



Optimización de redes neuronales mediante cálculo diferencial

Optimization of neural networks through differential calculus



Oscar Gabriel Toapanta Cunalata ^I

oscartoapantaambjlm@gmail.com



<https://orcid.org/0000-0002-5816-1785>

Instituto Superior Tecnológico Pelileo,
Pelileo – Ecuador

José Geovanny Vega Pérez ^{II}

jg.vega@uta.edu.ec



<https://orcid.org/0009-0004-9297-0158>

Facultad de Ciencia e Ingeniería en Alimentos y Biotecnología (FCIAB)
Universidad Técnica de Ambato (UTA)
Ambato - Ecuador

Luis Javier Tapia Vasco ^{III}

ljtapia@espol.edu.ec



<https://orcid.org/0000-0003-2898-5895>

Escuela Superior Politécnica del Litoral
Guayaquil - Ecuador

Jhonatan David Vistín Bastidas ^{IV}

jhonatan.vistin@epoch.edu.ec



<https://orcid.org/0009-0002-3171-4412>

Instituto Superior Tecnológico Pelileo,
Pelileo – Ecuador

Fecha de envío: 2026-06-18

Fecha de revisión: 2026-07-21

Fecha de aceptación: 2026-08-28

Resumen

El objetivo de la investigación fue determinar la incidencia del cálculo diferencial en la optimización de redes neuronales artificiales mediante una intervención didáctica apoyada en realidad aumentada, orientada al desarrollo de destrezas para interpretar gradientes, funciones de pérdida y procesos de retropropagación. La metodología correspondió a un estudio cuasi experimental, de enfoque correlacional descriptivo, con grupo de control y grupo experimental; participaron 80 estudiantes distribuidos en cuatro centros de entrenamiento académico. Se aplicó un test de base estructurada, validado por expertos, cuyo índice de confiabilidad mediante alfa de Cronbach fue 0.89, valor considerado muy alto. El análisis integró estadística descriptiva, correlación de Pearson, prueba t de Student para muestras independientes y tamaño del efecto mediante d de Cohen. Los resultados evidenciaron una mejora superior en el grupo experimental, especialmente en la comprensión de la regla de la cadena aplicada a la retropropagación, la lectura matemática del gradiente y la selección de tasas de aprendizaje. La diferencia de medias en el posttest fue estadísticamente significativa y el tamaño del efecto fue grande, lo que sugiere que el uso articulado del cálculo diferencial y la visualización inmersiva favoreció aprendizajes más transferibles hacia problemas de optimización computacional. Se concluye que el cálculo diferencial no debe presentarse como contenido aislado, sino como lenguaje operativo para explicar cómo aprende una red neuronal, cómo corrige sus pesos y cómo reduce el error durante el entrenamiento.

Palabras clave: cálculo diferencial; redes neuronales artificiales; optimización; realidad aumentada; aprendizaje automático.

Abstract

The objective of this study was to determine the influence of differential calculus on the optimization of artificial neural networks through an augmented reality-supported instructional intervention aimed at developing skills for interpreting gradients, loss functions, and backpropagation processes. The methodology followed a quasi-experimental design with a descriptive-correlational approach, including a control group and an experimental group; 80 students participated across four academic training centers. A structured test was applied and validated by experts, obtaining a Cronbach's alpha reliability index of 0.89, which is considered very high. The analysis integrated descriptive statistics, Pearson correlation, Student's t test for independent samples, and Cohen's d effect size. The results showed greater improvement in the experimental group,

particularly in understanding the chain rule applied to backpropagation, mathematical interpretation of gradients, and selection of learning rates. The posttest mean difference was statistically significant and the effect size was large, suggesting that the combined use of differential calculus and immersive visualization promoted more transferable learning for computational optimization problems. It is concluded that differential calculus should not be taught as an isolated topic, but as an operational language that explains how a neural network learns, how it adjusts its weights, and how it reduces error during training.

Keywords: differential calculus; artificial neural networks; optimization; augmented reality; machine learning.

Introducción

La optimización de redes neuronales artificiales (RNA) se ha convertido en un campo estratégico para comprender el aprendizaje automático contemporáneo. Una RNA no “aprende” por intuición ni por magia digital; aprende porque un conjunto de parámetros se ajusta iterativamente para reducir una función de pérdida. En ese proceso, el cálculo diferencial cumple una función estructural: permite medir variaciones, estimar pendientes, construir gradientes y actualizar pesos mediante reglas matemáticas que conectan el error observado con la corrección del modelo.

