

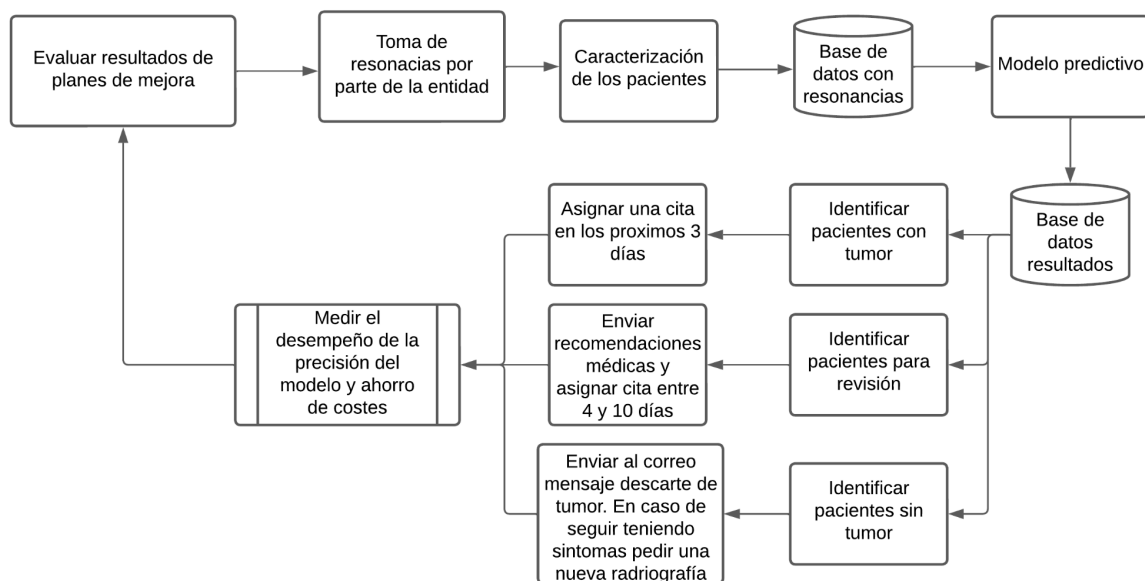
Problema de negocio

El tumor cerebral representa del 85% al 90% de todos los tumores primarios del sistema nervioso central, además cada año 11.700 personas son diagnosticadas con un tumor cerebral. Los tumores cerebrales se clasifican en tumor benigno, tumor maligno, tumor pituitario, entre otros. Los diagnósticos son realizados a través de resonancias magnéticas, estas imágenes son examinadas por radiólogos o neurocirujanos quienes toman la decisión. El sistema de salud actual presenta demoras para la lectura de estos exámenes y para dar citas prioritarias por parte de los profesionales del área, por otra parte, los médicos especialistas presentan sobrecarga y fatiga en muchas ocasiones por la gran cantidad de trabajo. Los problemas principales del hospital son el tiempo de asignación de citas para la lectura de las resonancias magnéticas y la alta ocupación que tiene el personal que está relacionada a esta área en el hospital, lo que ha generado las siguientes situaciones: complicaciones en pacientes que necesitan una intervención inmediata de su enfermedad, gastos inútiles en personal de radiología al descartar tumores en personas, aumento de PQRS por el déficit de atención, entre otras.

Problema analítico

Desarrollar un modelo analítico que determine si una persona tiene o no un tumor cerebral a partir de resonancias magnéticas. En las resonancias se incluyen diferentes tipos de tumores cerebrales como tumor benigno, tumor maligno, tumor pituitario y finalmente resonancias de personas sin el tumor.

Diseño de la solución



Gráfica 1: Diseño de la solución

Este proyecto pretende diseñar un modelo capaz de clasificar con precisión si una persona presenta o no un tumor cerebral. Las radiografías se ubicaran en una base de datos del hospital que se actualizará todas las noches con las radiografías tomadas el mismo día en la entidad, además, esta información hará parte de una caracterización que tiene el hospital de cada persona que se encuentra en el mismo servidor. La ejecución automática del programa se realizará todos los días a las 6 am, luego por medio de una persona que administre esta información, según la clasificación obtenida, enviará al correo las recomendaciones a cada persona. La predicción arroja la probabilidad para cada paciente y las personas con una probabilidad mayor al 90% se les asignará una cita prioritaria con un máximo de tres días con especialistas en la enfermedad, en donde se tomará una decisión de cirugía o tratamiento según sea el caso; las personas con una probabilidad menor al 10% serán descartadas de tener algún tumor cerebral, sin embargo se envían recomendaciones por si se presentan nuevamente síntomas relacionados a la enfermedad; finalmente, el resto de pacientes obtendrán sus citas con el método ya existente por parte de la entidad. El modelo se entrena todos los días con las nuevas imágenes tomadas del día anterior.

