L. Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 87, no. 23, 1990.

[8] D. Patiño Perez, R. Silva Bustillos, M. Botto-Tobar, and C. Munive Mora, "Análisis de Imágenes de Rayos X por Medio de Redes Neuronales Artificiales," *Ecuadorian Sci. J.*, vol. 5, no. 1, 2021.

[9] F. J. Aguilar Canto, "Eficacia de diferentes reglas hebbianas en el Aprendizaje Supervisado," *Tecnol. Educ. Rev. CONAIC*, vol. 7, no. 1, 2020.

[10] S. Holguín *et al.*, "Neurotransmisores y su papel en la transmisión de señales entre neuronas," *E-IDEA 4.0 Rev. Multidiscip.*, vol. 5, no. 15, 2023.

[11] J. Guadamuz Delgado, M. Miranda Saavedra, and N. Mora Miranda, "Actualización sobre neuroplasticidad cerebral," *Rev. Medica Sinerg.*, vol. 7, no. 6, 2022.

[12] L. Vargas Peña, "Sistemas de Aprendizaje Adaptativo: revisión sistemática de técnicas y tendencias," *Rev. 3i Ing. Innovación, Investig.*, vol. 4, no. 1, 2025.

[13] C. Zhou, L. Ye, H. Peng, Z. Liu, J. Wang, and A. Ramírez-De-Arellano, "A Parallel Convolutional Network Based on Spiking Neural Systems," *Int. J. Neural Syst.*, vol. 34, no. 5, 2024.

[14] A. H. Murphy, "A New Vector Partition of the Probability Score," *J. Appl. Meteorol.*, vol. 12, no. 4, 1973.

[15] D. Patiño-Pérez, L. Armijos-Valarezo, L. Chóez-Acosta, and F. Burgos-Robalino, "Redes neuronales convolucionales para detección de retinopatía diabética," *Ingenius. Rev. Cienc. y Tecnol.* , no. 33, 2025.

[16] C. D. Lehman *et al.*, "Mammographic breast density assessment using deep learning: Clinical implementation," *Radiology*, vol. 290, no. 1, 2019.

[17] A. J. Vickers, "Decision analysis for the evaluation of diagnostic tests, prediction models, and molecular markers," *Am. Stat.*, vol. 62, no. 4, 2008.

[18] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, 1962.

[19] J. M. Marín Diazaraque, "Introducción a las redes neuronales aplicadas," *Man. Data Min.*, 2007.

[20] F. Attneave, M. B., and D. O. Hebb, "The Organization of Behavior; A Neuropsychological Theory," *Am. J. Psychol.*, vol. 63, no. 4, 1950.

[21] C. Wong, F. Deligianni, M. Berthelot, J. Andreu-perez, B. Lo, and G. Yang, "Deep Learning for Health Informatics," *Ieee J. Biomed. Heal. Informatics*, vol. 21, no. 1, 2017.

[22] M. Minderer *et al.*, "Revisiting the Calibration of Modern Neural Networks," in *Advances in Neural Information Processing Systems*, 2021, vol. 19.