

Detección de hallazgos inesperados, un sistema para gestión de alertas clínicas.

M. Cirugeda Pablos¹, D. González Oviaño¹, I. Polimón Olabarrieta², T. Meizoso³, E. M. Cirugeda Roldán¹

¹Universidad Rey Juan Carlos, Fuenlabrada, España, {m.cirugeda.2019,d.gonzalezo.2019}@alumnos.urjc.es, eva.cirugeda@urjc.es

²Dermatología, Hospital Universitario de Móstoles, Móstoles, España, isabel.polimon@salud.madrid.org

³Anatomía Patológica, Hospital Universitario de Móstoles, Móstoles, España, telma.meizoso@salud.madrid.org

Resumen

El diagnóstico precoz de un hallazgo inesperado de malignidad es de vital importancia en el tratamiento de un paciente y clave para su recuperación. Actualmente es necesario que el facultativo que solicitó la prueba realice una búsqueda proactiva de los resultados de un hallazgo inesperado de malignidad en los informes emitidos por el departamento de Anatomía Patológica (AP). El principal problema de este sistema de comunicación radica en los tiempos de ejecución, además de en los posibles errores en la cadena de transmisión de información, como pueden ser fallos humanos que impidan que el facultativo obtenga el resultado con la celeridad deseada. El sistema de automatización propuesto en este trabajo se basa en un análisis automático mediante búsqueda de palabras clave de los informes generados por AP utilizando técnicas de Procesamiento del Lenguaje Natural. Los resultados ofrecen una alta capacidad discriminativa de clasificación lo que permite generar alertas de hallazgos inesperados de manera correcta. No obstante, dada la especificidad de la aplicación, ésta exige que sea orientada entrono al diccionario establecido por los responsables clínicos.

1. Introducción

Una precoz notificación de un hallazgo inesperado permite el diagnóstico inmediato e inicio del tratamiento cuando corresponde. Una mínima demora en alguno de los dos procesos repercute gravemente en la recuperación, calidad de vida y riesgo de mortalidad del paciente. Actualmente el departamento de Anatomía Patológica (AP) se encarga de esta notificación. Ellos redactan y gestionan todos los informes realizados en un sistema integrado llamado Vitropath [1]. Sin embargo, para comunicarse con otros servicios, el diagnóstico debe ser volcado en un sistema de información sanitaria (Selene) y el médico petionario acceder a él. En caso de ser un diagnóstico inesperado el Hospital Universitario de Móstoles no tiene sistematizada la notificación de estos hallazgos. Estos avisos se realizan, o bien por teléfono (el patólogo avisa al médico petionario) o enviando una copia en papel del informe a la consulta. En muchas ocasiones no se produce notificación alguna, y el conocimiento de estos resultados depende de la voluntad y disponibilidad de los médicos. Estos tienen que hacer una búsqueda proactiva de las muestras enviadas para comprobar si los resultados concuerdan con la sospecha diagnóstica de la solicitud.

El principal problema que presentan estos métodos es el retraso en el diagnóstico de estos hallazgos inesperados. Es decir, se pierde un tiempo hasta que el médico, que no

sospechaba lesión maligna, revise de manera rutinaria un resultado o a un paciente. Además, puede darse la circunstancia de que se pierda el seguimiento de un resultado, ya sea porque el paciente no acude a su revisión o el médico enferma. De tal forma, estos resultados se quedarían, inicialmente, sin valorar.

Para subsanar estas consecuencias y realizar una rápida notificación, se propone un sistema de gestión de alerta automático que minimice al máximo el número de fallos humanos y trasapelados. Se pretende que sistemáticamente, sin necesidad de intervención del médico petionario de una biopsia, se avise de aquellos resultados que sean inesperados. Estos hallazgos serán reconocidos dentro del informe por los algoritmos explicados en este proyecto. Cuando se detecte diagnóstico de malignidad, este informe será candidato para el envío de una alerta. Este será realizado de manera automática al médico petionario y al servicio correspondiente.