La literatura de aprendizaje profundo confirma que el cálculo diferencial no es un adorno teórico, sino el mecanismo que vuelve entrenable a la red. LeCun et al. (1998) demostraron que los métodos basados en gradiente podían sostener tareas complejas de reconocimiento visual; Hochreiter y Schmidhuber (1997) introdujeron arquitecturas recurrentes capaces de manejar dependencias temporales; Goodfellow et al. (2014) ampliaron el debate hacia modelos generativos adversariales; He et al. (2016) mostraron que la profundidad podía administrarse con conexiones residuales; y Vaswani et al. (2017) consolidaron la atención como una arquitectura capaz de desplazar muchas soluciones recurrentes.

El problema pedagógico aparece cuando estos conceptos se enseñan separados. Por un lado, el cálculo diferencial suele presentarse como una secuencia de reglas formales; por otro, las redes neuronales se explican desde diagramas, código o aplicaciones, sin hacer visible el puente matemático que une ambos dominios. Esta fragmentación produce aprendizajes frágiles: el estudiante puede derivar una función, pero no siempre reconoce qué significa esa derivada cuando una función de activación satura; puede ejecutar un

algoritmo, pero no siempre interpreta por qué una tasa de aprendizaje demasiado alta vuelve inestable el descenso del gradiente. Bottou et al.

En ese marco, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) ha insistido en que la transformación digital educativa debe orientarse a capacidades humanas, pensamiento crítico, equidad de acceso y uso responsable de la inteligencia artificial (IA). Su guía sobre IA generativa en educación y investigación plantea que las instituciones no deben limitarse a incorporar tecnología, sino formar criterios para comprender, evaluar y gobernar sus usos (UNESCO, 2023).

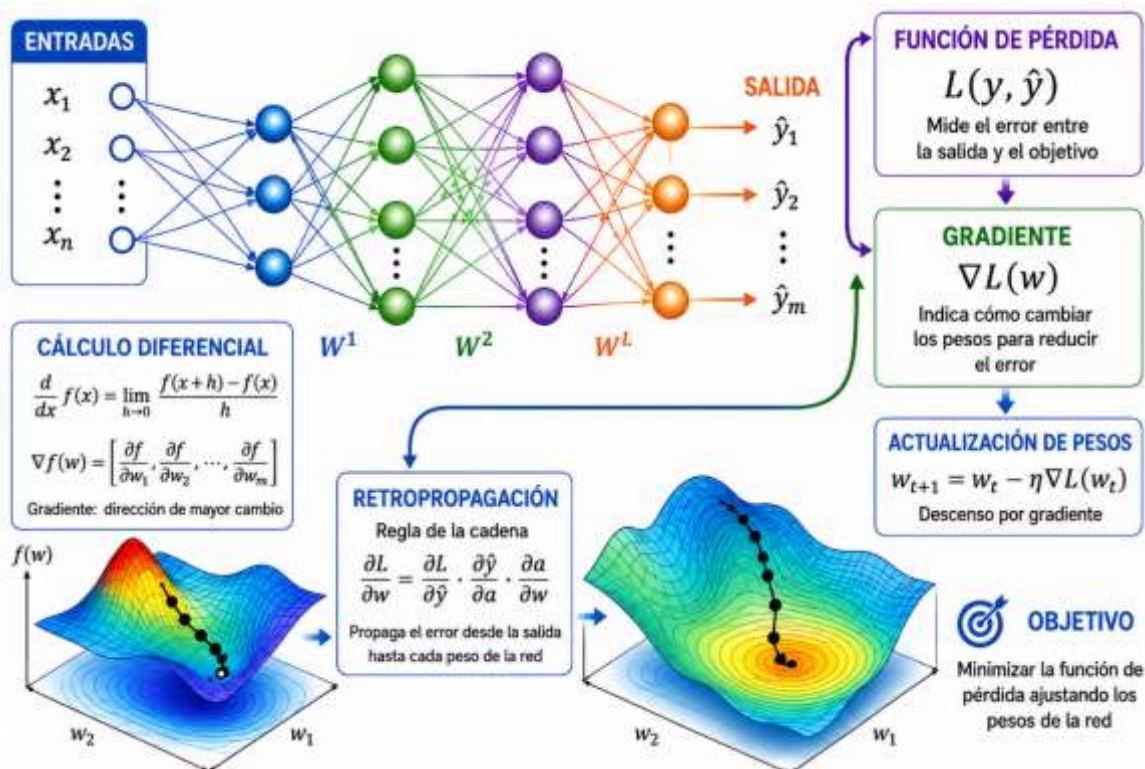
La pertinencia del tema también se explica por la expansión de experiencias de realidad aumentada (RA) en educación. Azuma (1997) definió la RA como la integración de objetos virtuales en entornos reales con interacción en tiempo real; posteriormente, Wu et al. (2013), Bacca et al. (2014), Akçayır y Akçayır (2017), Ibáñez y Delgado-Kloos (2018) y Garzón et al. (2020) reportaron que esta tecnología puede fortalecer la visualización, la motivación, la retención y la comprensión de fenómenos abstractos.