Las métricas de negocio son: los costos de la utilización de radiólogos; el costo que implica la respuesta de PQRS; el tiempo de espera para la asignación de citas e indemnizaciones por denuncias de usuarios. Traducidas en el problema analítico se busca una predicción rápida diaria para la clasificación de las personas y así poder dar sus citas, además esperar una precisión mayor al 80% en personas diagnosticadas con tumor y para personas sin presencia de un tumor. Es importante recordar que según la clasificación descrita en el párrafo anterior y la precisión esperada, para el hospital es muy importante que las predicciones del modelo sean precisas ya que asignar una cita con un especialista a un falso positivo implica mayores gastos que la revisión de un radiólogo en primera instancia, además, un falso negativo en relación a no tener tumor implicaría al largo plazo complicaciones en estas personas y por lo tanto posibles demandas que afecten la sostenibilidad del hospital.

Limpieza y transformación

Las radiografías tienen diferentes dimensiones dependiendo el tipo de tumor, en este caso la imagen original de no tumor tiene una dimensión en píxeles de 630 x 630 en tres canales, sin embargo se trabajará con una dimensión de 120 por 120 píxeles de forma que los modelos puedan capturar la mayor información sin afectar drásticamente los resultados. Por ejemplo:

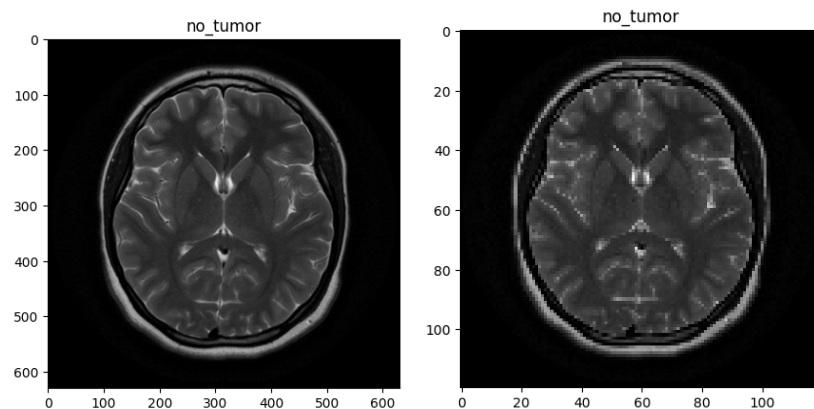


Imagen 1: Ejemplo radiografía de 630 x 630 a 120 x 120 píxeles

Selección de algoritmos y técnicas de modelado

Se inició creando dos modelos de machine learning y una red neuronal full connected donde se obtuvo:

Modelo	Data	Categoría	Presicion	Recall	F1-score	Accuracy
RL	Train	AVG	1	1	1	1
	Test	AVG	0.93	0.89	0.90	0.90
RFC	Train	AVG	1	1	1	1
	Test	AVG	0.91	0.85	0.87	0.88
RN-Simple	Train	AVG	1	1	1	1
	Test	AVG	0.93	0.90	0.91	0.91

Tabla 1: Valores de ajuste de modelos RL, RFC y RN-simple

Como se observa en ambos modelos RL y RFC hay sobreajuste, sin embargo se obtienen en general buenas métricas. Estos tipos de modelos dejan en el camino características propias de la imagen, por lo tanto se consideran otros modelos que pueden responder mejor a cambios en las imágenes de entrada y que además aprenden otros patrones. Se propone una red tradicional

En la red neuronal tradicional se obtiene menor sobreajuste en general como se observa en la imagen 2, solo en el epoch número dos la precisión en validación cayó un poco. Se realizó una regularización y dropout para diagnosticar el comportamiento de la red y poder disminuir el sobreajuste, evento que se logró como se puede ver en la imagen 3.