Este trabajo propone una metodología para la gestión de alertas de hallazgos inesperados basada en técnicas de procesamiento de textos y *Natural Language Processing* (NLP). Se ha elegido esta última técnica ya que es una disciplina en auge donde se realiza eficazmente el reconocimiento de palabras [2]. Ésta emplea técnicas computacionales con el fin de aprender y comprender el lenguaje humano. Los investigadores actuales la utilizan en la creación de sistemas de diálogo, traducción de voz o identificación de sentimientos y emociones en las redes sociales [3]. Asimismo, la aplicación de técnicas de NLP está cobrando cada vez mayor importancia en el ámbito clínico en distintas aplicaciones como la mejora del flujo de trabajo en el servicio de Radiología [4] y en la identificación distintas patologías en Nefrología [5].

2. Metodología

2.1. Base de datos.

Se dispone de una muestra de 70 informes clínicos anonimizados correspondientes a análisis de biopsias solicitadas por Dermatología. El fichero proporcionado cumple la estructura mostrada en la Tabla 1 para todos los casos registrados.

Nombre campo	Tipo campo (tamaño)
Estudio	Varchar(32)
Diag_clin	Mediumtext
Registro_tumores	Char(1)
Conservación	Char(1)

Tabla 1. Estructura de la base de datos de cada informe clínico.

Se etiqueta como estudio al identificador único del informe ya encriptado, una cadena de longitud variable cuya máxima longitud son 32 caracteres. El diagnóstico clínico corresponderá al diagnóstico realizado por el patólogo, una cadena de caracteres que distingue entre minúsculas y mayúsculas y con una longitud máxima de 16.777.215 caracteres. En este campo se encontrarán indicadas las patologías observadas por estudio. La variable Registro_tumores es un indicador de malignidad que solo puede albergar un dato. En caso de ser positivo se indicará con una S, en caso contrario se dejará vacío. Por último, el campo conservación se trata de una casilla de verificación propuesta por el servicio de AP para determinar los hallazgos inesperados. Al igual que Registro_tumores, solo puede adquirir dos posibles valores (S, ""). Cada campo descrito en la Tabla 1 se encuentra separado por el carácter especial "|".

2.2. Adecuación de la base de datos.

Los datos en bruto contienen caracteres no deseados. Antes de aplicar cualquier técnica, se realiza un análisis de textos donde se depura la información de la base de datos y se obtiene el contenido necesario para la implementación de los algoritmos.

2.3. Diccionario específico.

Dado que no todos los identificativos tumorales son malignos, el servicio AP ha definido un diccionario de palabras clave que definen un hallazgo inesperado y palabras clave que lo descartan. Se observa en la Tabla 2 ambos diccionarios.

Palabras clave
Carcinoma, adenocarcinoma, melanoma, linfoma, tumor neuroendocrino, positivo, carcinoide, neoplasia, mielodisplásico, mieloproliferativo, mieloma, sarcoma, micosis fungoide, angiosarcoma, liposarcoma, leucemia, cistoadenocarcinoma, colangiocarcinoma.
Palabras excepción
Carcinoma basocelular

Tabla 2. Diccionario de palabras clave y excepción.

2.4. Generación de alerta.

El sistema debe generar alarma en tres casos característicos, en el resto, no se generará.

- Caso 1: Se hayan detectado las palabras clave en el informe, y ambos *checks* estén activados.
- Caso 2: Se hayan detectado las palabras clave en el informe y ninguno de los *checks* estén activados.

- Caso 3: Se hayan detectado las palabras clave en el informe y el *check* de conservación está activado y el *check* de registro de tumores esté desactivado.

En la Tabla 3 se muestra la estructura de generación de alerta donde 1 significa la activación del campo. Para generar estas alertas, se han desarrollado dos técnicas, una de ellas basada en análisis de caracteres strings y lógica combinacional y otra de ellas en técnicas NLP.

Registro de Tumores	Conservación	Palabra clave	Alerta
1	1	1	1
0	1	1	1
0	0	1	1

Tabla 3. Generación de alerta con campos binarios.

2.5. Técnica basada en NLP.

Esta técnica se basa en la implementación de NLP. Se apoya en la librería *spacy*, una librería de software de código abierto diseñada para facilitar tareas de NLP [6]. A diferencia de expresiones regulares en texto, *spacy* permite crear vectores de coincidencias encontrando palabras y frases [7].

