

## Análisis bucodental empleando sensores inerciales mediante algoritmos de deep learning

Oral analysis using inertial sensors using a Deep Learning model

**Luis Oswaldo Betancourt Safla, Ing. <sup>(1)</sup>**  
**Carlos Alfredo Silva Villafuerte, Msc. <sup>(2)</sup>**

<sup>(1)</sup> Universidad Técnica de Manabí, Facultad de Posgrado, Maestría Ciencia de Datos y Máquinas de Aprendizaje. Portoviejo, Manabí, Ecuador. Correo: lbetancourt0733@utm.edu.ec, Código Orcid: <https://orcid.org/0009-0009-9242-3708>

<sup>(2)</sup> Universidad Técnica de Manabí, Portoviejo – Ecuador, Correo: carlos.silva@utm.edu.ec, Código Orcid: <https://orcid.org/0000-0002-0328-8033>

**Contacto:** lbetancourt0733@utm.edu.ec

**Recibido:** 7 de abril de 2025

**Aprobado:** 10 de octubre de 2025

### Resumen

La higiene bucodental continúa siendo un desafío de salud pública, especialmente en determinar la correcta técnica de cepillado durante la limpieza bucal. Frente a esta problemática, las tecnologías emergentes como los sensores inerciales y los algoritmos de inteligencia artificial ofrecen nuevas oportunidades para monitorear la actividad del cepillado dental. En la presente investigación se realizó un análisis empleando señales capturadas durante la ejecución de la actividad de cepillado bucal empleando sensores inerciales (acelerómetro y giroscopio) y modelos de aprendizaje profundo (CNN y LSTM). Se empleó una metodología mixta, de tipo aplicada y experimental. Se trabajó con muestra de 57 participantes mayores de edad, clasificados en dos grupos de acuerdo al correcto cepillado. La recolección de las señales se la realizó mediante sensores MetaMotionR, acoplados a cepillos de uso personal. Los datos fueron procesados mediante técnicas de limpieza, codificación y escalado para su análisis con modelos CNN, LSTM y CNN+LSTM desarrollados en Python, para la recolección de datos se utilizó una aplicación móvil, llamada Metabase, dada por MetaMotionR. La fase cualitativa incluyó observación directa para complementar la interpretación de patrones de comportamiento. Los resultados obtenidos muestran una alta *Recall* (0.9935), *accuracy* (0.9965) y *F1-score* (0.9965), manifestando que un modelo de inteligencia artificial es capaz de distinguir/ clasificar los movimientos correctos e incorrectos. Se concluye que el sistema representa un paso preliminar en la incorporación de IA al monitoreo del cepillado dental. Sin embargo, se identifican limitaciones como la falta de una base de datos con patrones específicos de cepillado o sensores de presión, aspectos que deben abordarse en investigaciones futuras.

**Palabras clave:** *Cepillado dental, CNN-LSTM, Deep learning, Salud bucodental, Sensores inerciales.*

### Abstract

Oral hygiene continues to be a public health challenge, especially in determining the correct brushing technique during oral hygiene. Faced with this problem, emerging technologies such as inertial sensors and artificial intelligence algorithms offer new opportunities to monitor tooth brushing activity. In this research, an analysis was conducted using signals captured during oral brushing using inertial sensors (accelerometer and gyroscope) and deep learning models (CNN and LSTM). A mixed methodology, both applied and experimental, was employed. The study involved a sample of 57 adult participants, classified into two groups based on their correct brushing behavior. Signal collection was performed using MetaMotionR sensors attached to personal toothbrushes. The data were processed using cleaning, encoding, and scaling techniques for analysis with CNN, LSTM, and CNN+LSTM models developed in Python. A mobile application called Metabase, provided by the manufacturer, was used for data collection. The

<https://www.itsup.edu.ec/sinapsis>



qualitative phase included direct observation to complement the interpretation of behavioral patterns. The results obtained show a high Recall (0.9935), Accuracy (0.9965), and F1-score (0.9965), demonstrating that an artificial intelligence model is capable of distinguishing/classifying correct and incorrect movements. It is concluded that the system represents a preliminary step in the incorporation of AI into tooth brushing monitoring. However, limitations are identified, such as the lack of a database with specific brushing patterns or pressure sensors, aspects that should be addressed in future research.

**Keywords:** *Tooth brushing, CNN-LSTM, Deep learning, Oral health, Inertial sensors.*

## Introducción

El mantenimiento de una salud bucodental es parte integral del bienestar y su falta de atención puede resultar en complicaciones que van desde caries a enfermedades periodontales crónicas. La Organización Mundial de la Salud (OMS) estima que aproximadamente 3.500 millones de personas en el mundo sufren alguna enfermedad bucodental, siendo la caries en dientes permanentes la patología más frecuente (Organización Mundial de la Salud, 2022). Al mismo tiempo, reforzar rutinas de prevención, educación e intervención tecnológica sobre higiene bucal cobra mayor importancia en niños y adolescentes.

En países como Ecuador, la situación respecto a la salud bucal refleja avances y desafíos en curso. Los datos históricos indican un declive gradual en la prevalencia de caries dental entre escolares del 88.2% en 1996 al 75.6% en 2009. Sin embargo, estudios más recientes indican que aún existe aproximadamente un 70% de tasa de prevalencia de caries dental entre niños de ocho a diez años (Parise-Vasco et al., 2020) lo que indica que los esfuerzos de políticas públicas continúan siendo inadecuados para abordar este problema. La situación se agrava por la ausencia de información actualizada que permita la toma de decisiones basada en evidencia para el diseño de intervenciones (López-Pérez, 2021). Además, Armas-Vega et al. (2023) examinaron la salud bucodental de escolares de las Islas Galápagos y hallaron que, pese a las campañas preventivas locales, la prevalencia de caries seguía siendo considerablemente alta, lo que subraya la necesidad de reforzar programas de educación y control odontológico continuos en esa población insular.

En este ámbito, el uso de tecnologías emergentes como la inteligencia artificial (IA) y los sensores iniciales ofrecen la oportunidad de romper con los paradigmas existentes en el cuidado bucodental. El uso de sistemas inteligentes ha sido explorado en otros campos de la medicina, como la dermatología o la radiología (Topol, 2019) y en años recientes se comenzó a aprovechar en odontología para el control y evaluación de hábitos de higiene bucal. Incorporar sensores iniciales a cepillos de dientes que registren movimientos y patrones temporalmente permite avanzar hacia métodos personalizados para el aprendizaje y perfeccionamiento de técnicas higiénicas orales.

En un estudio reciente, Chen et al. (2021) alcanzaron una elevada precisión del 99.08% en el reconocimiento de técnicas de cepillado utilizando cepillos de dientes inteligentes equipados con sensores iniciales, lo que se logró utilizando redes neuronales probabilísticas recurrentes (RNN). Aunque este enfoque logró resultados destacables, investigaciones como la presente demuestran que arquitecturas como las redes neuronales convolucionales (CNN) y las de memoria a largo y corto plazo (LSTM) también poseen una sólida capacidad para identificar patrones de cepillado. De hecho, estas arquitecturas han mostrado un desempeño robusto en distintos contextos, lo cual valida su utilidad práctica en sistemas inteligentes de monitoreo de la higiene oral.

Si bien el potencial de la inteligencia artificial en la salud bucodental ha sido ampliamente demostrado en investigaciones internacionales, su adopción práctica en América Latina aún enfrenta obstáculos significativos. En muchas regiones, los sistemas de IA aplicados al diagnóstico o monitoreo del cepillado aún no forman parte de los programas clínicos ni de intervención educativa, debido a la escasa infraestructura tecnológica y a la brecha de formación digital entre profesionales de la salud. Esta situación contrasta con países donde la IA ya se integra en rutinas escolares o clínicas, promoviendo hábitos saludables desde edades tempranas. En este sentido, Lira et al. (2023) destacan que intervenir durante la infancia mediante educación personalizada y continua no solo es más costo-efectivo, sino que también genera beneficios

sostenibles en la calidad de vida bucodental. Por tanto, se requiere no solo validar modelos inteligentes, sino también establecer estrategias de implementación en contextos locales.

A este respecto, Mejía et al. (2022) también mencionan que una técnica deficiente del cepillado puede contribuir a la acumulación de placa bacteriana, dando origen a enfermedad cariosa y gingivitis crónica. Por esta razón, el uso potencial que ofrecen las tecnologías basadas en IA para un diagnóstico temprano y autocorrección constituye un avance importante desde perspectivas pedagógicas y asistenciales.

De esta forma, se parte de una hipótesis: es factible integrar sensores iniciales y modelos de inteligencia artificial y esto tendrá un impacto positivo en la evaluación de las técnicas de cepillado. Este método no solo proporciona eficacia al diagnóstico, sino que abre posibilidades hacia una mayor personalización e implementación estratégica debido a su adaptación al comportamiento del usuario. La IA ha sido aplicada con gran éxito en la clasificación del movimiento humano (Al-Saffar et al., 2020), particularmente en el ámbito de la salud donde su implementación sigue aumentando.

Si bien se han desarrollado IA en algunos ambientes clínicos, su uso rutinario para la evaluación de hábitos, particularmente en poblaciones vulnerables como infantes y adolescentes, está estancado. Por eso, en este caso el propósito general de la investigación es evaluar la capacidad de los modelos de deep learning para la clasificación de movimientos adecuados o inadecuados durante la higiene bucal.

