Dr. Carlos Minutti Martínez

Doctor en Ciencia en Ingenería de la Computación

Candidata o Candidato a Investigadora o Investigador Nacional (SNII-C)

Datos de contacto:

carlos.minutti@infotec.mx

Inteligencia Artificial Generativa para el Análisis Transparente de Imágenes Médicas: Modelos de Aprendizaje Profundo Explicables.

Datos Generales

Resumen ejecutivo

Los modelos de aprendizaje profundo (AP) en el análisis de imágenes médicas ofrecen capacidades notables para la detección y diagnóstico de enfermedades. Sin embargo, su naturaleza de "caja negra" genera preocupaciones sobre la explicabilidad y sesgos, dificultando su adopción clínica. Este trabajo propone un enfoque basado en IA generativa para abordar estos desafíos. Aprovechamos los autoencoders para aprender representaciones comprimidas (espacio latente), capturando características clave utilizadas para la reconstrucción y clasificación de imágenes. El análisis e interpretación del espacio latente proporciona información sobre el proceso de toma de decisiones del modelo, permitiendo relacionar las variables del espacio latente con cambios visuales en la imagen. Esto permite la identificación y mitigación de sesgos sin necesidad de reentrenamiento del modelo o modificación de datos. Además, al crear mapas de atención y técnicas de conciencia del contexto para permitir que el modelo se centre en la información pertinente relevante para la tarea de clasificación actual, permite mejorar la precisión de la clasificación y también reduce la dependencia de datos etiquetados, permitiendo así un enfoque semi-supervisado. Es asi que el marco propuesto tiene como objetivo mejorar la transparencia y equidad de los modelos de aprendizaje profundo en el análisis de imágenes médicas.

Línea General y Aplicación del Conocimiento (LGAC)

8. Analítica de datos e información

Palabras clave

Análisis de Imágenes Médicas, Inteligencia Artificial Explicable, Inteligencia Artificial Generativa

Objetivo General

Desarrollar un marco de aprendizaje profundo generativo novedoso y explicable para el análisis de imágenes médicas que permita la identificación y mitigación de sesgos, mejorando la transparencia, la confianza y la adopción en el dominio clínico.

Objetivos especificos

Diseñar un modelo de autoencoders con espacio latente interpretable para la representación de imágenes médicas. - Integrar técnicas de atención y modelado de contexto para mejorar la precisión de la clasificación y reducir la dependencia de datos etiquetados. - Desarrollar métodos para identificar y mitigar sesgos en el modelo utilizando el espacio latente interpretable. - Evaluar el rendimiento del modelo propuesto en tareas de clasificación de imágenes médicas del mundo real.

Datos del proyecto

Descripción

El análisis de imágenes médicas es crucial para los sistemas de diagnóstico asistido por computadora (CAD) y detección de enfermedades. El aprendizaje profundo, especialmente las redes neuronales convolucionales (CNN), ha demostrado un rendimiento de vanguardia en diversas tareas de análisis de imágenes médicas, incluida la detección de enfermedades, la segmentación de lesiones y la clasificación de imágenes. Sin embargo, la falta de explicabilidad y transparencia es un obstáculo crítico para su adopción generalizada en el cuidado de la salud. Estos modelos a menudo se tratan como cajas negras, lo que dificulta la comprensión y explicación de sus procesos de toma de decisiones. Nuestro marco propuesto pretende abordar estos desafíos mediante el aprendizaje profundo generativo. Se basa en autoencoders para aprender representaciones del espacio latente que capturan características clave de las imágenes médicas. Analizando e interpretando este espacio latente, podemos relacionar las variables latentes con características visuales en las imágenes de entrada. Mediante el análisis de este espacio latente se busca desarrollar técnicas para la identificación y mitigación de sesgos sin necesidad de reentrenamiento del modelo o modificación de datos.