El vocabulario aplicado cuenta con una sintaxis predefinida. De tal manera, se podrá llamar al documento y generar el objeto NLP. Este método demanda diccionarios específicos como medio para detectar coincidencias, por ello, se establecen tres listas de diccionarios para generar el buscador, etiquetadas como *pattern*. Los diccionarios mencionados describirán las palabras clave, las palabras excepción y el valor S. Se define el comparador basado en reglas, etiquetado como *matcher*. Además, se fija el vocabulario que debe coincidir con el documento en el que trabajará. El *matcher* se añade al *pattern* y se buscan las coincidencias.

Se evalúa cada *matcher* en cada muestra y se estudia su longitud. Si se obtiene una longitud del *matcher* de palabra clave mayor o igual a 1, se entenderá la existencia de malignidad en dicha muestra. De igual forma, se procederá para el resto de *matchers*.

El resultado proporciona el campo estudio y cuatro campos binarios, registro de tumores, conservación, palabra clave y alerta. La alerta es implementada atendiendo a los 3 casos mencionados con bucles *for* y lógica combinacional.

Finalmente, se añadirá el campo estudio junto con la alerta a un fichero. Se definirán los estudios que deben ser notificados, es decir, estudios que contengan un 1 en el campo alerta. Este fichero será usado por el Servicio de Informática que alberga las herramientas y el sistema necesario para generar la alerta al peticionario del estudio.

2.6. Técnica de análisis de string y lógica combinacional.

La técnica está basada en reconocimiento de strings y lógica combinacional. Se obtiene el encriptado para cada estudio y, mediante bucles *for* anidados, se evalúa si el estudio contiene el valor S en los posibles campos. En caso de que exista

supondrá la activación del mismo. De igual forma, se recorre cada muestra valorando si existen palabras clave en el diagnóstico.

La estructura del resultado es análoga a la distribución observada en la Tabla 3 a excepción del campo alerta. En este método se obtendrá dicho campo atendiendo a los casos propuesto con bucles *for* y lógica combinacional. Asimismo, se enviará el fichero correspondiente al Servicio de Informática con la misma metodología descrita en la técnica NLP.

2.7. Métodos de caracterización.

Por último, a fin de evaluar las prestaciones del sistema propuesto se ha considerado realizar la matriz de confusión, curva ROC [8] y el Área Bajo la Curva, valorando la capacidad discriminativa del sistema. También, para comparar ambas técnicas implementadas, se han calculado los tiempos de ejecución realizando 5 diferentes particiones de las 70 muestras anonimizadas, (10, 25, 35, 50 y 70). Dividiéndolos por el número de muestras, se obtienen los tiempos medios de ejecución por informe. Finalmente, se calcula la media (μ) de las 5 particiones y la desviación estándar (σ). Se observan los resultados en la Tabla 5.

3. Resultados

Se dispone de la base de datos anotada por el servicio de AP. En ella se describen los registros que deberían producir una etiqueta de alerta positiva.

El sistema desarrollado para el procesamiento de texto e importación de los informes clínicos realiza adecuadamente su función. Las etiquetas representadas por *check* conservación y *check* Registro_tumores fueron correctamente marcadas para cada informe.

A continuación, en la Tabla 4, se muestra la matriz de confusión asociada a esta tarea de clasificación. Esta refleja dos errores cometidos.

	Alerta +	Alerta -
Predicción +	6	2
Predicción -	0	62

Tabla 4. Matriz de confusión

A partir de la Tabla 4, se pueden calcular los estadísticos de Sensibilidad (SEN=100%), Especificidad (ESP=96.88%), Valor Predictivo Positivo (VPP=75%) y Valor Predictivo Negativo (VPN=100%).

Para evaluar la clasificación binaria del sistema, se muestra la curva ROC en la Figura 1. Esta muestra la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos. A partir de esta se calculó el Área Bajo la Curva (AUC=0.98) que representa la capacidad del modelo para distinguir entre las clases positivas y negativas.

Finalmente, la Tabla 5 muestra los tiempos medios de ejecución en milisegundos de los algoritmos por registro.