La relación entre cálculo diferencial y redes neuronales también exige considerar la arquitectura cognitiva del aprendizaje. Sweller (1988) explicó que la sobrecarga cognitiva aparece cuando el estudiante debe procesar demasiadas relaciones simultáneas sin apoyos adecuados; Kirschner et al. (2006) advirtieron que la exploración sin guía puede ser insuficiente para aprender contenidos complejos; y Mayer (2009) sostuvo que la combinación pertinente de representaciones verbales y visuales favorece la construcción de modelos mentales.

En el plano técnico, las funciones de activación, la regla de la cadena y la actualización de parámetros constituyen un itinerario didáctico natural. Ioffe y Szegedy (2015) mostraron que normalizar activaciones puede estabilizar el aprendizaje; He et al. (2015) estudiaron la inicialización y activaciones rectificadas para redes profundas; Glorot y Bengio (2010) analizaron la dificultad de entrenar redes profundas cuando las señales se desvanecen o explotan; y Srivastava et al. (2014) propusieron mecanismos de regularización para reducir el sobreajuste. Como se observa en el Gráfico 1, el cálculo diferencial constituye el fundamento matemático del proceso de optimización de una red neuronal, puesto que permite cuantificar el error mediante una función de pérdida, calcular el gradiente, propagarlo a través de las capas y actualizar los pesos en la dirección que favorece la reducción del error.

Gráfico 1

Relación entre el cálculo diferencial y el proceso de optimización de una red neuronal artificial.



Nota. La figura representa el cálculo de la pérdida, el gradiente, la retropropagación y la actualización de pesos durante el entrenamiento neuronal.

Desde el punto de vista metodológico, evaluar una intervención de este tipo exige algo más que opiniones de satisfacción. El aprendizaje debe medirse mediante instrumentos válidos y confiables. Cronbach (1951) estableció criterios clásicos para estimar consistencia interna; Schober et al. (2018) precisaron el uso interpretativo de la correlación de Pearson; Student (1908) abrió el camino para contrastar diferencias de medias; Cohen (1992), Lakens (2013) y Sullivan y Feinn (2012) recordaron que el valor p debe acompañarse de tamaños del efecto para interpretar relevancia práctica.

A partir de lo anterior, este artículo sostiene que el cálculo diferencial puede convertirse en un puente formativo entre la matemática universitaria y la inteligencia artificial aplicada. El estudio propone una intervención didáctica basada en NeuroDerivAR, una aplicación de realidad aumentada diseñada para representar funciones de pérdida, derivadas, gradientes y trayectorias de optimización en escenarios manipulables. La pregunta central no es si la tecnología llama la atención, porque eso lo hace cualquier

pantalla bien iluminada; la pregunta sería es si mejora la comprensión matemática necesaria para optimizar una red neuronal.

Objetivo

Determinar la incidencia del cálculo diferencial, mediado por la aplicación de realidad aumentada NeuroDerivAR, en el desarrollo de destrezas para la optimización de redes neuronales artificiales en participantes formados en cuatro centros académicos de entrenamiento, mediante la comparación entre un grupo de control y un grupo experimental.

Metodología

En el Gráfico 2 se puede observar el proceso que la investigación se desarrolló como un estudio cuasi experimental de enfoque correlacional descriptivo, debido a que se trabajó con grupos previamente organizados en contextos formativos reales y no con una asignación aleatoria absoluta de participantes.

Gráfico 2

Metodología aplicada al trabajo investigativo



Se consideró pertinente este diseño porque el propósito no fue únicamente describir niveles de desempeño, sino comparar el avance entre un grupo de control y un grupo experimental después de una intervención didáctica centrada en el uso del cálculo diferencial para optimizar redes neuronales artificiales. La muestra estuvo conformada por 80 participantes, distribuidos en cuatro centros de entrenamiento académico, con 20 participantes por centro; en cada centro se organizó una sección de control y una sección experimental, manteniendo criterios equivalentes de horario, contenidos base y duración. El grupo de control trabajó con una estrategia convencional basada en explicación magistral, resolución de ejercicios y análisis de código, mientras que el grupo experimental trabajó con la aplicación de realidad aumentada NeuroDerivAR, diseñada para visualizar funciones de pérdida, derivadas parciales, vectores gradiente, regla de la cadena, trayectorias de descenso y variación de la tasa de aprendizaje.

Resultados

Los resultados de la investigación se presentan mediante el análisis comparativo de las mediciones pretest y posttest, considerando la variación cuantitativa del desempeño entre el grupo de control y el grupo experimental, así como la estimación de diferencias de medias, asociaciones entre variables y magnitud del efecto atribuible a la intervención didáctica.

