



ugr

Universidad
de **Granada**

APLICACIONES EN CIENCIA DE DATOS E INGENIERÍA DE COMPUTADORES

MÁSTER EN CIENCIA DE DATOS E INGENIERÍA DE COMPUTADORES

Imágenes Biomédicas e Identificación Forense

Autora

Lidia Sánchez Mérida



Curso 2021 - 2022

Granada, Junio de 2022

Descripción del artículo

Debido a la cantidad masiva de datos que generan diferentes disciplinas relativas a la medicina, en este [artículo](#) se pretende realizar una revisión del estado del arte con la que identificar las técnicas más prometedoras aplicadas a la **clasificación de imágenes biomédicas**. Gracias a los relevantes progresos e investigaciones en las áreas de la Inteligencia Artificial y el Aprendizaje Automático, se han desarrollado algoritmos y metodologías capaces de detectar enfermedades y mejorar sus tratamientos asociados de forma eficiente. En particular, este artículo se centra en el análisis del rendimiento relativo a cada técnica en función del volumen de datos a tratar, estableciendo un importante énfasis en el uso de algoritmos de **Big Data** con los que aumentar tanto la precisión como la velocidad de los procedimientos involucrados.

Relevancia del problema

La evaluación de diferentes tipos de información conlleva una desmesurada **inversión de recursos** personales, monetarios y temporales. Concretamente, la gestión, organización y el estudio de imágenes para la detección de distintas dolencias es especialmente costosa. Sin embargo, empleando métodos de computación y aprendizaje automáticos a partir del conocimiento de expertos, puede ayudar en la construcción de modelos predictivos capaces de realizar este procedimiento de manera más veloz y con menor esfuerzo, convirtiéndose en sistemas de ayuda en la toma de decisiones. Por lo tanto este artículo persigue el objetivo de acelerar los diagnósticos, monitorizar su progreso y explorar los diversos tratamientos aplicables intentando seleccionar el más óptimo, reduciendo así la sobrecarga existente en el ámbito de la salud.

Análisis de resultados

Existen multitud de algoritmos de clasificación de imágenes que han sido desarrollados durante los últimos años de investigación en este campo. Algunos de los pioneros que han demostrado proporcionar mejores resultados en base a distintas métricas de calidad, como la precisión y el número de muestras positivas (*sensitivity*) y negativas (*specificity*) bien clasificadas, han sido las **Redes Neuronales** y las **Support Vector Machine** (SVM). Muchos de los trabajos analizados en este resumen se encuentran orientados a problemas específicos, como la detección de tumores cerebrales, por lo que no son aplicables a cualquier ámbito de la medicina. Otros aplican reglas de **Lógica Difusa** con el cometido de medir el grado de aportación de cada uno de los atributos al problema de clasificación. No obstante, una de las

principales limitaciones de estas estrategias reside en el volumen de datos que gestionan, puesto que si es demasiado alto comienzan a perder tanto eficiencia como capacidad de predicción.

Una de las propuestas más competentes para solventar la *maldición de la dimensionalidad* reside en las **Redes Neuronales Convolutivas** (CNN), pertenecientes al ámbito del *Deep Learning*. Su objetivo consiste en realizar una extracción de características a diferentes niveles de abstracción con la habilidad de registrar el nivel de contribución de cada atributo a la resolución del problema. Gracias a sus exitosas aplicaciones en la clasificación de imágenes de distinta índole, también se desarrollaron otro tipo de operaciones como la localización de entidades o la segmentación de imágenes. Con el fin de **comparar** la habilidad resolutoria en clasificación de imágenes biomédicas de sendas técnicas, en el trabajo analizado presentan un flujo de trabajo con el que llevar a cabo las etapas de entrenamiento y evaluación de modelos. La principal conclusión destacable de los resultados reside en el característico elevado rendimiento que demuestra el clasificador CNN en datasets de **mayor tamaño**.

Como se ha comentado anteriormente, este artículo impulsa la introducción de metodologías de **Big Data** con las que incrementar más aún el rendimiento de los algoritmos de clasificación sobre conjuntos masivos de información. Para ejemplificar el modo de funcionamiento han creado un segundo *workflow* empleando el ecosistema de *Spark* por su capacidad de paralelización e integración de multitud de librerías con algoritmos y técnicas de Big Data. Tras cargar el listado de imágenes biomédicas en el sistema de ficheros HDFS, haciendo uso del paradigma **MapReduce** se plantean dos principales actividades, siendo la primera consistente en generar los **descriptores** de las imágenes que contienen las características más relevantes, pudiendo aplicar operaciones matemáticas para transformar sus propiedades visuales en vectores en el espacio. Esta aproximación puede decrementar considerablemente su eficiencia si existen multitud de imágenes de tamaños ínfimos o si su contenido difiere demasiado. Existen dos posibles alternativas para paliar ambas situaciones a partir de la integración de la extracción de características secuencialmente o mediante el empleo de técnicas de segmentación de imágenes. Utilizando la misma estrategia de paralelización se afrontan las dos siguientes etapas de entrenamiento y validación de modelos predictivos usando distintos algoritmos. Finalmente como conclusiones subrayan un **incremento del rendimiento y eficacia** de los clasificadores obtenidos sobre conjuntos de datos de gran volumen por su capacidad de aprendizaje autónoma desde la experiencia y sin la necesidad de prácticamente ninguna intervención humana.

Comentario crítico

La clasificación de imágenes en las distintas áreas de la salud ha sido una de las temáticas principales que más repercusión ha tenido en la última década y a la que se le ha dedicado un gran número de esfuerzos y recursos. Los objetivos que persigue proponen beneficios de distinta naturaleza, como la detección precoz de enfermedades o la monitorización y estudio de los tratamientos existentes con los que encontrar los más eficaces para cada paciente de manera personalizada. Con herramientas capaces de efectuar análisis detallados a partir de un volumen masivo de información se podrían evitar múltiples inconvenientes característicos de la sanidad actual, como las interminables esperas para la atención de especialistas, la sobrecarga de los sanitarios o la detección tardía de enfermedades que provoca un tremendo impacto negativo en el tratamiento de los pacientes.

Por otro lado he podido observar que el artículo se encuentra fundamentado en una considerable cantidad de referencias bibliográficas, por lo que parece que este resumen ha sido generado de forma minuciosa y concienzuda intentando abarcar la mayor parte de las técnicas que constituyen el estado del arte pasado y presente. Si bien recopilan diversas pruebas de rendimiento entre los algoritmos estudiados, a mi parecer falta la adición de las conclusiones extraídas por los artículos revisados, de modo que enriquezca más aún la información que representan. En particular, en la segunda tabla comparativa no se proporciona ningún tipo de detalle acerca de las métricas utilizadas para el análisis de ambos algoritmos, los conjuntos de datos involucrados o los resultados comparativos. Si bien el artículo trata de realizar un resumen acerca de las mejores técnicas para clasificación de imágenes biomédicas, no termino de comprender las conclusiones extraídas de ciertas secciones o los motivos explicativos del apoyo que ejerce a la introducción de algoritmos y métodos de Big Data.

Finalmente considero que las conclusiones redactadas al final del artículo son prácticamente una repetición de las generadas al final de los apartados anteriores, por lo que no aportan información de valor. El trabajo futuro planteado que consiste en la aplicación del flujo de trabajo desarrollado en Spark debería haber sido incluido en este artículo para demostrar la relevancia de la combinación propuesta entre Aprendizaje Automático y Big Data, ya que a mi juicio su justificación se encuentra considerablemente escasa y definida.