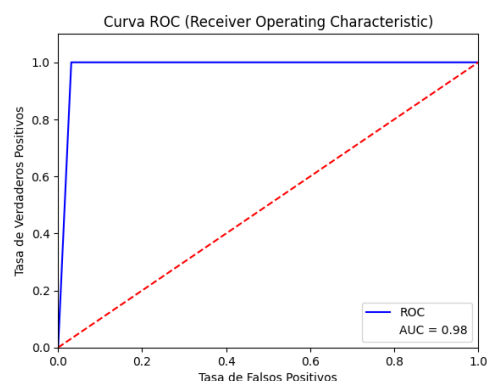


Figura 1. Curva ROC

	10	25	35	50	70	μ	σ
String	0,058	0,059	0,057	0,079	0,061	0,06	0,01
NLP	28.43	24.74	25.69	27.56	26.42	26.59	1.31

Tabla 5. Tiempos de ejecución (ms/informe)

4. Discusión

De acuerdo con los resultados mostrados, el algoritmo realiza una correcta clasificación del 97% de los registros considerados. Estos dos errores se deben a la localización de la palabra “carcinoma” dentro de las oraciones “carcinoma de células basales” y “sin evidencia de carcinoma”. En lo referido a la primera frase, se relaciona directamente con el diagnóstico benigno de “carcinoma basocelular”. En cuanto a la frase de “sin evidencia de”, realizando una correcta interpretación, puede descartarse el hallazgo inesperado, aunque exista una palabra clave de malignidad.

Se observa que estos errores se deben principalmente a las características de redacción del patólogo. Como solución particular, se propuso la adición en el diccionario de exclusiones de las palabras “carcinoma de células basales” y “sin evidencia”. Estas son soluciones muy específicas que provocan que una utilización más generalizada sea más difícil de implantar.

Alternativamente, existen técnicas basadas en inteligencia artificial como puede ser la Representación de Codificador Bidireccional de Transformadores [9] o BERT por sus siglas en inglés. Estas permiten realizar un análisis contextual de palabras ambiguas y una interpretación más amplia de su significado. De esta forma podrían aumentar la robustez en la clasificación generando únicamente alarmas en aquellos casos en que fuera estrictamente necesario.

En cuanto a los errores de clasificación, los resultados de la curva ROC muestran que se emitirán alarmas de hallazgos inesperados en casos en los que no debieran emitirse, pero nunca se perderán alarmas que sí se corresponden con hallazgos de malignidad inesperados. De esta forma, aunque puede existir una sobrecarga de alarmas, el médico receptor de la alarma puede descartarla sin que suponga un detrimento en la salud o recuperación del paciente asociado.

En general, estos resultados, muestran una alta capacidad de discriminación de clasificación del sistema. No obstante, se han llevado a cabo soluciones específicas que hacen que este modelo se pueda ajustar a los informes proporcionados.

Por otro lado, se han podido descubrir los errores llevados a cabo por el clínico en la activación de *checks*. Este error sucede si se detecta palabra clave y no *check* Registro_tumores. Esta equivocación humana se ha producido 2 veces y ambas han sido detectadas por los algoritmos.

Además, se han podido detectar 3 casos en los el *check* Registro_tumores estaba activado, pero no se ha detectado palabra clave correspondiente a esa malignidad. Gracias a la base de datos anotada, estos 3 diagnósticos se corresponden con “nevus melanocítico”, “carcinoma basocelular” y “fibroxantoma”. Más tarde se aseguró que se correspondían con diagnósticos malignos. En base a este hallazgo cabe destacar que la actualización continua del diccionario de las palabras clave es de vital importancia. Ya que en caso de no detectar palabras clave, no se enviaría aviso, independientemente de los *checks* marcados.

En cuanto a los tiempos de ejecución de la Tabla 5, se observa que la media de procesamiento por informe para el método por comparación de listas de *strings* es menor. Esto se debe a que este método se ciñe a la detección de palabras clave, mientras que NLP, integra también la detección de los *checks* y la creación de las alarmas. A pesar de su rapidez, el método tradicional por comparativas de *strings* es más complejo computacionalmente. La necesidad de numerosos bucles *for* consecutivos hace que esta metodología sea más propensa a errores. El uso de librerías de Python como *spacy* simplifican y sintetizan notablemente el algoritmo de NLP y su detección de palabras clave.

5. Conclusión

El presente trabajo describe el diseño, desarrollo y evaluación de un sistema de alerta de los resultados inesperados de los informes emitidos por el servicio de Anatomía Patológica del Hospital Universitario de Móstoles de Madrid.

Este sistema evita la necesidad de la intervención proactiva del dermatólogo peticionario de una biopsia. Se pretende reducir al máximo el número de fallos humanos y fallos en la cadena de transmisión de información.