### **Material y Métodos**

Este estudio se llevó a cabo utilizando un enfoque mixto que combina componentes cuantitativos y cualitativos con el fin de analizar de manera integral la técnica de cepillado. Desde la perspectiva cuantitativa, se emplearon los datos capturados mediante los sensores iniciales durante el cepillado dental, empleando métricas para evaluar la precisión y sensibilidad del modelo de inteligencia artificial.

El enfoque cualitativo en este caso permitió explicar patrones de comportamiento asociados con las técnicas de cepillado que se utilizaron y enriqueció el análisis al proporcionar retroalimentación verbal de los participantes monitoreados.

El tipo de investigación fue aplicada y de diseño experimental, orientada a resolver un problema práctico vinculado con la mejora de la salud bucodental mediante el uso de tecnologías automatizadas accesibles. Para el procesamiento y análisis de datos, se utilizó Python 3.10, empleando bibliotecas como Pandas y NumPy para la gestión de datos, la librería Scikit-learn para la evaluación de métricas de clasificación, y TensorFlow/Keras para la implementación y entrenamiento de los modelos de aprendizaje profundo basados en redes neuronales convolucionales (CNN y LSTM). Todo el proceso fue controlado en un entorno experimental, lo que garantizó la replicabilidad y fiabilidad de los resultados obtenidos.

La población objetivo incluía adultos jóvenes en formación o auto-práctica de procedimientos de higiene oral centrándose en la habilidad de control del cepillo. Se identificó que 20 participantes voluntarios no poseían buenos hábitos de limpieza luego de ser validados por el dentista mediante videos de grabación.

La clasificación dentro de cada grupo se estructuró en torno a tener rutinas de cuidado oral adecuadas o inadecuadas. El primer grupo contenía 57 sujetos con prácticas rutinaria normalmente como realizan su cepillado (todo lo que se realiza regularmente). El segundo grupo de los 57 sujetos se elige 14 participantes que práctica el cepillado supervisado por medio de un video educativo (a través de evaluación clínica). Los participantes fueron seleccionados mediante una técnica de muestreo estratificado donde se eligieron individuos en función de comportamientos específicos para asegurar la representación proporcional en el comportamiento relacionado con las prácticas de higiene oral. Esta segmentación fortaleció el entrenamiento del modelo hacia la generalización en lugar de adaptarlo.

Este estudio adoptó el modelo metodológico CRISP-LM-Q (Proceso Estándar de la Industria Cruzada para Máquinas de Aprendizaje con Calidad), una adaptación del modelo tradicional CRISP-DM. Es particularmente útil en investigación aplicada que involucra inteligencia artificial, aprendizaje automático y procesamiento de señales. Las siguientes fases se completaron:

#### **Fase 1: Comprender los datos**

<https://www.itsup.edu.ec/sinapsis>



En esta fase, se realizó una exploración de los datos recolectados de los 57 sujetos durante el proceso de cepillado dental. Los datos provinieron de dos tipos de sensores: acelerómetros y giroscopios, los cuales registraron las señales de movimiento en tres dimensiones: X, Y, y Z. Los sujetos participaron en varias pruebas, denominadas trial\_1, trial\_2, trial\_3 y trial\_4, bajo diferentes condiciones de supervisión, lo que permite analizar su rendimiento tanto en situaciones normales como en condiciones supervisadas.

Los datos de acelerómetro y giroscopio fueron organizados en un *DataFrame*, donde las columnas clave incluyen *epoch* (ms), *time* (-14:00), *elapsed* (s), *x-axis* (g), *y-axis* (g), *z-axis* (g), *subject\_id*, y *trail*. La variable “*subject\_id*” identifica al participante, mientras que *trail* distingue entre las diferentes pruebas realizadas. La columna *elapsed* (s) indica el tiempo transcurrido desde el inicio de la medición, y las columnas de aceleración y giroscopio registran los valores en las tres dimensiones (X, Y, Z) durante cada prueba.

En un análisis preliminar, se identificaron valores nulos en algunas columnas, especialmente en aquellos registros correspondientes a movimientos no representativos del cepillado (por ejemplo, cuando el sensor se desconecta o cuando se detectaron movimientos no deseados). Este resultado llevó a la aplicación de un proceso de limpieza de datos, en el cual se eliminaron los registros incompletos y se corrigieron valores atípicos generados por desconexiones del sensor o movimientos abruptos. La **Figura 1** muestra la ubicación y las coordenadas del sensor en el cepillo de dientes, proporcionando un contexto visual sobre cómo se obtuvo la información de las señales de acelerómetro y giroscopio durante el cepillado.

**Figura 1.** Ubicación y coordenadas del sensor en el cepillo



## Fase 2: Preparación de datos

La fase de preparación de los datos es un paso crucial para asegurar que los datos sean apropiados y estén listos para ser utilizados en el modelado. Durante esta fase, se llevaron a cabo varias actividades para transformar y estructurar los datos brutos recolectados por los sensores de acelerómetro y giroscopio, de manera que puedan ser utilizados en el entrenamiento de los modelos de aprendizaje automático.

### Carga y unión de los datos

El primer paso en la preparación de los datos fue la carga y unión de las señales de acelerómetro y giroscopio para cada uno de los 57 sujetos. Para ello, se implementó una función que permitió leer los archivos CSV correspondientes a cada sujeto y prueba, y combinarlos en un único *DataFrame* (Estructura de datos en la librería *Pandas*). Este *DataFrame* contenía las señales de acelerómetro (en los tres ejes: X, Y, Z) y de giroscopio (también en los tres ejes: X, Y, Z), además de la información adicional, como el *subject\_id* (que identifica al participante) y *trail* (que indica los diferentes ensayos o pruebas por sujeto).

Se verificó que los datos de acelerómetro y giroscopio tuvieran la misma longitud para cada prueba (mediante un proceso de segmentación), garantizando que ambos conjuntos de datos estuvieran sincronizados en cuanto a las muestras registradas.

### Segmentación por Ventanas

Una vez cargados y combinados los datos de todos los sujetos, el siguiente paso fue la segmentación de las señales en ventanas de 2 segundos y 3 segundos. La cantidad de muestras de cada ventana se definió en función de la frecuencia de muestreo de los sensores, que era de 50 Hz. Esto significa que cada ventana de datos contiene 100 muestras (50 muestras por segundo durante 2 segundos). Además, se utilizó un 50% de solapamiento entre las ventanas para mejorar la captura de patrones continuos en los movimientos durante el cepillado.

La segmentación permitió transformar el conjunto de datos continuo en una serie de ventanas temporales que se utilizarían como entradas para los modelos de aprendizaje automático. Al aplicar este método, se generaron un total de 10776 ventanas de datos para todos los sujetos, las cuales fueron etiquetadas con la actividad correspondiente (normal o supervisada).

### Normalización de los Datos

Una parte clave de la preparación de los datos fue la normalización, dado que las señales de acelerómetro y giroscopio pueden tener diferentes rangos y unidades de medida (por ejemplo, la aceleración se mide en g y la rotación en deg/s), era necesario normalizar los datos para asegurar que todas las características estuvieran en la misma escala.

Se utilizó el *StandardScaler* de *sklearn* para normalizar las características de cada ventana de datos. Esto evitó que las diferencias en las escalas de las características afectaran negativamente el desempeño de los modelos. La normalización por característica es esencial para el rendimiento de los algoritmos de aprendizaje automático, especialmente en redes neuronales y otros modelos sensibles a la escala de los datos.

### Generación del conjunto de datos

El resultado de la segmentación y normalización fue un conjunto de datos estructurado y listo para el modelado. Se generaron dos matrices principales: X y Y. La matriz X contiene las ventanas de datos normalizadas, con una forma de (10776, 100, 6), donde 10776 es el número de ventanas generadas, 100 es el tamaño de cada ventana (equivalente a 2 segundos de datos), y 6 es el número de características (tres para el acelerómetro y tres para el giroscopio). La matriz Y contiene las etiquetas correspondientes a cada ventana de datos, con una forma de (10776), donde cada valor indica si la ventana corresponde a una prueba normal o supervisada.

### Fase 3: Modelado

#### Modelo CNN

El modelo de *Convolutional Neural Network* (CNN) se entrenó utilizando datos segmentados en ventanas de 2 segundos y normalizados para las señales de acelerómetro y giroscopio. El modelo mostró un buen desempeño durante un entrenamiento inicial, las primeras épocas alcanzaron una *accuracy* de 85.19% de entrenamiento. A medida que el modelo avanzaba en el número de épocas, la *accuracy* de la validación aumentó significativamente, alcanzando un valor máximo de 99.98% en la época 150, lo que indica que el modelo fue capaz de generalizar bien en los datos de prueba. En términos de pérdida, el modelo mostró una reducción considerable desde el inicio del entrenamiento, pasando de una pérdida de 0.3337 a 0.0009 en la última época, lo que demuestra una convergencia estable y una buena optimización durante el entrenamiento. Esto se reflejó en las métricas de validación, con la *accuracy* de validación alcanzando un 99.12%.

El modelo también presentó un buen desempeño en cuanto a la reducción de la pérdida de clasificación, manteniendo un comportamiento estable a lo largo de las épocas. Sin embargo, se observó cierta variabilidad en las métricas de validación hacia las últimas épocas.