En primer lugar, se compara el comportamiento de ambos grupos con el propósito de identificar el avance asociado al uso de la aplicación de realidad aumentada NeuroDerivAR. Posteriormente, se analizan las ganancias de aprendizaje por destreza matemática, las correlaciones entre fundamentos del cálculo diferencial e indicadores de optimización neuronal, así como las diferencias estadísticas entre grupos mediante la prueba t de Student y el tamaño del efecto.

La comparación descriptiva evidencia que ambos grupos iniciaron el proceso con medias muy similares en el pretest: 10.36 para el grupo de control y 10.41 para el grupo experimental. Esta equivalencia inicial es metodológicamente relevante, ya que permite interpretar que las diferencias observadas en el posttest no responden a una ventaja previa del grupo experimental, sino al efecto de la intervención didáctica aplicada.

Gráfico 3

Desempeño global en pretest y postest según estrategia formativa aplicada.



Nota. El mayor desempeño se registró en el postest del grupo experimental NeuroDerivAR, con una media global de 16.72 sobre 20, evidenciando una ganancia de +6.31 puntos respecto al pretest.

El grupo de control incrementó su media de 10.36 a 12.18, con una ganancia de +1.82 puntos, lo que refleja una mejora moderada asociada al trabajo convencional mediante explicación, ejercicios y análisis de código. En contraste, el grupo experimental aumentó de 10.41 a 16.72, alcanzando una ganancia de +6.31 puntos, lo que demuestra un avance considerablemente mayor. Este resultado sugiere que el uso de NeuroDerivAR favoreció una comprensión más sólida de los procesos matemáticos vinculados con la optimización de redes neuronales, especialmente al integrar visualización, cálculo diferencial y análisis del descenso del gradiente.

Tabla 1

Ganancia de aprendizaje según destreza matemática y módulo de realidad aumentada aplicado.

Módulo RA	Destreza	Inicial	Final	Ganancia	Variación
Superficie 3D	Mínimos y dirección	10.2	17.1	+6.9	67.6 %
Gradiente vectorial	Derivadas parciales y pesos	10.7	17.4	+6.7	62.6 %
Cadena retroprop.	Regla cadena multicapa	9.8	16.9	+7.1	72.4 %

Tasa dinámica	Convergencia y oscilación	10.5	16.3	+5.8	55.2 %
Activaciones	Saturación y sensibilidad	10.6	15.9	+5.3	50.0 %

Nota. La mayor ganancia absoluta se registró en la aplicación de la regla de la cadena en capas sucesivas, con +7.1 puntos.

Los resultados por módulo evidencian que la mayor mejora se produjo en la comprensión de la regla de la cadena aplicada a la retropropagación, una de las destrezas más críticas para conectar cálculo diferencial y redes neuronales. Esto es coherente con la dificultad habitual del tema: muchos estudiantes pueden derivar funciones simples, pero se pierden cuando deben explicar cómo el error de salida se propaga hacia capas anteriores.

Tabla 2

Correlación de Pearson entre fundamentos de cálculo diferencial e indicadores de optimización neuronal.

Fundamento	Indicador	r de Pearson	p valor	Lectura
Derivada de activación	Saturación y sensibilidad	0.62	< 0.001	Asociación moderada-alta
Derivadas parciales	Peso con mayor error	0.71	< 0.001	Asociación alta
Regla de la cadena	Retropropagación multicapa	0.78	< 0.001	Asociación alta
Pendiente de pérdida	Dirección de actualización	0.69	< 0.001	Asociación alta
Razón de cambio	Tasa de aprendizaje	0.59	0.002	Asociación moderada
Análisis de curvatura local	Evita oscilación	0.55	0.004	Asociación moderada

Nota. La correlación más alta correspondió a la regla de la cadena y la explicación de la retropropagación multicapa, con $r = 0.78$.

Las correlaciones confirman que las destrezas de cálculo diferencial se vinculan de manera directa con la capacidad de optimizar redes neuronales. La asociación más fuerte, $r = 0.78$, aparece entre regla de la cadena y explicación de la retropropagación multicapa, lo que tiene pleno sentido matemático: la retropropagación es, en esencia, una aplicación organizada de la regla de la cadena para calcular cómo cambia la pérdida respecto de cada peso.

Tabla 3

Comportamiento de la intervención en los cuatro centros de entrenamiento académico.

Centro	Sesiones	Ganancia media	Tareas resueltas	Estabilidad	Observación
Centro Norte	96 %	+6.4	84 %	Alta	Mejor lectura de gradientes
Centro Sur	93 %	+5.9	79 %	Media-alta	Avance en pérdida
Centro Este	98 %	+6.8	88 %	Alta	Dominio de regla cadena
Centro Oeste	91 %	+5.6	76 %	Media	Dificultad con tasa

Nota. El Centro Este presentó la mayor ganancia media, con +6.8 puntos y 88 % de tareas resueltas.