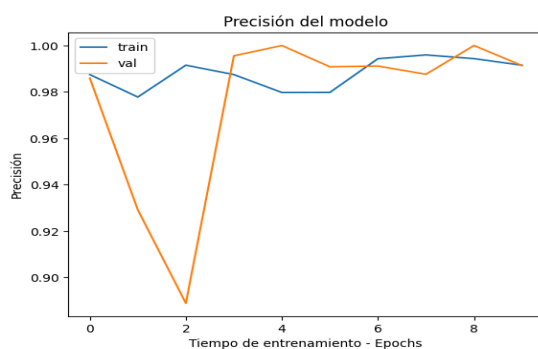


Imagen 2: Ajuste red neuronal tradicional

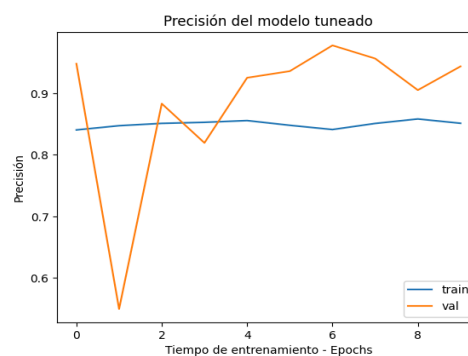


Imagen 3: Red regularizada y dropout

Ahora se entrenará un red neuronal convolucional, primero se creó una red convolucional sin dropout y regularización donde se obtuvo sobreajuste como se observa en la imagen 4. Esto pudo ser originado por la similitud entre las imágenes de entrenamiento y las imágenes de test, por lo tanto se probó otra red neuronal con regularización y dropout, en donde se obtuvo una pequeña mejora, además se observa que la red es sensible a estos métodos que disminuyen el sobreajuste, se puede evidenciar en la imagen, por lo que finalmente se realiza el tuning de hyperparameters para encontrar el mejor ajuste para el modelo.

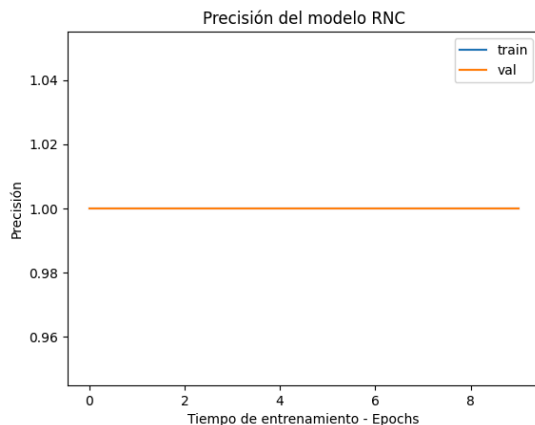


Imagen 4: Ajuste de red neuronal convolucional

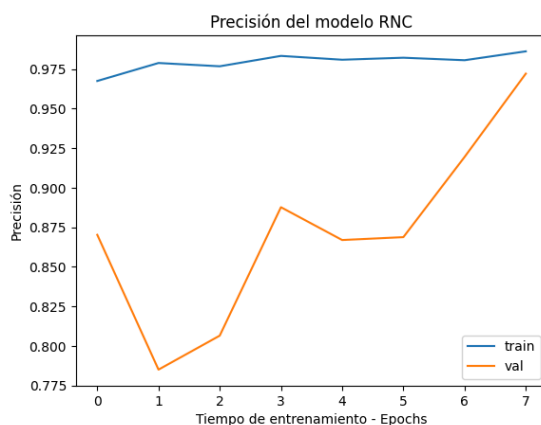


Imagen 5: Red convolucional regularizada

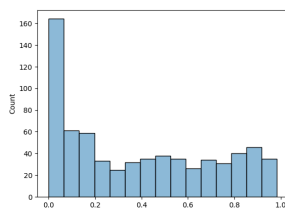
Afinamiento de hyper parámetros y evaluación

Luego de los diagnósticos anteriores se realizaron cuatro tuning diferentes al modelo, para observar comportamientos y diferencias entre ellos, la descripción es la siguiente:

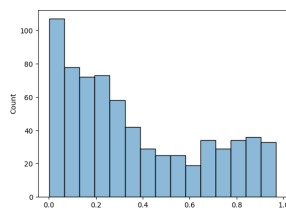
Primer y segundo tuning: para estos tuning inicialmente se realizó una disminución de la dimensionalidad de las imágenes a 150 x 150 píxeles (No a 120x120), además solo se tunearon los hyperparameters dropout, la fuerza de la regularización y el optimizador para el primer tuning, para el segundo los mismos hyperparameters anteriores y además la activación de las capas, que no era tan necesario debido a que Relu tenia un buen comportamiento, pero se quiso realizar para comparar resultados y además el learning rate. Analizando estos dos modelos se observa en la gráfica 2 que las probabilidades arrojadas por el modelo están distribuidas aproximadamente constante luego del 10% y por lo tanto si se tuviera un umbral del 50% la mayoría de los pacientes no tuvieran la presencia del tumor, este comportamiento se puede estar originando por la variabilidad de los tipos de tumor que se encuentran en los datos que complican el aprendizaje de la red. Los resultados del segundo tuning son parecidos al tuning anterior, sin embargo si se analiza el mismo umbral las predicciones estarían más cerca de no tener tumor. Luego se intentó balancear las clases.