#### Modelo LSTM

El modelo LSTM (*Long Short-Term Memory*) fue entrenado utilizando los datos segmentados en ventanas de 2 segundos y normalizados para las señales de acelerómetro y giroscopio. Este modelo está diseñado para aprender secuencias temporales, lo que lo hace adecuado para capturar las dinámicas temporales de los datos de movimiento durante el cepillado.

El modelo mostró un desempeño robusto durante las épocas de entrenamiento, alcanzando una *accuracy* de 83.15% en la primera época. A medida que avanzaron las épocas, la *accuracy* de validación aumentó significativamente, llegando a un máximo de 100.00% en la época 150. Este resultado indica que el modelo fue capaz de generalizar bien a los datos de validación y aprender patrones útiles de las señales temporales.

En términos de pérdida, el modelo comenzó con una pérdida de 0.4517 en la primera época, y a lo largo del entrenamiento, esta disminuyó considerablemente hasta llegar a 0.0001 en la época 150, mostrando una convergencia estable. La *accuracy* de validación, que fluctuó entre el 91.09% y 98.28%, refleja que el modelo estuvo constantemente mejorando su capacidad de clasificación a medida que avanzaba el entrenamiento. Este comportamiento indica que el modelo fue capaz de aprender y generalizar de manera efectiva los patrones del cepillado, aunque podría

beneficiarse de algunos ajustes, como regularización adicional o más datos, para mejorar la estabilidad.

#### Modelo combinado CNN + LSTM

El modelo combinado de CNN + LSTM fue entrenado para capturar tanto las características locales (usando la capa Conv1D) como las dependencias temporales (con las capas LSTM) en las señales de acelerómetro y giroscopio. Este enfoque híbrido aprovechó lo mejor de ambos mundos, permitiendo que el modelo aprendiera patrones espaciales a nivel de ventana temporal (CNN) y también capturara la dinámica temporal de los movimientos del cepillo a través de las secuencias (LSTM).

El modelo CNN + LSTM consistió en una capa convolucional 1D seguida de un MaxPooling, lo que permitió extraer características locales de las señales. Posteriormente, se incorporaron capas LSTM para modelar las relaciones secuenciales en los datos. Finalmente, se añadió una capa densa con salida softmax, adecuada para tareas de clasificación binaria, como la detección de cepillado normal o supervisada. La arquitectura completa tuvo un total de 80.386 parámetros entrenables.

Durante el entrenamiento, el modelo mostró una *accuracy* inicial de 86.32% en la época 1, y a medida que avanzaban las épocas, la *accuracy* aumentó significativamente. Después de 139 épocas de entrenamiento, la *accuracy* alcanzó un 100.0%, lo que indica que el modelo fue capaz de aprender los patrones de movimiento del cepillo con gran exactitud. La pérdida de entrenamiento también disminuyó progresivamente, pasando de 0.3275 en la época 1 a 0.0005 en la época 139, lo que refleja la optimización continua del modelo.

En cuanto a la *accuracy* de validación, el modelo también mostró un excelente desempeño, alcanzando un valor máximo de 99.35% en la época 143. Aunque en algunas épocas se observó una ligera caída en la *accuracy* de validación, el rendimiento general siguió siendo muy alto, lo que sugiere que el modelo estaba aprendiendo correctamente los patrones de los datos.

Los valores de *accuracy* y pérdida obtenidos durante el entrenamiento son muy prometedores, lo que indica que el modelo combinado CNN + LSTM logró aprender los patrones tanto espaciales como temporales de las señales de acelerómetro y giroscopio. La alta *accuracy* en el conjunto de validación también sugiere que el modelo es capaz de generalizar bien a datos nuevos.

Aunque hubo algunas fluctuaciones menores en la *accuracy* de validación en ciertas épocas, el modelo mostró un comportamiento estable y convergió a un rendimiento excelente al final del entrenamiento. Esto es indicativo de que el modelo es muy efectivo para clasificar correctamente el cepillado normal y supervisado, basado en las señales de los sensores.

La fase de implementación y evaluación del modelo: El rendimiento del sistema se evaluó utilizando métricas cuantitativas incluyendo *precision*, *recall*, *accuracy*, *F1-score*, especificidad con ajuste de hiperparámetros orientado a maximizar la capacidad predictiva. La evaluación del experimento se realizó una comparación de métricas de los modelos.

Entre las estrategias más notables implementadas está la observación directa, por medio de la cual se pudieron registrar cualitativamente el comportamiento de los sujetos durante la actividad de cepillado dental. La observación fue complementada con una captura de datos, considerada esencial, que consistió en un dispositivo portátil montado sobre los cepillos capaces de medir aceleraciones y giros. Además, se realizó video grabación y/o fotos durante las sesiones de cepillado lo cual posibilitó una posterior revisión sobre cada usuario asegurando que existía coincidencia entre los datos automáticos y las imágenes obteniendo veracidad.

En cuanto a los dispositivos utilizados, se escogió el sensor inercial de MetaMotionR debido a su bajo consumo energético, uso de energía móvil, y precisión (MbientLab, 2024). Con el fin de facilitar su uso y no limitar la experiencia del cepillado dental, estos sensores fueron adheridos a cepillos dentales estándar. Para el procesamiento de los datos, se llevaron a cabo trabajos en Python 3.10 donde scripts fueron creados para procesar información con ayuda de bibliotecas especializadas como pandas y numpy en manipulación de datos, scikit-learn para tareas de machine learning, así como tensorflow para la creación de modelos basados en redes neuronales profundas como *Convolutional Neural Networks* (CNN) o *Long Short-Term Memory* (LSTM). La aplicación Metabase fue utilizado para la recolección de los valores del sensor, lo cual generó un archivo en formato csv.

La evaluación del rendimiento del modelo de clasificación y la veracidad de los patrones detectados requería realizar un análisis estadístico que se condujera con los datos obtenidos.

Para evaluar el rendimiento del modelo de inteligencia artificial desarrollado, se utilizaron cuatro métricas: *precision* (ver Formula 1), *recall* (ver Formula 2), *accuracy* (ver Formula 3) y *F1-Score* (ver Formula 4). La precisión se calculó como la proporción de verdaderos positivos sobre el total de predicciones positivas, es decir, la capacidad del modelo para clasificar correctamente los cepillados apropiados entre todos los casos identificados como tales. Por su parte, la sensibilidad (también conocida como *recall*) se definió como la proporción de verdaderos positivos respecto al total de casos realmente positivos, lo que permite medir qué tan eficazmente el modelo identifica los cepillados correctos entre todos los que realmente lo son. La exactitud (conocida como *Accuracy*) mide el porcentaje, sumando las predicciones verdaderas positivas sobre la suma de todas las predicciones del caso evaluados. F1-score calcula el rendimiento del producto entre precisión y sensibilidad sobre la suma de las mismas métricas. Estos cálculos se realizaron a partir de la matriz de confusión generada durante la etapa de validación del modelo. Se efectuó un control estadístico mediante un nivel de significancia del 5% ( $p < 0.05$ ), con el fin de garantizar la robustez metodológica y la validez de los resultados obtenidos.

Formula 1. Precisión (*Precision*)

$$Precision = \frac{VP}{VP + FP}$$

**Nota:** ¿Qué tan confiable es el modelo cuando dice que un cepillado fue adecuado?

Se define como la proporción de verdaderos positivos (VP) frente al total de predicciones positivas (VP + FP). Esta métrica responde a la pregunta: ¿Qué tan confiable es el modelo cuando predice que un cepillado fue adecuado?

Formula 2. Sensibilidad (*Recall*)

$$Sensibilidad = \frac{VP}{VP + FN}$$

**Nota:** ¿Qué tan bueno es el modelo para encontrar todos los cepillados realmente adecuados?

Representa la proporción de verdaderos positivos frente al total de casos realmente positivos (VP + FN). Evalúa la capacidad del modelo para identificar todos los cepillados correctos. ¿Qué tan bueno es el modelo para encontrar todos los cepillados realmente adecuados?

Formula 3. Exactitud (*Accuracy*)

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

**Nota:** ¿Qué tan efectivo es el modelo para clasificar adecuadamente tanto los casos positivos como los negativos?

Es la relación entre las predicciones acertadas (verdaderos positivos y verdaderos negativos) y el total de casos analizados. Evalúa el rendimiento general del modelo para determinar correctamente todas las clases que están presentes.

Formula 4. *F1-score*

$$F1 - score = \frac{Precisión \times Recall}{Precisión + Recall} \times (2)$$

**Nota:** ¿Cuán efectivo es el modelo para identificar correctamente los casos positivos, sin cometer muchos errores y manteniendo un balance entre precisión y sensibilidad?

Expresa la media armónica entre la sensibilidad y la precisión, proporcionando un único valor que sintetiza el equilibrio entre las dos. Es útil cuando las clases están desproporcionadas y se busca un balance entre detectar los positivos y prevenir los falsos positivos.

Esta investigación todos los participantes dieron consentimiento informado. La confianza y anonimato de los datos personales se mantuvo mediante la anonimización de registros, así como la extracción intelectual de la información a través de seudónimos para garantizar la confidencialidad. Además, se garantizó que la participación en la investigación no implicara riesgo físico o psicológico, y que los sensores utilizados no interferirían con su cuidado durante las sesiones de cepillado.