La distribución por centros permite verificar que la intervención no dependió de un solo ambiente académico. Los cuatro centros mostraron ganancias positivas, aunque con diferencias razonables asociadas al cumplimiento de sesiones y al ritmo de apropiación de los contenidos. El Centro Este obtuvo la mayor ganancia y el mayor porcentaje de tareas resueltas, lo que puede explicarse por su cumplimiento de sesiones de 98 % y por el dominio alcanzado en la regla de la cadena.

Tabla 4

Perfil de logro por dimensión evaluada en la optimización de redes neuronales artificiales.

Dimensión evaluada	Nivel inicial	Nivel final	Cambio observado	Indicador clave
Funciones de activación	Básico	Alto	De fórmulas a sensibilidad	Saturación identificada
Función de pérdida	Básico	Alto	De cálculo a trayectoria	Mínimo aproximado explicado
Gradiente	Intermedio bajo	Alto	De pendiente a vector	Dirección de descenso justificada
Retropropagación	Básico	Alto	De mecánica a dependencias	Error propagado correctamente
Hiperparámetros	Básico	Intermedio alto	De ensayo a argumento	Tasa de aprendizaje regulada

Nota. El cambio más robusto se observó en retropropagación, donde el nivel final predominante fue alto.

El perfil por dimensiones muestra que la intervención no solo incrementó puntajes, sino que modificó la calidad del razonamiento. En funciones de activación, los participantes pasaron de memorizar derivadas a interpretar sensibilidad y saturación; en función de pérdida, dejaron de ver una fórmula de error como resultado final y comenzaron a leerla

como un terreno de búsqueda; en gradiente, transformaron una idea intuitiva de pendiente en un vector de actualización; y en retropropagación, comprendieron que el error.

Tabla 5

Indicadores de transferencia hacia resolución de problemas de entrenamiento neuronal.

Caso	Criterio	Control	NeuroDerivAR	Brecha
Sigmoide saturada	Pérdida de gradiente	48 %	82 %	+34 puntos
Tasa excesiva	Oscilación del error	44 %	79 %	+35 puntos
Convergencia lenta	Ajuste gradual	50 %	76 %	+26 puntos
Error en capas ocultas	Regla de cadena	41 %	83 %	+42 puntos
Curvas de pérdida	Entrenamiento estable	53 %	86 %	+33 puntos

Nota. La mayor brecha favorable al grupo NeuroDerivAR se presentó en el caso de error en capas ocultas, asociado con la aplicación de la regla de la cadena, donde el grupo experimental alcanzó 83 % frente al 41 % del grupo de control, con una diferencia de +42 puntos porcentuales.

Los indicadores de transferencia permiten valorar si el aprendizaje logrado por los participantes puede aplicarse a situaciones distintas del ejercicio mecánico o rutinario. En este sentido, la tabla evidencia que el grupo NeuroDerivAR obtuvo mejores resultados en todos los casos evaluados, especialmente en la identificación del error en capas ocultas, la interpretación de curvas de pérdida, el diagnóstico de tasas excesivas y la comprensión de la pérdida de gradiente en funciones sigmoideas saturadas. Estos resultados sugieren que la intervención con realidad aumentada no solo fortaleció la resolución procedimental, sino también la capacidad de interpretar problemas propios del entrenamiento neuronal. La brecha más alta, registrada en la aplicación de la regla de la cadena para explicar el error en capas ocultas, indica que NeuroDerivAR favoreció una comprensión más estructural de la retropropagación, entendida como un sistema de dependencias matemáticas entre capas y no únicamente como una secuencia algorítmica de cálculo.

Gráfico 4

Prueba t de Student para muestras independientes en el posttest global.



Nota. El resultado más significativo fue $p < 0.001$, lo que indica diferencia estadísticamente significativa entre grupos.

La prueba t de Student confirma inferencialmente lo observado en la descripción: la media del grupo experimental en el posttest fue significativamente mayor que la del grupo de control. El valor $t = 10.72$, con 78 grados de libertad y $p < 0.001$, permite rechazar la hipótesis nula de igualdad de medias. La diferencia de 4.54 puntos en una escala de 20 es educativa y estadísticamente relevante, porque no representa una variación marginal sino un salto de desempeño asociado a la intervención.

Tabla 6

Tamaño del efecto mediante d de Cohen según destrezas de optimización neuronal.

Destreza	d de Cohen	g de Hedges	Magnitud	Implicación
Gradientes	1.42	1.40	Grande	Lectura sólida del cambio
Regla cadena	1.56	1.53	Grande	Transferencia multicapa
Función de pérdida	1.31	1.29	Grande	Comprensión de convergencia
Tasa de aprendizaje	1.08	1.06	Grande	Diagnóstico de estabilidad
Activaciones	0.96	0.94	Grande	Sensibilidad y saturación
Posttest global	2.40	2.37	Muy grande	Efecto integral

Nota. El posttest global presentó el mayor tamaño del efecto, con $d = 2.40$.

