

El triaje de los incidentes de urgencias y emergencias extrahospitalarias representa un reto difícil, principalmente debido a las limitaciones temporales—que exigen una evaluación rápida de las prioridades en un estrecho margen de tiempo—y a la incertidumbre—tomar decisiones con la información disponible. Además, los errores en este proceso pueden tener graves consecuencias para los pacientes, con el consiguiente riesgo de muerte. Por lo tanto, cualquier protocolo, herramienta o estrategia novedosa que haya demostrado mejorar estos procesos puede ofrecer un valor sustancial en términos de atención al paciente y gestión global de los incidentes médicos de urgencias y emergencias extrahospitalarias.

La hipótesis fundamental en la que se basa esta tesis es que el Aprendizaje Automático, concretamente el Aprendizaje Profundo, puede mejorar significativamente estos procesos proporcionando estimaciones de la gravedad de los incidentes médicos de urgencia y emergencia extrahospitalaria, teniendo en cuenta la información de la que dispone el operador en el momento del triaje del incidente durante la llamada de emergencia. Mediante el análisis de millones de datos derivados de llamadas de emergencia de la Comunitat Valenciana (España) que abarcan desde 2009 hasta 2019, planteamos que los modelos de Machine Learning podrían extraer patrones que pueden conferir capacidad predictiva a esta tarea.

Por ello, esta tesis profundiza en el diseño y desarrollo de varios modelos de Aprendizaje Automático, concretamente modelos de Aprendizaje Profundo Multitarea que aprovechan los datos multimodales asociados a eventos de urgencias y emergencias extrahospitalarias. Nuestro objetivo principal era predecir tres etiquetas indicativas de la gravedad del incidente, influyendo así en su priorización. Estas etiquetas englobaban si el incidente suponía una situación de riesgo vital, la demora admisible de la respuesta (desde no demorable hasta minutos, horas o días) y si era competencia del sistema de emergencias o de atención primaria. Utilizando datos disponibles entre 2009 y 2012, los resultados obtenidos fueron prometedores. Se observaron mejoras sustanciales en las métricas macro F1, con ganancias del 12.5% para la clasificación de riesgo vital, del 17.5% para la demora en la respuesta y del 5.1% para la clasificación por jurisdicción, en comparación con el protocolo interno de triaje de la Comunidad Valenciana.

Sin embargo, es esencial tener en cuenta que los sistemas, los protocolos de triaje y las prácticas operativas evolucionan de forma natural con el tiempo. Los modelos que mostraron un rendimiento excelente con el conjunto de datos inicial de 2009 a 2012 no demostraron la misma eficacia cuando se evaluaron con datos que abarcaban de 2014 a 2019 (los datos de 2013 no estaban disponibles). Este último conjunto de datos había sufrido modificaciones significativas en comparación con el anterior. Estas modificaciones provocaron cambios en el conjunto de datos, lo que dio lugar a variaciones en las distribuciones de probabilidad, que hemos caracterizado e investigado meticulosamente en esta tesis, centrándonos en su impacto en el rendimiento del modelo.

Continuando con nuestra investigación, nuestro objetivo era proporcionar un rendimiento sostenible del modelo a lo largo del tiempo o, como mínimo, mitigar los efectos adversos de las inevitables variaciones en la distribución de los datos de la forma más eficaz posible. Para hacer frente a este reto, nos centramos en el Aprendizaje Continuo Profundo. Al incorporar el paradigma del Aprendizaje Continuo a nuestros

diseños y desarrollos, pudimos mitigar sustancialmente los efectos adversos sobre el rendimiento y comprender mejor cómo gestionar el despliegue de modelos a lo largo del tiempo en un centro de atención a la llamada de urgencias y emergencias médicas extrahospitalarias. Los resultados de nuestra investigación indican que, al considerar el Aprendizaje Continuo Profundo, si bien no elimina por completo las fluctuaciones de rendimiento a lo largo del tiempo, las mantiene efectivamente dentro de un rango manejable. En particular, con respecto a la métrica F1, cuando las variaciones distribucionales son ligeras o moderadas, el comportamiento se mantiene estable, sin variar más de un 2,5%, como se observa en nuestros datos de incidentes médicos extrahospitalarios. Por lo tanto, bajo estas condiciones, el rendimiento de nuestros modelos es operativamente aceptable.

Además, nuestra tesis demuestra la viabilidad de construir herramientas auxiliares que permitan a los operadores interactuar con estos complejos modelos. En consecuencia, sin interrumpir el flujo de trabajo de los profesionales, se hace posible proporcionar retroalimentación mediante predicciones de probabilidad para cada clase de etiqueta de gravedad y tomar las medidas adecuadas en función de estas predicciones.

Por último, los resultados de esta tesis tienen implicaciones directas en la gestión de las urgencias y emergencias extrahospitalarias en la Comunidad Valenciana. El modelo final resultante de nuestra investigación está previsto que se integre en los centros de atención de llamadas asociadas a urgencias y emergencias médicas de la Comunidad Valenciana. Este modelo utilizará los datos proporcionados por los operadores telefónicos para calcular automáticamente las predicciones de gravedad, que luego se compararán con las generadas por el protocolo de triaje interno. Cualquier disparidad entre estas predicciones desencadenará la derivación del incidente a un coordinador médico, que supervisará su tratamiento. Por lo tanto, es evidente que nuestra tesis, además de realizar importantes contribuciones al campo de la Investigación en Aprendizaje Automático Biomédico, también conlleva implicaciones sustanciales para mejorar la gestión de las urgencias y emergencias extrahospitalarias en el contexto de la Comunidad Valenciana.