## Resultados

## Evaluación del modelo con señal Acelerómetro y ventanas de 2 segundos, 50% solapamiento.

### Modelo CNN

Durante la fase de evaluación del modelo *Convolutional Neural Network* (CNN) aplicado al análisis bucodental mediante sensores iniciales, se obtuvieron resultados que evidencian la capacidad del modelo para distinguir entre actividades "normal" y "supervisada". La matriz de confusión muestra que de un total de 768 instancias reales clasificadas como "normal", el modelo identificó correctamente 762 y cometió 6 errores. En el caso de la clase "supervisada", de 394 instancias, 390 fueron correctamente clasificadas y solo 4 fueron mal etiquetadas. Esta distribución indica un bajo índice de error y una alta efectividad en ambas clases.

Clase	Precision	Recall	F1-score	Accuracy	Soporte
normal	0.9948	0.9922	0.9935	-	768
supervisada	0.9848	0.9898	0.9873	-	394
Weighted avg	0.9914	0.9914	0.9912	0.9912	1162

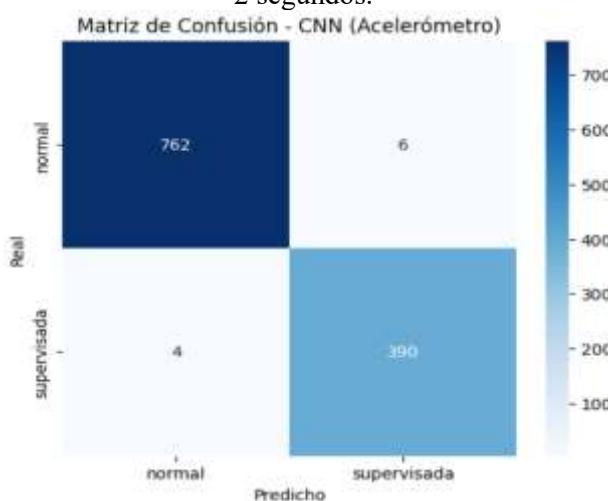
En cuanto a las métricas de evaluación, el modelo alcanzó una precisión del 99.48% para la clase normal y del 98.48% para la clase "supervisada", lo que significa que prácticamente todas las predicciones positivas realizadas para cada clase fueron correctas. Asimismo, el recall fue del 99.22% en "normal" y del 98.98% en "supervisada", demostrando una gran capacidad para recuperar casi todas las instancias verdaderas de cada categoría. El *F1-score*, que combina *precision* y *recall*, fue de 99.35% para "normal" y de 98.73% para "supervisada", ratificando la solidez del desempeño.

Además, el *accuracy* global del modelo fue del 99.14%, reflejando que casi la totalidad de las predicciones realizadas por el modelo fueron correctas. El promedio ponderado de las métricas es calculado con la **fórmula 5**, para *precision*, *recall* y *F1-score* también se situaron en 99.14%, lo que muestra que el modelo mantiene un rendimiento equilibrado incluso cuando las clases tienen diferente número de muestras.

**Formula 5.** Promedio ponderado (Weighted average)

$$\text{weighted avg} = \frac{(\sum i \text{ metric } i \times \text{support } i)_{\text{normal}} + (\sum i \text{ metric } i \times \text{support } i)_{\text{supervisada}}}{\sum i \text{ support } i_{(\text{normal, supervisada})}}$$

**Figura 1.** Matriz de Confusión, modelo CNN, empleando la señal del acelerómetro, ventanas de 2 segundos.



### Modelo LSTM

Las métricas de evaluación del modelo *Long Short-Term Memory* (LSTM) evidencian un rendimiento muy variable, el modelo tuvo una *precision* del 96.18% normal y 96.55% supervisada, mientras que *recall* 98.31% normal y 92.39% supervisada, *F1-score* para la clase

normal 97.23%, mientras que la clase “supervisada” alcanzó 94.42%, reflejando una ligera disminución en sensibilidad sin afectar significativamente la exactitud general. El modelo logró un *accuracy* total del 96.30%, lo que confirma su eficacia es menor para la clasificación de actividades bucodentales a partir de datos secuenciales de sensores iniciales.

**Tabla 2.** Evaluación del modelo LSTM

Clase	Precision	Recall	F1-score	Accuracy	Soporte
normal	0.9618	0.9831	0.9723	-	768
supervisada	0.9655	0.9239	0.9442	-	394
Weighted avg	0.9630	0.9630	0.9628	0.9429	1162

La matriz de confusión del modelo LSTM muestra un rendimiento muy sólido, con 755 clasificaciones correctas para la clase “normal” y 364 para la clase “supervisada”. Se cometieron 13 errores al predecir como “supervisada” muestras normales y 30 al predecir como “normal” las muestras supervisadas. Estas cifras se reflejan en métricas de medio nivel manteniendo un equilibrio del promedio ponderado entre las métricas.

**Figura 2.** Matriz de confusión del modelo LSTM, empleando la señal del acelerómetro, ventanas de 2 segundos.



### Modelo CNN + LSTM

El modelo combinado CNN + LSTM obtuvo resultados sobresalientes, alcanzando un *accuracy* del 98.97%, lo que indica una clasificación precisa de los datos bucodentales recogidos mediante sensores iniciales. La clase “normal” alcanzó un *F1-score* de 99.22% y la clase “supervisada” un 98.48%, con una precisión de 99.48% normal y 97.99% supervisada y *recall* con valores muy equilibrado de 98.98%. Estas métricas reflejan que la combinación de la extracción espacial de CNN con la modelación secuencial de LSTM permite una clasificación robusta, generalizando eficazmente sobre los datos de entrada. El promedio ponderado se mantiene en 98.97% en todas las métricas respectivamente, consolidando a esta arquitectura como la más precisa entre los modelos evaluados.

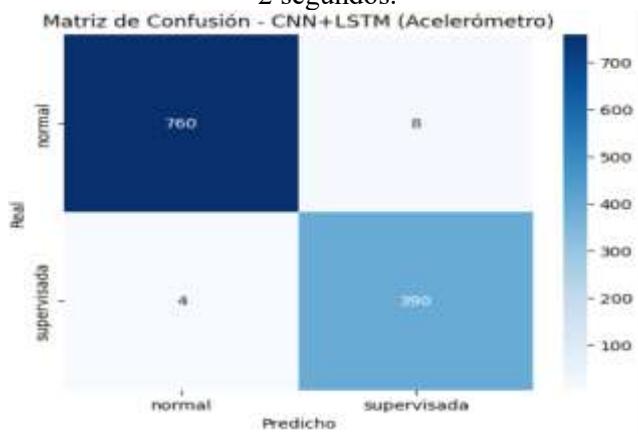
**Tabla 3.** Evaluación CNN + LSTM

Clase	Precision	Recall	F1-score	Accuracy	Soporte
normal	0.9948	0.9896	0.9922	-	768
supervisada	0.9799	0.9898	0.9848	-	394
Weighted avg	0.9897	0.9897	0.9897	0.9897	1162

La matriz de confusión del modelo combinado CNN + LSTM refleja un rendimiento excepcional en la clasificación de actividades bucodentales. De las 768 muestras reales clasificadas como “normal”, 760 fueron correctamente identificadas y solo 8 fueron clasificadas erróneamente como “supervisada”. En cuanto a la clase “supervisada”, de un total de 394 muestras, 390 fueron correctamente clasificadas y únicamente 4 resultaron mal etiquetadas. Estos valores confirman la efectividad del modelo híbrido, el cual combina la capacidad de detección espacial de CNN con

la sensibilidad secuencial de LSTM, logrando minimizar los errores de clasificación en ambas categorías.

**Figura 3.** Matriz de confusión CNN + LSTM, empleando la señal del acelerómetro, ventanas de 2 segundos.



**Tabla 4.** Resultados empleando la señal del acelerómetro, ventanas de 2 segundos.

Modelo	Accuracy	Precisión	Recall	F1-score
CNN	0.9914	0.9914	0.9914	0.9914
LSTM	0.9630	0.9630	0.9630	0.9628
CNN+LSTM	0.9897	0.9897	0.9897	0.9897

Evaluación del modelo con señal Giroscopio y ventanas de 2 segundos, 50% solapamiento.

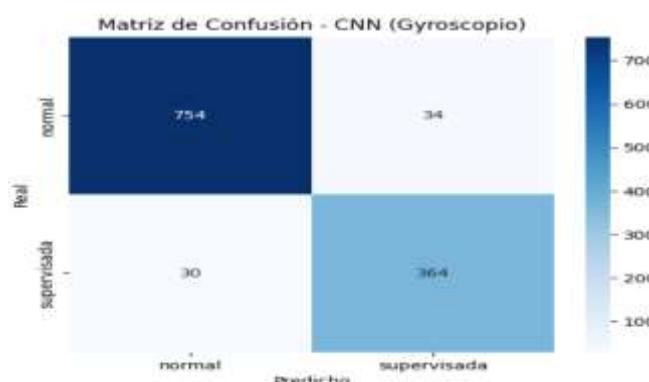
#### Modelo CNN

Se obtuvo resultados no muy favorables, de las 788 muestras reales clasificadas como “normal”, 754 fueron correctamente identificadas y solo 34 fueron clasificadas erróneamente como “supervisada”. En cuanto a la clase “supervisada”, de un total de 394 muestras, 364 fueron correctamente clasificadas y únicamente 30 resultaron mal etiquetadas. La clase “normal” obtuvo un *F1-score* de 95.93%, mientras que la clase “supervisada” alcanzó 91.92%, el *Recall* en la clase “normal” se obtuvo 95.69%, mientras en la clase “supervisada” se obtuvo un valor favorable de 92.39%, en *accuracy* total del 94.59%.