Antecedentes del problema a resolver

Los modelos de aprendizaje profundo han revolucionado el análisis de imágenes, pero su falta de explicabilidad y transparencia dificulta su adopción en áreas críticas como la medicina. La incapacidad para comprender cómo estos modelos toman decisiones genera desconfianza y limita su utilidad en aplicaciones de alto riesgo. Además, los modelos existentes de aprendizaje profundo, como las redes neuronales convolucionales y los transformers de visión, a menudo se entrenan con conjuntos de datos de imágenes naturales, como ImageNet. Sin embargo, las imágenes médicas tienen características y patrones visuales muy diferentes a las imágenes naturales, lo que puede dificultar la transferencia efectiva de conocimiento de estos modelos pre-entrenados al dominio médico.

Justificación y pertinencia

Los modelos entrenados exclusivamente con imágenes naturales pueden no ser óptimos para tareas de análisis de imágenes médicas y podrían requerir un ajuste sustancial o un reentrenamiento completo. Además, estos modelos suelen ser computacionalmente costosos y requieren grandes cantidades de datos etiquetados para su entrenamiento, lo cual es un desafío en el dominio médico donde los conjuntos de datos anotados son escasos y su obtención es costosa. Por lo tanto, es crucial desarrollar modelos específicamente diseñados y pre-entrenados para el análisis de imágenes médicas. Estos modelos deben aprovechar las características y patrones visuales únicos de las imágenes médicas, lo que permitiría una transferencia de conocimiento más efectiva y un mejor rendimiento en tareas clínicas. Además, deben ser computacionalmente eficientes para permitir su implementación en diversos entornos, incluyendo dispositivos de recursos limitados. Un enfoque prometedor es pre-entrenar estos modelos en conjuntos de datos de imágenes médicas de gran escala, aprovechando técnicas de aprendizaje semisupervisado y autosupervisado. Esto permitiría al modelo capturar representaciones relevantes para el dominio médico, reduciendo la necesidad de grandes cantidades de datos anotados durante el entrenamiento final en tareas específicas. El desarrollo de modelos explicables para el análisis de imágenes médicas es crucial para mejorar la atención médica. La transparencia en la toma de decisiones por parte de los modelos de IA fomenta la confianza y la adopción en el dominio clínico. Al proporcionar información sobre el proceso de toma de decisiones y permitir la detección y mitigación de sesgos, este enfoque generativo podría contribuir al desarrollo de sistemas más confiables, justos y aplicables en la práctica clínica. En resumen, este proyecto aborda desafíos críticos como la interpretabilidad, la equidad, la escasez de datos y la falta de modelos específicos y eficientes para el análisis de imágenes médicas. Al generar modelos computacionalmente eficientes y pre-entrenados en el dominio médico, se reduce la necesidad de grandes conjuntos de datos anotados, facilitando su adopción en entornos con recursos limitados y acelerando el avance de la atención médica asistida por IA.

Metas

-Desarrollar un marco generativo de IA basado en autoencoders para el análisis de imágenes médicas. - Analizar e interpretar el espacio latente aprendido por el modelo para comprender su proceso de toma de decisiones. - Integrar mapas de atención y técnicas de conciencia contextual para mejorar la precisión de la clasificación y reducir la dependencia de datos etiquetados. - Validar el marco propuesto en un conjunto de datos diverso y evaluar su efectividad para identificar y mitigar sesgos. - Diseñar plataformas web que permitan el uso y exploración de los modelos desarrollados.