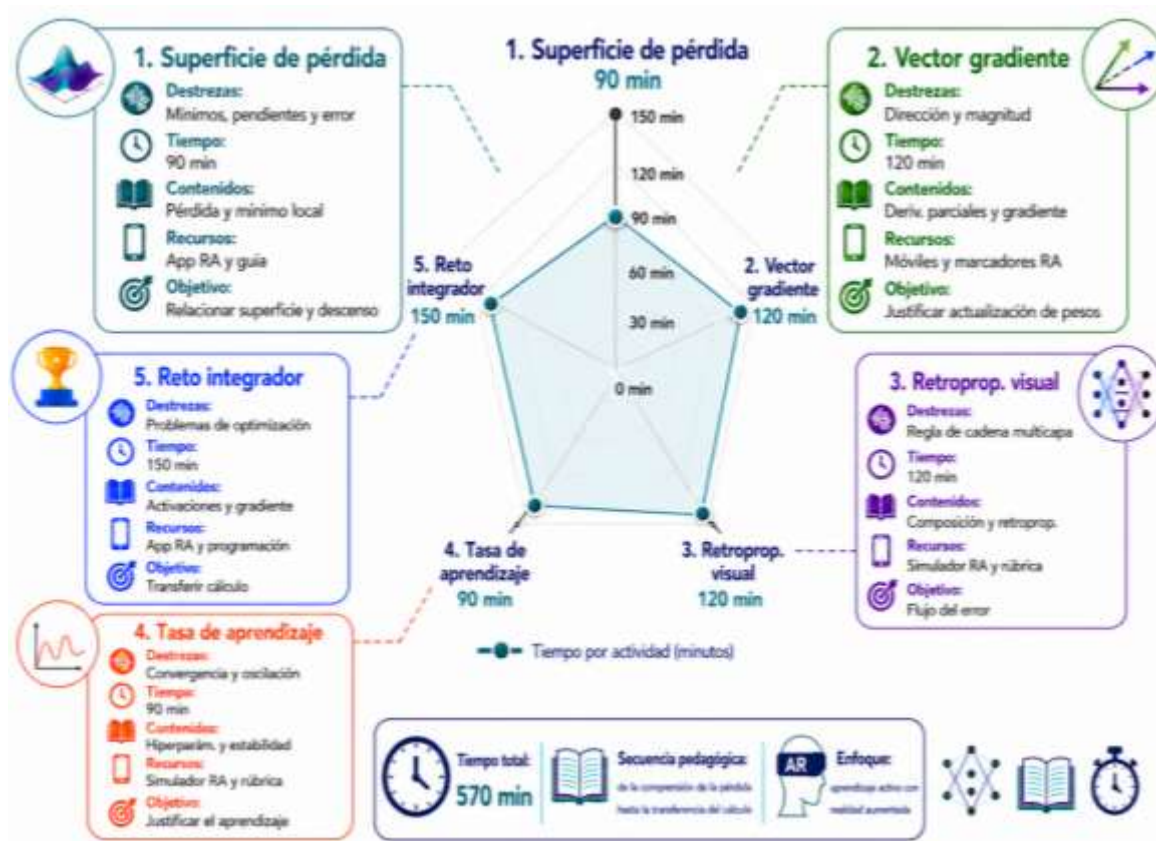
La d de Cohen permitió valorar la relevancia práctica de la intervención. Los tamaños del efecto fueron grandes en todas las destrezas y muy grande en el postest global; por ello, la diferencia entre el grupo experimental y el grupo de control no solo fue significativa, sino sustantiva desde una perspectiva educativa.

Propuesta didáctica implementada

La propuesta NeuroDerivAR fue diseñada como una secuencia didáctica de integración entre cálculo diferencial, visualización inmersiva y optimización de redes neuronales artificiales. Antes de su aplicación, la propuesta fue revisada por diez expertos en matemática aplicada, inteligencia artificial, estadística educativa, didáctica universitaria y tecnologías inmersivas.

Gráfico 5

Secuencia de actividades de la propuesta NeuroDerivAR para optimización de redes neuronales artificiales.



Nota. La actividad integradora concentró el mayor tiempo porque exigió transferencia simultánea de cálculo, análisis de gradiente y decisión algorítmica.

La secuencia se estructuró de lo simple a lo complejo: primero se observó la superficie de pérdida, luego se interpretó el gradiente, después se aplicó la regla de la cadena y finalmente se resolvieron casos completos de entrenamiento. Esta progresión evitó que la RA se convirtiera en una demostración vistosa pero vacía. Cada actividad exigió una evidencia de desempeño: explicar, justificar, comparar o resolver. El mayor valor de la propuesta fue su capacidad de conectar representaciones visuales con decisiones matemáticas.