**Tabla 5.** Evaluación CNN, empleando la señal del giroscopio, ventanas de 2 segundos.

Clase	Precision	Recall	F1-score	Accuracy	Soporte
normal	0.9617	0.9569	0.9593	-	788
supervisada	0.9146	0.9239	0.9192	-	394
Weighted avg	0.9460	0.9459	0.9459	0.9459	1182

**Figura 4.** Matriz de confusión CNN, empleando la señal del giroscopio, ventanas de 2 segundos.



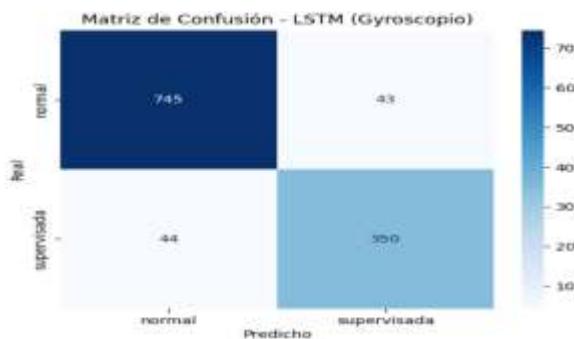
### Modelo LSTM

El modelo *Long Short-Term Memory* (LSTM) tuvo un desempeño inferior en todas las métricas, lo que sugiere que, al trabajar solo con señales del acelerómetro, el modelo basado en memoria secuencial pierde efectividad frente a arquitecturas convolucionales o mixtas. Esto podría atribuirse a la menor complejidad temporal de las señales cuando no están enriquecidas con giroscopio.

**Tabla 6.** Evaluación LSTM, empleando la señal del giroscopio, ventanas de 2 segundos.

Clase	Precision	Recall	F1-score	Accuracy	Soporte
normal	0.9442	0.9454	0.9448	-	788
supervisada	0.8906	0.8883	0.8895	-	394
Weighted avg	0.9264	0.9264	0.9264	0.9126	1182

**Figura 5.** Matriz de confusión LSTM, empleando la señal del giroscopio, ventanas de 2 segundos.



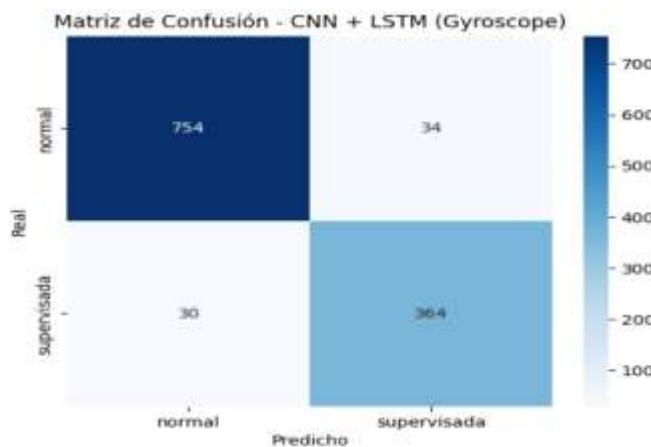
### Modelo CNN + LSTM

El acelerómetro por sí solo proporciona una base suficiente para una clasificación precisa, especialmente con modelos CNN o CNN+LSTM. Sin embargo, estos resultados confirman que el acelerómetro es una fuente válida y eficiente para la detección de patrones en tareas de cepillado dental supervisado vs. normal.

**Tabla 7.** Evaluación CNN + LSTM, empleando la señal del giroscopio, ventanas de 2 segundos.

Clase	Precision	Recall	F1-score	Accuracy	Soporte
normal	0.9617	0.9569	0.9593	-	788
supervisada	0.9146	0.9239	0.9192	-	394
Weighted avg	0.9460	0.9459	0.9459	0.9459	1182

**Figura 6.** Matriz de confusión CNN + LSTM, empleando la señal del giroscopio, ventanas de 2 segundos.



**Tabla 8.** Resultados empleando la señal del giroscopio, ventanas de 2 segundos.

Modelo	Accuracy	Precisión	Recall	F1-score
CNN	0.9459	0.9460	0.9459	0.9459
LSTM	0.9264	0.9264	0.9264	0.9264
CNN+LSTM	<b>0.9662</b>	<b>0.9666</b>	<b>0.9662</b>	<b>0.9663</b>

#### Evaluación del modelo con señal Giroscopio + Acelerómetro con ventanas de 2 segundos.

##### Modelo CNN

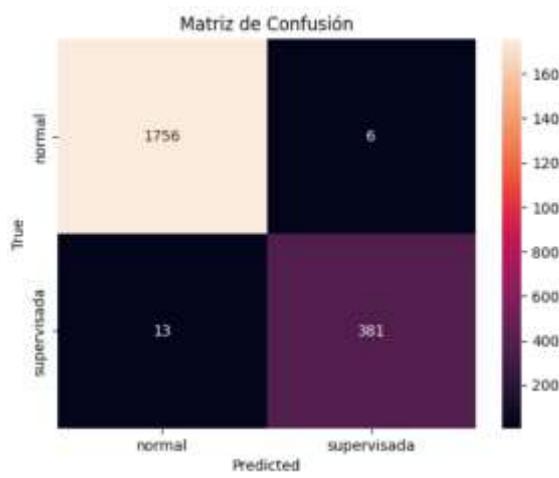
El modelo *Convolutional Neural Network* (CNN) se obtuvieron resultados satisfactorios que evidencian la capacidad del modelo para distinguir entre actividades "normal" y "supervisada". La matriz de confusión muestra que de un total de 1762 instancias reales clasificadas como "normal", el modelo identificó correctamente 1756 y 6 errores. En el caso de la clase "supervisada", de 394 instancias, 381 fueron correctamente clasificadas y solo 13 fueron etiquetadas erróneas. Esta distribución indica un bajo índice de error y una alta efectividad en ambas clases, para diferenciar de forma efectiva entre ambas categorías a partir de los datos de sensores iniciales.

El modelo CNN ha demostrado ser una herramienta altamente eficaz y confiable para la clasificación de actividades bucodentales a partir de datos de sensores iniciales. Su capacidad para identificar patrones complejos con bajo margen de error lo convierte en una opción sólida para aplicaciones prácticas en contextos de clasificación.

**Tabla 9.** Evaluación del modelo CNN, empleando la señal combinada del acelerómetro y del giroscopio, ventanas de 2 segundos.

Clase	Precision	Recall	F1-score	Accuracy	Soporte
normal	0.9927	0.9966	0.9946	-	1762
supervisada	0.9845	0.9670	0.9757	-	394
Weighted avg	0.9912	0.9912	0.9912	0.9912	2156

**Figura 7.** Matriz de confusión CNN, empleando la señal combinada del acelerómetro y del giroscopio, ventanas de 2 segundos.



##### Modelo LSTM

Las métricas evidencian un rendimiento promedio confiable, con valores equilibrados entre *precision*, *recall* y *F1-score* para ambas clases, *F1-score* de 98.69% en clase normal, reflejando una ligera disminución en recall 98.64%, sin afectar significativamente la *precision* en general. El modelo logró un accuracy total del 97.87%, lo que confirma su eficacia para la clasificación de actividades bucodentales a partir de datos secuenciales de sensores iniciales.

**Tabla 10.** Evaluación del modelo LSTM, empleando la señal combinada del acelerómetro y del giroscopio, ventanas de 2 segundos.

Clase	Precision	Recall	F1-score	Accuracy	Soporte
normal	0.9875	0.9864	0.9869	-	1762
supervisada	0.9394	0.9442	0.9418	-	394
Weighted avg	0.9787	0.9787	0.9787	0.9787	2156

La matriz de confusión del modelo LSTM muestra un rendimiento muy sólido, con 1738 clasificaciones correctas para la clase “normal” y 372 para la clase “supervisada”. Se cometieron 24 errores al predecir como “supervisada” muestras normales y 22 al predecir como “normal” las muestras supervisadas. Estas cifras se reflejan en métricas el nivel: *precision* del 98.75% para “normal” y 93.94% para “supervisada”, con un *F1-score* promedio ponderado de 97.87%. En conjunto, el modelo demuestra una capacidad efectiva para distinguir entre ambas clases con un *accuracy* general del 97.87%, manteniendo un equilibrio adecuado entre precisión y sensibilidad.

**Figura 8.** Matriz de confusión LSTM, empleando la señal combinada del acelerómetro y del giroscopio, ventanas de 2 segundos.



### Modelo CNN + LSTM

Con este se obtuvo resultados sobresalientes, alcanzando un *accuracy* del 99.35%, lo que indica una clasificación altamente precisa de los datos bucodentales recogidos mediante sensores inerciales. La clase “normal” alcanzó un *F1-score* de 99,60% y la clase “supervisada” un 98.24%, la *precision* en la clase normal obtuvo 99.83% y supervisada 97.26%, mientras que *recall* en la clase normal 99.38% y supervisado 99.24%. Estas métricas reflejan que la combinación de la extracción espacial de CNN con la modelación secuencial de LSTM permite una clasificación robusta, generalizando eficazmente sobre los datos de entrada. El promedio ponderado se mantiene en 99.35% y 99.36% respectivamente, consolidando a esta arquitectura como la más precisa entre los modelos evaluados.