Metodologías

- Diseño e implementación de la arquitectura de autoencoders: Se explorará y optimizará una arquitectura de autoencoders personalizada, incorporando elementos de redes neuronales eficientes como ShuffleNet y técnicas de aprendizaje de representaciones desacopladas. - Incorporación de mecanismos de atención y modelado de contexto: Se integrarán mapas de atención optimizados y técnicas de modelado de contexto, como el borrado aleatorio de parches de imagen, para mejorar la precisión de clasificación y reducir la dependencia de datos etiquetados. - Pre-entrenamiento y transferencia de aprendizaje: Se utilizará un conjunto de datos de imágenes médicas de gran tamaño, como MiMeta, para realizar un pre-entrenamiento del modelo, aprovechando técnicas de aprendizaje supervisado y no supervisado. Posteriormente, se realizará una transferencia de aprendizaje a tareas específicas de análisis de imágenes médicas. - Análisis e interpretación del espacio latente: Se desarrollarán técnicas para visualizar y analizar el espacio latente aprendido por el autoencoder, relacionando las variables latentes con características visuales en las imágenes de entrada. - Detección y mitigación de sesgos: Utilizando las capacidades de interpretabilidad del enfoque generativo, se implementarán estrategias para detectar y mitigar sesgos en el modelo, sin necesidad de volver a entrenarlo o modificar los datos. - Evaluación y validación: El marco de trabajo propuesto se evaluará y validará utilizando múltiples conjuntos de datos públicos y privados de imágenes médicas, incluyendo tareas de clasificación, detección y segmentación de enfermedades.

Resultados esperados

- Un marco de trabajo de Inteligencia Artificial generativa basado en autoencoders, capaz de aprender representaciones latentes interpretables de imágenes médicas. - Mejora en la precisión de clasificación y detección de enfermedades en tareas de análisis de imágenes médicas, en comparación con los enfoques existentes. - Capacidad para visualizar y analizar las variables latentes aprendidas, relacionándolas con características visuales en las imágenes de entrada. - Técnicas efectivas para la detección y mitigación de sesgos en el modelo, aprovechando las capacidades de interpretabilidad del enfoque generativo. - Reducción de la dependencia de datos etiquetados mediante el uso de técnicas semisupervisadas y de modelado de contexto. - Validación del marco de trabajo propuesto en múltiples conjuntos de datos de imágenes médicas y tareas de análisis. - Distintas versiones de los modelos para diferentes capacidades de cómputo.

Cronograma de trabajo

#	Entregable(s) comprometido(s)	Fecha inicio	Fecha fin
1	Artículo de investigación en revista especializada	01/04/2024	31/12/2024
2	Divulgación	01/02/2024	31/12/2024
3	Difusión	01/02/2024	31/12/2024
4	Vinculación	01/04/2024	31/12/2024
5	Innovación de la invención.	01/10/2024	31/12/2024

Bibliografía relevante

Burgess, Christopher & Higgins, Irina & Pal, Arka & Matthey, Loic & Watters, Nick & Desjardins, Guillaume & Lerchner, Alexander. Understanding disentangling in β-VAE. (2018). https://doi.org/10.48550/arXiv.1804.03599 Minutti-Martinez, C., Escalante-Ramírez, B., Olveres-Montiel, J. PumaMedNet-CXR: An Explainable GenerativeArtificial Intelligence for the Analysis and Classification of Chest X-Ray Images. Lecture Notes in ComputerScience, (2023). vol 14392. Springer, Cham. https://doi.org/10.1007/978-3-031-47640-2_18 Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretablemodels instead. Nat Mach Intell 1, 206–215 (2019). https://doi.org/10.1038/s42256-019-0048-x Sarvamangala, D.R., Kulkarni, R.V. Convolutional neural networks in medical image understanding: a survey. Evol. Intel. 15, 1–22 (2022). https://doi.org/10.1007/s12065-020-00540-3 Solatidehkordi, Z.; Zualkernan, I. Survey on Recent Trends in Medical Image Classification Using SemiSupervised Learning. Appl. Sci. (2022), 12, 12094. https://doi.org/10.3390/app122312094 Suganyadevi, S., Seethalakshmi, V. & Balasamy, K. A review on deep learning in medical image analysis. Int JMultimed Info Retr 11, 19–38 (2022). https://doi.org/10.1007/s13735-021-00218-1