Discusión

Los resultados obtenidos evidencian que el cálculo diferencial, cuando se enseña como lenguaje funcional de la optimización, favorece el desarrollo de destrezas necesarias para comprender el entrenamiento de redes neuronales artificiales. La mejora del grupo experimental coincide con la tradición iniciada por McCulloch y Pitts (1943), Rosenblatt (1958) y Rumelhart et al. (1986), quienes situaron la formalización matemática en el centro de los modelos neuronales. También dialoga con LeCun et al. (1998), Hochreiter y Schmidhuber (1997), Goodfellow et al. (2014), He et al. (2016) y Vaswani et al.

La relación positiva entre fundamentos de cálculo diferencial e indicadores de optimización también confirma que el aprendizaje de inteligencia artificial requiere una base matemática sólida. La correlación más alta entre regla de la cadena y explicación de retropropagación coincide con el carácter composicional de las redes profundas: cada capa transforma la entrada y la salida final depende de una cadena de funciones. Este hallazgo puede interpretarse junto con los aportes de Hochreiter (1998), Pascanu et al. (2013), Bengio et al. (1994) y Goodfellow et al.

Los resultados también son consistentes con la investigación educativa sobre realidad aumentada. Azuma (1997) estableció la base conceptual de la RA al integrar objetos virtuales en entornos reales; Wu et al. (2013), Bacca et al. (2014), Akçayır y Akçayır (2017), Ibáñez y Delgado-Kloos (2018), Garzón et al. (2020), Cheng y Tsai (2013) y Dunleavy et al. (2009) han reportado beneficios en visualización, motivación, comprensión y aprendizaje situado. En este estudio, esos beneficios se expresaron en destrezas concretas: interpretar un vector gradiente, leer una superficie de pérdida y diagnosticar una tasa de aprendizaje.

Desde la teoría cognitiva, los resultados pueden explicarse por la reducción de carga cognitiva y el fortalecimiento de representaciones múltiples. Sweller (1988) argumentó que aprender contenidos complejos exige administrar la carga mental; Kirschner et al.

(2006) advirtieron que la guía instruccional es clave en dominios de alta complejidad; Mayer (2009) defendió el valor de combinar palabras e imágenes bajo principios de coherencia y señalización. NeuroDerivAR permitió distribuir el procesamiento entre ecuaciones, visualizaciones y manipulación guiada.

La significancia estadística y los tamaños del efecto refuerzan la solidez del hallazgo. La prueba t de Student mostró una diferencia clara entre grupos, mientras que la d de Cohen evidenció una magnitud práctica grande. Esta lectura coincide con Cohen (1992), Lakens (2013), Sullivan y Feinn (2012), Cumming (2014) y Faul et al. (2007), quienes advierten que la interpretación de resultados no debe limitarse al valor p. En el estudio, la diferencia de medias fue acompañada por efectos grandes, lo que permite sostener que la intervención no solo produjo una variación estadísticamente detectable, sino una mejora formativa relevante.

El análisis por centros aporta otra dimensión. Los cuatro centros mostraron ganancias, aunque no idénticas, lo que sugiere transferibilidad con sensibilidad contextual. Esta lectura se articula con los planteamientos de UNESCO (2023), CEPAL y UNESCO (2020), UNESCO, CEPAL y UNICEF (2022) y Ministerio de Educación del Perú (2016), que insisten en la necesidad de integrar tecnología con pertinencia pedagógica y condiciones reales de implementación. La educación digital en América Latina no necesita más promesas infladas; necesita diseños verificables, medibles y adaptables.

Una implicación científica del estudio es que el cálculo diferencial puede funcionar como eje integrador para formar competencias en inteligencia artificial aplicada. Frente a una enseñanza fragmentada, la propuesta organizó derivadas, gradientes, funciones de pérdida y ajuste de parámetros dentro de un mismo problema: reducir el error de una red neuronal. Esta integración coincide con Sun et al. (2019), Nocedal y Wright (2006), Bottou et al. (2018), Duchi et al. (2011), Zeiler (2012) y Smith (2017), quienes ubican la optimización como columna vertebral del aprendizaje automático.

También se observa que las dimensiones con menor efecto relativo, como análisis de activaciones y ajuste de tasa de aprendizaje, no deben interpretarse como debilidad de la propuesta, sino como contenidos de mayor complejidad conceptual. La saturación de funciones no lineales, la elección de tasas y la estabilidad del descenso implican razonamiento sobre sensibilidad, escala y comportamiento dinámico. Estos temas se conectan con Ioffe y Szegedy (2015), Glorot y Bengio (2010), He et al. (2015), Srivastava et al.