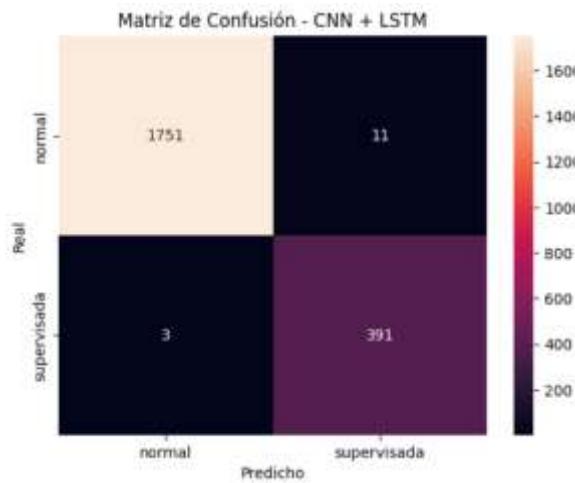
**Tabla 11.** Evaluación CNN + LSTM, empleando la señal combinada del acelerómetro y del giroscopio, ventanas de 2 segundos.

Clase	Precision	Recall	F1-score	Accuracy	Soporte
normal	0.9983	0.9938	0.9960	-	1762
supervisada	0.9726	0.9924	0.9824	-	394
Weighted avg	0.9936	0.9935	0.9935	0.9855	2156

La matriz de confusión del modelo combinado CNN + LSTM refleja un rendimiento excepcional en la clasificación de actividades bucodentales. De las 1762 muestras reales clasificadas como “normal”, 1751 fueron correctamente identificadas y solo 11 fueron clasificadas erróneamente como “supervisada”. En cuanto a la clase “supervisada”, de un total de 394 muestras, 391 fueron

correctamente clasificadas y únicamente 3 resultaron mal etiquetadas. Estos valores confirman la efectividad del modelo híbrido, el cual combina la capacidad de detección espacial de CNN con la sensibilidad secuencial de LSTM, logrando minimizar los errores de clasificación en ambas categorías.

**Figura 9.** Matriz de confusión LSTM, empleando la señal combinada del acelerómetro y del giroscopio, ventanas de 2 segundos.



**Tabla 12.** Resultados empleando la señal combinada del acelerómetro y del giroscopio, ventanas de 2 segundos.

Modelo	Accuracy	Precisión	Recall	F1-score
CNN	0.9912	0.9912	0.9912	0.9912
LSTM	0.9787	0.9787	0.9787	0.9787
<b>CNN + LSTM</b>	<b>0.9935</b>	<b>0.9936</b>	<b>0.9935</b>	<b>0.9935</b>

**Resumen de la evaluación de los resultados con ventanas de 2 segundo y 50% solapamiento.** La comparación de las métricas de rendimiento entre los modelos CNN, LSTM y la arquitectura combinada CNN+LSTM revela que este último ofrece el mejor desempeño global utilizando la combinación de las señales accelerometer + gyroscope.

**Tabla 13.** Comparación de resultado de los modelos de 2 segundo (100 muestra)

Señales	Accelerometer			Gyroscope			Accelerometer +Gyroscope			
	Algoritmo	Accuracy	Recall	F1-score	Accuracy	Recall	F1-score	Accuracy	Recall	F1-score
CNN	0.9914	0.9914	0.9914	0.9459	0.9459	0.9459	0.9459	0.9912	0.9912	0.9912
LSTM	0.9630	0.9630	0.9628	0.9264	0.9264	0.9264	0.9264	0.9787	0.9787	0.9787
CNN+LSTM	0.9897	0.9897	0.9897	0.9662	0.9662	0.9663	0.9663	0.9935	0.9935	0.9935

#### Evaluación del modelo con señal Giroscopio + Acelerómetro con ventana de 3 segundo.

##### Modelo CNN

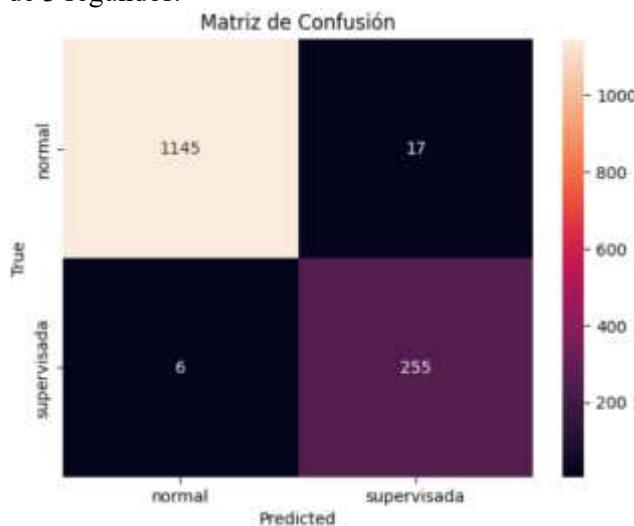
El modelo logró resultados satisfactorios que demuestran su habilidad para diferenciar entre las actividades "normal" y "supervisada". La matriz de confusión figura 10 revela que, de 1162 casos reales clasificados como "normal", el modelo identificó correctamente 1145 y cometió 17 errores. De 261 instancias de la clase "supervisada", se clasificaron correctamente 255 y solo 6 fueron etiquetadas incorrectamente. Esta distribución muestra una elevada eficacia en las dos categorías y un bajo índice de error.

El modelo CNN ha probado ser un instrumento sumamente confiable y eficaz para clasificar actividades bucodentales a partir de datos obtenidos por sensores inerciales. Su habilidad para detectar patrones complejos con un margen de error mínimo lo hace una opción sólida para aplicaciones prácticas en escenarios clínicos.

**Tabla 14.** Resultado de CNN, empleando la señal combinada del acelerómetro y del giroscopio, ventanas de 3 segundos.

Clase	Precision	Recall	F1-score	Accuracy	Soporte
normal	0.9948	0.9854	0.9918	-	1162
supervisada	0.9375	0.9770	0.9568	-	261
Weighted avg	0.9843	0.9838	0.9840	0.9838	1423

**Figura 10.** Matriz de confusión CNN, empleando la señal combinada del acelerómetro y del giroscopio, ventanas de 3 segundos.



### Modelo LSTM

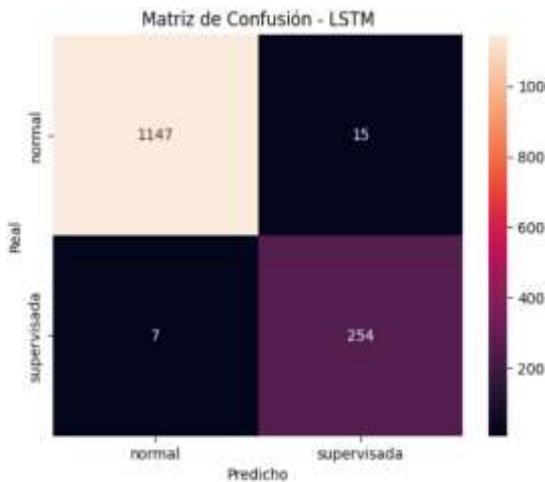
Las métricas muestran un desempeño no confiable, con cifras muy variante entre la *precision*, *recall* y *F1-score* para las dos clases. La clase "normal" logró un F1-score de 99.05%, en tanto que la clase "supervisada" llegó a 95.85%; este modelo mostró reducción en la precisión teniendo un impacto notable en la *accuracy* en general. El modelo alcanzó una *precision* total del 98.45%, lo que demuestra meno eficiencia en la clasificación de actividades bucodentales usando datos secuenciales de sensores iniciales.

**Tabla 15.** Resultado de LSTM, empleando la señal combinada del acelerómetro y del giroscopio, ventanas de 3 segundos.

Clase	Precision	Recall	F1-score	Accuracy	Soporte
normal	0.9939	0.9871	0.9905	-	1162
supervisada	0.9442	0.9732	0.9585	-	261
Weighted avg	0.9848	0.9845	0.9846	0.9845	1423

El desempeño del modelo LSTM, según su matriz de confusión, es muy sólido, con 1162 clasificaciones precisas para la categoría "normal" y 261 para la "supervisada". Se cometieron 15 errores al pronosticar las muestras supervisadas como "normales" y 7 al pronosticar las muestras normales como "supervisadas". Estas cifras se ven en métricas de alto nivel: una *precision* del 99.39% para "normal" y del 94.42% para "supervisada", así como un *F1-score* promedio del 98.46%. En resumen, el modelo muestra una *accuracy* del 98,45%.

**Figura 11.** Matriz de confusión LSTM, empleando la señal combinada del acelerómetro y del giroscopio, ventanas de 3 segundos.