El trabajo propone dos diseños diferentes para su implementación basados en la detección de diagnósticos de malignidad en el informe. Uno se apoya en la comparación de listas de *strings* y otra se basa en técnicas básicas de NLP. Debido a su aplicación con una orientación muy específica, las metodologías presentan ciertas limitaciones. Se exigen una cierta normalización en la redacción de los informes y un amplio diccionario de palabras claves. Aunque los resultados son buenos, estos podrían mejorarse y reducir los problemas encontrados utilizando métodos basados en inteligencia artificial que utilicen técnicas más sofisticadas de NLP como pueden ser la BERT. Dichas técnicas podrían realizar una interpretación de la frase en un contexto más general una vez detectada una palabra clave asociada a la malignidad y poder determinar un veredicto más ajustado.

Procedimientos más generalistas permitirían implementar este sistema en otras áreas clínicas que colaboren con la división de Anatomía Patológica. Para ello se debería

ampliar el diccionario de palabras clave, de acuerdo con la terminología usada en el servicio en cuestión.

A nivel funcional, el sistema ayudaría a agilizar las peticiones sin la necesidad de la disponibilidad del médico, el cual invierte mucho tiempo en analizar informe a informe. Por otro lado, se evitarían retrasos en el diagnóstico del paciente permitiendo un contacto rápido y un inmediato inicio del tratamiento preciso. Esto repercutiría directamente sobre la calidad de su vida, puesto que una demora en la detección o en el tratamiento provoca el empeoramiento de la salud del paciente e incluso la muerte.

Agradecimientos

Por último, agradecer al Servicio de Informática el Hospital Universitario de Móstoles por su colaboración.

Referencias

- [1] Vitro. (s. f.). Vitropath marcando el cambio. Indubuilding. <http://www.vitropath.com/download/Vitropath-marcando-el-cambio.pdf>
- [2] Briega, R. E. L. (2017, 23 septiembre). Procesamiento del Lenguaje Natural con Python. <https://relopezbriega.github.io/blog/2017/09/23/procesamiento-del-lenguaje-natural-con-python/>.
- [3] Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245), 261-266.
- [4] López-Úbeda, P., Martín-Noguerol, T., Juluru, K., & Luna, A. (2022). Natural Language Processing in Radiology: Update on Clinical Applications. *Journal of the American College of Radiology: JACR*, 19(11), 1271-1285. <https://doi.org/10.1016/j.jacr.2022.06.016>.
- [5] Van Vleck, T. T., Farrell, D., & Chan, L. (2022). Natural Language Processing in Nephrology. *Advances in chronic kidney disease*, 29(5), 465-471. <https://doi.org/10.1053/j.ackd.2022.07.001>.
- [6] datos.gob.es. (2022, 23 junio). Probamos SpaCy: mucho más que una librería para crear proyectos reales de procesamiento del lenguaje. [https://datos.gob.es/es/blog/probamos-spacy-mucho-mas-que-una-libreria-para-crear-proyectos-reales-de-procesamiento-del#:~:text=SpaCy%20es%20una%20librer%C3%ADa%20de,de%20bajo%20nivel%20muy%20eficiente\).](https://datos.gob.es/es/blog/probamos-spacy-mucho-mas-que-una-libreria-para-crear-proyectos-reales-de-procesamiento-del#:~:text=SpaCy%20es%20una%20librer%C3%ADa%20de,de%20bajo%20nivel%20muy%20eficiente).)
- [7] Rule-based matching · spaCy Usage Documentation. (s. f.). Rule-based matching. <https://spacy.io/usage/rule-based-matching>.
- [8] Statologos: El sitio web para que aprendas estadística en Stata, R y Python. (2021). Cómo trazar una curva ROC en Python (paso a paso). Statologos: El sitio web para que aprendas estadística en Stata, R y Python. <https://statologos.com/plot-roc-curve-python/>.
- [9] Liu, H., Zhang, Z., Xu, Y., Wang, N., Huang, Y., Yang, Z., Jiang, R., & Chen, H. (2021). Use of BERT (Bidirectional Encoder Representations from Transformers)-Based Deep Learning Method for Extracting Evidences in Chinese Radiology Reports: Development of a Computer-Aided Liver Cancer Diagnosis Framework. *Journal of medical Internet research*, 23(1), e19689. <https://doi.org/10.2196/19689>