Finalmente, los hallazgos abren una línea de investigación didáctica con valor para ingeniería, ciencia de datos y educación matemática. La optimización de redes neuronales es un tema técnico, pero su aprendizaje depende de decisiones pedagógicas precisas: qué se visualiza, qué se calcula, qué se interpreta y qué evidencia demuestra transferencia. NeuroDerivAR mostró que la realidad aumentada puede ser un medio potente cuando se diseña con fidelidad matemática, validación experta y análisis estadístico.

Conclusiones

La investigación permite concluir que el cálculo diferencial constituye un fundamento esencial para comprender la optimización de redes neuronales artificiales, especialmente cuando su enseñanza se apoya en recursos visuales e interactivos como NeuroDerivAR. El grupo experimental mostró un desempeño superior en la interpretación de gradientes, la aplicación de la regla de la cadena, el análisis de funciones de pérdida y el diagnóstico de tasas de aprendizaje, lo que evidencia una mejora tanto conceptual como procedimental.

La principal contribución del estudio radica en el diseño y la evaluación de un puente didáctico entre el cálculo diferencial, las redes neuronales artificiales y la realidad aumentada. Los resultados indican que la visualización de la superficie de pérdida, el vector gradiente y la propagación del error favorece una comprensión más profunda de los procesos de optimización. Sin embargo, su efectividad depende de una integración pedagógica planificada, sustentada en objetivos de aprendizaje definidos, validación experta e instrumentos de evaluación válidos y confiables.

Referencias

- Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review*, 20, 1-11. <https://doi.org/10.1016/j.edurev.2016.11.002>
- Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4), 355-385. <https://doi.org/10.1162/pres.1997.6.4.355>
- Bacca, J., Baldiris, S., Fabregat, R., Graf, S., & Kinshuk. (2014). Augmented reality trends in education: A systematic review of research and applications. *Educational Technology & Society*, 17(4), 133-149.

- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157-166. <https://doi.org/10.1109/72.279181>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289-300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bottou, L., Curtis, F. E., & Nocedal, J. (2018). Optimization methods for large-scale machine learning. *SIAM Review*, 60(2), 223-311. <https://doi.org/10.1137/16M1080173>
- CEPAL, & UNESCO. (2020). La educación en tiempos de la pandemia de COVID-19. Comisión Económica para América Latina y el Caribe y Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. <https://repositorio.cepal.org/handle/11362/45904>
- Cheng, K. H., & Tsai, C. C. (2013). Affordances of augmented reality in science learning: Suggestions for future research. *Journal of Science Education and Technology*, 22, 449-462. <https://doi.org/10.1007/s10956-012-9405-9>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334. <https://doi.org/10.1007/BF02310555>
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7-29. <https://doi.org/10.1177/0956797613504966>
- Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12, 2121-2159.
- Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology*, 18, 7-22. <https://doi.org/10.1007/s10956-008-9119-1>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191. <https://doi.org/10.3758/BF03193146>

- Garzón, J., Pavón, J., & Baldiris, S. (2019). Systematic review and meta-analysis of augmented reality in educational settings. *Virtual Reality*, 23, 447-459. <https://doi.org/10.1007/s10055-019-00379-9>
- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 249-256.
- Goodfellow, I. J., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672-2680. <https://doi.org/10.48550/arXiv.1406.2661>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. *Proceedings of the IEEE International Conference on Computer Vision*, 1026-1034. <https://doi.org/10.1109/ICCV.2015.123>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(2), 107-116. <https://doi.org/10.1142/S0218488598000094>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Ibáñez, M. B., & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. *Computers & Education*, 123, 109-123. <https://doi.org/10.1016/j.compedu.2018.05.002>
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on Machine Learning*, 448-456. <https://doi.org/10.48550/arXiv.1502.03167>
- Johnson-Glenberg, M. C., Birchfield, D., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments:

- Two science studies. *Journal of Educational Psychology*, 106(1), 86-104.
<https://doi.org/10.1037/a0034008>
- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations*.
<https://doi.org/10.48550/arXiv.1412.6980>
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work. *Educational Psychologist*, 41(2), 75-86.
https://doi.org/10.1207/s15326985ep4102_1
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science. *Frontiers in Psychology*, 4, Article 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
<https://doi.org/10.1109/5.726791>
- Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. *International Conference on Learning Representations*.
<https://doi.org/10.48550/arXiv.1711.05101>
- Mayer, R. E. (2009). *Multimedia learning*. Cambridge University Press.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.
<https://doi.org/10.1007/BF02478259>

Declaración

Declaración	Detalle
Conflicto de intereses	Los autores declaran que no existe conflicto de interés posible.
Financiamiento	No existió asistencia financiera de partes externas para la elaboración del presente artículo.
Agradecimiento	N/A
Nota	El artículo no es producto de una publicación anterior.