### Modelo CNN + LSTM

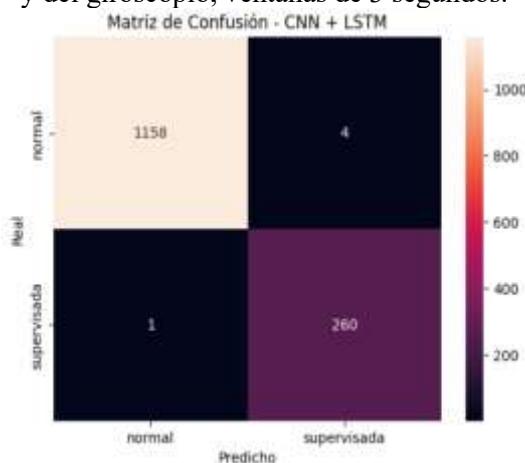
Se lograron resultados excepcionales con este, llegando a un 99.65% de *accuracy* (exactitud), lo que significa una clasificación muy precisa de los datos bucodentales recolectados a través de sensores inerciales. La clase "normal" obtuvo un puntaje *F1-score* de 99.78% y la clase "supervisada" uno de 99.05%, con cifras muy balanceada en *recall* en las dos clases. Estos parámetros demuestran que la conjunción de la extracción espacial de CNN + LSTM posibilita una clasificación sólida, que generaliza efectivamente a partir de los datos iniciales. El promedio macro y el ponderado se sostienen en 99.65%, lo que confirma a esta arquitectura como la más exacta entre los modelos analizados.

**Tabla 16.** Resultado de CNN + LSTM, empleando la señal combinada del acelerómetro y del giroscopio, ventanas de 3 segundos.

Clase	Precision	Recall	F1-score	Accuracy	Soporte
normal	0.9991	0.9966	0.9978	-	1162
supervisada	0.9848	0.9962	0.9905	-	261
Weighted avg	0.9965	0.9965	0.9965	0.9965	1423

El modelo combinado CNN + LSTM muestra un desempeño sobresaliente en la clasificación de actividades bucodentales, según lo indica su matriz de confusión. De un total de 1162 muestras reales, clasificadas como "normal", 1158 fueron identificadas correctamente y 4 fueron clasificadas equivocadamente como "supervisada". Con respecto a la clase "supervisada" de 261 muestras en total, 260 fueron clasificadas adecuadamente y solo 1 se etiquetaron incorrectamente. La eficacia del modelo híbrido, que mezcla la habilidad de detección espacial de las CNN con la sensibilidad secuencial de las LSTM, se confirma con estos valores, ya que consigue reducir al mínimo los errores de clasificación en los dos grupos.

Figura 12. Matriz de confusión CNN + LSTM, empleando la señal combinada del acelerómetro y del giroscopio, ventanas de 3 segundos.



**Tabla 17.** Resultado de gyroscope + accelerometer, empleando la señal combinada del acelerómetro y del giroscopio, ventanas de 3 segundos.

Modelo	Accuracy	Precisión	Recall	F1-score
CNN	0.9838	0.9843	0.9838	0.9840
LSTM	0.9845	0.9848	0.9845	0.9846
CNN + LSTM	<b>0.9965</b>	<b>0.9965</b>	<b>0.9965</b>	<b>0.9965</b>

**Resumen de la evaluación de los resultados con ventanas de 2 segundo y 3 segundo, 50% solapamiento.**

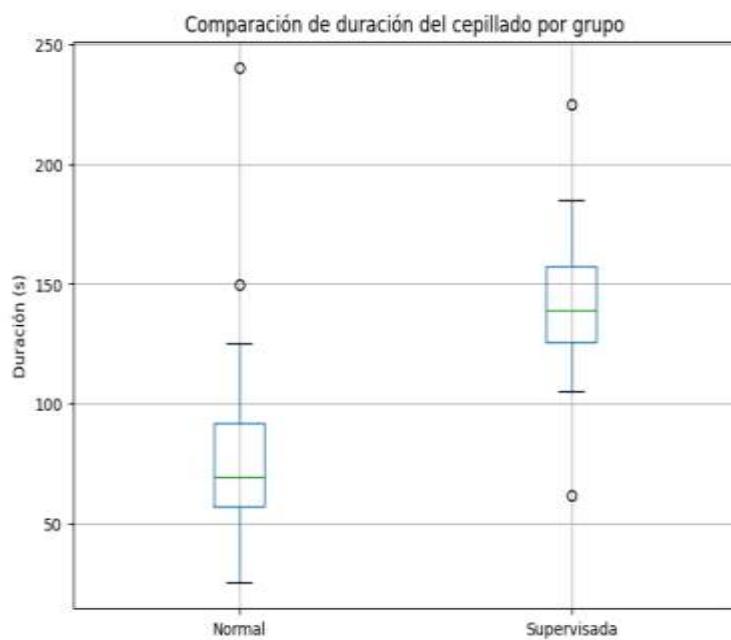
La comparación de las métricas de rendimiento entre los modelos CNN, LSTM y la combinada CNN+LSTM, en la tabla 18 detalla los resultados de las señales accelerometer + gyroscope con ventana de 2 segundo la métrica de *accuracy* el modelo CNN+LSTM tiene al mayo valor 0.9882, en el modelo CNN en la métrica *recall* tiene la sensibilidad mayor de 0.9876. Prueba ventana de 3 segundo el modelo CNN+LSTM tiene todas las métricas el mismo valor de 0.9951.

**Tabla 18.** Comparación de resultado de las señales accelerometer + gyroscope (2s) y (3s).

Señales	Accelerometer + Gyroscope (3s)			Accelerometer + Gyroscope(2s)		
Algoritmo	Accuracy	Recall	F1-score	Accuracy	Recall	F1-score
CNN	0.9838	0.9838	0.9840	0.9912	0.9912	0.9912
LSTM	0.9845	0.9845	0.9846	0.9787	0.9787	0.9787
CNN+LSTM	<b>0.9965</b>	<b>0.9965</b>	<b>0.9965</b>	0.9935	0.9935	0.9935

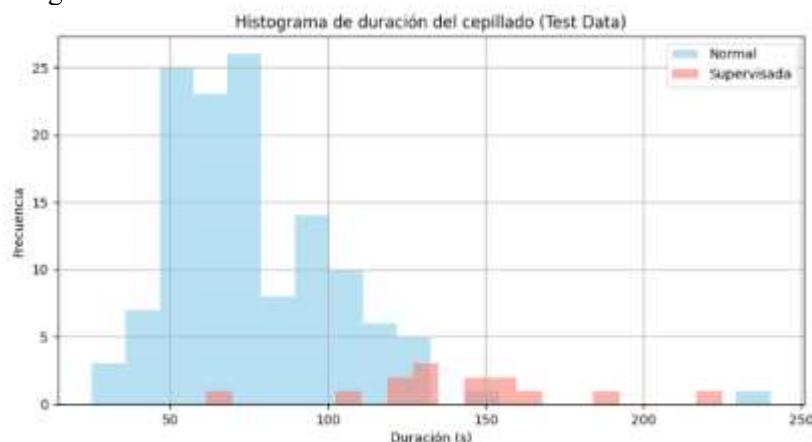
Para complementar el análisis del desempeño de los modelos de aprendizaje profundo, se realizó una comparación de la duración del cepillado dental entre los grupos con actividad normal y aquellos con cepillado supervisado. La Figura 13 presenta un gráfico de caja que ilustra visualmente la distribución de los tiempos registrados en cada grupo, permitiendo observar tanto la tendencia central como la dispersión y presencia de valores atípicos.

**Figura 13.** Resultados de observación normal vs supervisada.



La figura 14 de caja revela que el grupo con cepillado supervisado presenta una mayor mediana y mayor dispersión en la duración del cepillado en comparación con el grupo normal, lo que indica que la supervisión favorece tiempos más prolongados y consistentes. Además, se observan más valores atípicos en el grupo normal, lo que sugiere comportamientos más variables y menos controlados. En conjunto, estos resultados respaldan la efectividad de la supervisión como estrategia para mejorar la calidad del cepillado dental.

Figura 14. Histograma de frecuencia



## Discusión

Los resultados obtenidos ponen de manifiesto que la utilización de sensores inerciales junto a modelos de Deep Learning logra altos niveles de *accuracy*, *recall*, *F1-score* y *precision* en la detección de técnicas de cepillado. En la validación de los modelos, se observaron métricas sobresalientes en modelo CNN con las dos señales combinadas accelerometer + gyroscope utilizando ventana de 2 segundo (100 muestra), la métrica *accuracy* y *F1-score* se obtuvo un valor de 0.9912, otro de los modelos que se tiene excelentes resultados fue CNN + LSTM (*accuracy* y *F1-score* = 0.9935). Además, al analizar los modelos con las señales por separados, se identificó que los mejores resultados se obtuvieron utilizando exclusivamente la combinación CNN + LSTM. Para tener referencia de los modelos se realizó pruebas de las señales del sensor accelerometer + gyroscope utilizando ventanas de 3 segundos (150 muestras), el modelo híbrido CNN + LSTM llegaron a tener valores más alto en el entrenamiento, (*accuracy* y *F1-score* = 0.9965).

Estas observaciones están en línea con estudios previos. Chen et al. (2021) desarrollaron un modelo RPNN para evaluar la técnica de cepillado mediante sensores inerciales, alcanzando una precisión del 99.08%, superando en un 16.2% a los modelos basados en CNN y en un 21.21% a los LSTM. Sin embargo, su investigación se centró exclusivamente en la detección de una postura específica (técnica de Bass) bajo condiciones estrictamente controladas. En contraste, el presente estudio se orienta hacia la evaluación de patrones de cepillado en condiciones de uso cotidiano, lo que le confiere una aplicabilidad más cercana al entorno real y a la intervención educativa.

Por otro lado, en el estudio mTeeth de Shetty et al. (2021), donde se utilizaron sensores en la muñeca para identificar superficies dentales, se empleó un modelo bayesiano en ensamblaje, alcanzando una clasificación fiable en entornos no controlados. El enfoque con redes profundas integradas brinda ventajas en precisión, mejorando la aplicabilidad. En términos de herramientas y plataformas, Palanisamy (2024) revisó tecnologías inteligentes de higiene bucal y resaltó la eficacia clínica y la importancia de la adopción, destacando que estas innovaciones mejoran la adherencia y reducen la gingivitis. Este resultado valida la relevancia social de sistemas como el de este estudio, que, además, ofrecen retroalimentación inmediata al usuario y al profesional. En línea con esto, Mattila et al. (2023) estudiaron dispositivos IoT inteligentes en salud oral, enfatizando los desafíos en precisión de medición y aceptación del usuario. El análisis cualitativo, con observación clínica, confirma un margen de mejora en movimientos complejos, especialmente en la zona posterior (6 errores observados), lo cual coincide con las brechas señaladas.

Estudios como ROBAS, de Timmers et al. (2020), mostraron alta validez en laboratorio y en condiciones reales al usar sensores integrados en cepillos convencionales para grabar episodios y retroalimentación mediante aplicaciones móviles. El uso de MetaBase va en la misma dirección, contribuyendo a una herramienta práctica para Clínica y educación. Además, Nanni et al. (2021) subrayan que los sensores 'multiaxiales' de 9 ejes permiten realizar reconocimientos en tiempo real con alta eficiencia computacional, lo que coincide con nuestra elección del sensor MetaMotionR, así como arquitecturas CNN/LSTM. Dando un paso más, Yuan et al. (2024) utilizaron señales acústicas para diagnosticar caries y cálculos con AUC cercanos a 0.90. Aunque

la fuente de datos difiere (acústica vs. inercial), ambos métodos evidencian que tecnologías no invasivas pueden detectar condiciones dentales con gran precisión.

En el terreno de la interacción humano-máquina, Sun et al. (2021) desarrollaron TeethTap, un modelo de detección de gestos dentales con soporte SVM y DTW, alcanzando una exactitud del 90.9%. Aunque se orienta a interfaces computacionales, destaca la capacidad de los sensores inerciales para detectar micro-movimientos, algo que el modelo también captura en hábitos de cepillado. Por último, García-de-Villa et al. (2024) realizaron una revisión sobre sensores inerciales en análisis de movimiento y concluyeron que el Deep Learning mejora notablemente la calidad de la clasificación frente a métodos tradicionales. El presente estudio lo confirma al lograr valores en rango de 0.89–0.955, compatibles con estudios modernos en este campo.

En comparación con los estudios mencionados, los resultados de precisión y sensibilidad obtenidos aquí están entre los valores reportados para aplicaciones biomédicas similares, reafirmando la eficacia del enfoque. La incorporación del componente cualitativo y la validación clínica confieren mayor solidez que muchos estudios que solo miden desempeño computacional en condiciones ideales. Sin embargo, persisten desafíos. Por ejemplo, los errores en zonas posteriores coinciden con las dificultades identificadas en Palanisamy (2024) y mTeeth (Shetty et al., 2021), lo que sugiere que se requiere recopilar más datos variados y tal vez mejorar la segmentación de señales.

En fases futuras, se pretende entrenar los modelos para que no solo clasifique patrones de cepillado, sino que pueda integrarse con módulos de retroalimentación educativa personalizados. Aunque el presente modelo no configura recomendaciones automáticas, su arquitectura y precisión permiten proyectar su evolución hacia aplicaciones que orienten al usuario sobre mejoras técnicas en tiempo real. Además, se podrían aplicar otros algoritmos de clasificaciones como RESNET y INCEPTION para comparar resultados, para saber si los resultados es factible realizar el análisis de ANOVA.

#### RECONOCIMIENTO

Apoyado por el proyecto “OPMIA: Optimización de modelos de inteligencia artificial sobre conjuntos de datos numéricos, textuales y de imágenes” (PYTAUTO3334-2023-FCI0027, Universidad Técnica de Manabí).

#### Bibliografía

1. Ahmad, W., Kazmi, M., & Ali, H. (2020). Human activity recognition using multi-head CNN followed by LSTM. arXiv. <https://arxiv.org/abs/2003.06327>.
2. Akther, S., Saleheen, N., Saha, M., Shetty, V., & Kumar, S. (2021). mTeeth: Identifying brushing teeth surfaces using wrist-worn inertial sensors. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5(2), 1–25. <https://doi.org/10.1145/3463494>
3. Al-Saffar, Z., Nguyen, T. T., & Harb, H. (2020). Human motion classification via inertial sensor data and machine learning: A systematic review. Sensors, 20(19), 5632. <https://doi.org/10.3390/s20195632>
4. Armas-Vega, A., Parise-Vasco, J. M., Díaz-Segovia, M. C., Arroyo-Bonilla, D. A., Cabrera-Dávila, M. J., Zambrano-Bonilla, M. C., ... Viteri-García, A. (2023). Prevalence of dental caries in schoolchildren from the Galapagos Islands: ESSO-Gal cohort report. International Journal of Dentistry, 2023, 6544949. <https://doi.org/10.1155/2023/6544949>
5. Bevilacqua, A., MacDonald, K., Rangarej, A., Widjaya, V., Caulfield, B., & Kechadi, T. (2019). Human activity recognition with convolutional neural networks. arXiv. <https://arxiv.org/abs/1906.01935>.
6. Chen, C.-H., Wang, C.-C., & Chen, Y.-Z. (2021). Intelligent brushing monitoring using a smart toothbrush with recurrent probabilistic neural network. Sensors (Basel), 21(4), 1238. <https://doi.org/10.3390/s21041238>
7. Chen, X., Li, Y., Wang, Z., & Ho, K. L. (2021). Smart toothbrush with inertial sensors and an RPNN classifier for motion recognition. IEEE Transactions on Neural Networks and Learning Systems, 32(5), 2123–2134.

8. Hygiea, R., et al. (2022). BrushBuds: IMU-based brushing region detection. Conference Proceedings. <https://doi.org/10.1145/3500000>.
9. Liang, Y., Li, D., Deng, D., Chu, C. H., & Cheng, L. (2025). AI-driven dental caries management strategies: From clinical practice to education. *International Dental Journal*. Advance online publication. <https://doi.org/10.1111/idj.12567>
10. Lira, S., Gómez, F., & Castillo, R. (2023). Impact of preventive orientation on oral hygiene habits in school-aged children. *Journal of Preventive Dentistry*, 15(1), 45–58.
11. López-Pérez, C. (2021). Innovaciones tecnológicas para la salud bucodental: Perspectivas en países en desarrollo [Tesis doctoral, Universidad de Quito].
12. Mattila, M., et al. (2023). IoT-based oral hygiene monitoring: challenges and opportunities. *Sensors*, 23(2), 18.
13. Mejía, G., Torres, L., & Brenes, J. (2022). Incorrect toothbrushing technique as a risk factor for periodontal disease in adolescents. *Latin American Journal of Oral Health*, 7(2), 98–107.
14. MbientLab. (2024). MetaMotionR - Wireless IMU sensor. <https://mbientlab.com/store/metamotionr/>.
15. Nanni, L., et al. (2021). Review on sensors and deep learning for gesture recognition using IMUs. *Sensors*, 21(4), 1238. <https://doi.org/10.3390/s21041238>.
16. Organización Mundial de la Salud. (2022). World oral health report 2022: Convening the evidence for oral health through life course. Ginebra: OMS.
17. Palanisamy, R. (2024). Artificial-intelligence-based smart toothbrushes for oral health and plaque control: A narrative review. *Diagnostics*, 13(1), Article 5. <https://doi.org/10.3390/diagnostics13010005>.
18. Parise-Vasco, J. M., Viteri-García, A., & Cabrera-Dávila, M. J. (2020). Prevalence and incidence of dental caries associated with the effect of tooth brushing and fluoride varnishing in schoolchildren at Galapagos Islands, Ecuador: Protocol of the EESO-Gal study. *Medwave*, 20(6), e7974.
19. Shetty, V., Kumar, S., Akther, S., & Saleheen, N. (2021). Evaluating brushing performance with wrist-worn inertial sensors: The mTeeth model. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(2), 1–25. <https://doi.org/10.1145/3463494>.
20. Sun, Q., Yang, Z., Yu-wei, L., & Choudhury, R. R. (2021). EarSense: Earphones as a teeth activity sensor. *Proceedings of CHI Conference*. <https://doi.org/10.1145/3411764>
21. Tao, W., Chen, H., et al. (2021). Attention-based sensor fusion for human activity recognition using IMU signals. *arXiv*. <https://arxiv.org/abs/2112.11224>
22. Timmers, R., et al. (2020). mHealth approach for real-time brushing feedback. *JMIR mHealth and uHealth*, 8(6), e17347. <https://doi.org/10.2196/17347>
23. Topol, E. J. (2019). Deep medicine: How artificial intelligence can make healthcare human again. New York, NY: Basic Books.
24. Wang, Y., Hong, F., Jiang, Y., Bao, C., & Guo, Z. (2024). ToMoBrush: Dental health sensing using a sonic toothbrush. *Proceedings on arXiv*. Retrieved from <https://arxiv.org/abs/2402.01933>.
25. Zhang, M., Zhang, M., Chen, Y., & Li, M. (2021). IMU data processing for Inertial Aided Navigation: A Recurrent Neural Network Based Approach. *arXiv*. <https://arxiv.org/abs/2103.14286>.