Tercer tuning: Para este tuning se consideraron los mismo hyper parámetros que el segundo tuning, solo cambio los píxeles por imagen, ahora son 120 x 120 píxeles, además se consideró hacer un balanceo de clases con la técnica de muestreo SMOTEENN, en donde se paso para la categoría tumor de 2475 imágenes a 2097 imágenes y para la categoría no tumor se pasó de 395 imágenes a 2384 imágenes. Como se observa en la gráfica 4 se obtiene un resultado similar a las redes convolucionales anteriores, se destaca de que la probabilidad de sus predicciones se encuentran un poco más cercanas a ambos límites, sin embargo este modelo puede ser peligroso ya que se entrenó con imágenes ficticias generadas por el SMOTEENN. A partir de aquí y teniendo en cuenta que las redes neuronales se les dificulta encontrar patrones en datos donde hay más columnas que ejemplos o número de imágenes, se decidió hacer un esfuerzo en balancear ambas clases con la búsqueda de imágenes reales en internet de personas que no tuvieran el tumor.

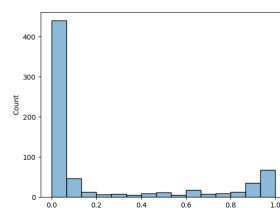
Cuarto tuning: la clase no tumor pasó de 395 a 1595 imágenes. Finalmente este modelo muestra una buena clasificación y está construido sobre más imágenes, lo que pudo originar una pérdida en las métricas en pequeñas unidades, como se observa en la tabla el tuning 4 en el promedio de la precisión se encuentra detrás del tuning 1 y el tuning 2 por pocas unidades, por lo que se decide continuar con este tuning para el modelo.



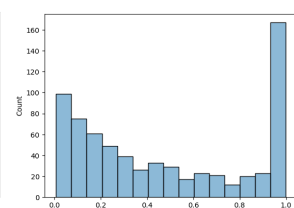
Gráfica 2: Tuning 1



Gráfica 3: Tuning 2



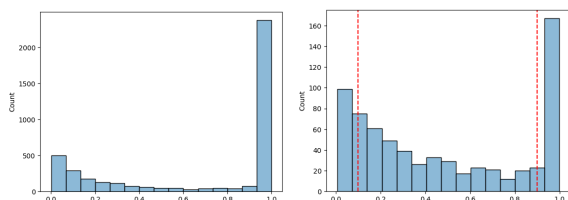
Gráfica 4: Tuning 3



Gráfica 5: Tuning 4

Análisis del modelo

Se crearon dos umbrales en donde el modelo automáticamente gracias a la buena precisión va a predecir en tres categorías (gráfica 6), quienes definitivamente deben de pasar a una cita prioritaria con especialistas, quienes no tienen un tumor y el resto tendrá su cita asignada bajo el sistema actual del hospital, además se observa que el comportamiento del test es similar al comportamiento de las distribuciones en entrenamiento imagen. Se analizan las métricas obtenidas para cada clase según lo anterior, como se observa en la tabla 2:



Gráfica 6: distribución en train y umbrales en test

Tuning	Data	Metrica	Presicion	Recall	F1-score	Accuracy
1	Train	Macro avg	0.95	0.94	0.95	0.95
	Test	Macro avg	0.81	0.80	0.80	0.81
2	Train	Macro avg	0.93	0.93	0.93	0.93
	Test	Macro avg	0.79	0.76	0.76	0.78
3	Train	Macro avg	0.95	0.95	0.95	0.95
	Test	Macro avg	0.81	0.73	0.73	0.77
4	Train	Macro avg	0.94	0.91	0.92	0.93
	Test	Macro avg	0.77	0.77	0.77	0.78
4-Grupo 1	Test	No tumor	0.96	0.33	0.50	0.60
4-Grupo 2	Test	Tumor	0.93	0.57	0.78	0.80

Tabla 2: Análisis de métricas obtenidas

Teniendo en cuenta el diseño propuesto inicialmente y analizando los resultados de prueba del modelo en la tabla anterior se tiene que: para la clasificación sin tumor, de los que se predijo que no tenían tumor se acertó el 96%, de las personas que no tienen tumor se identificó el 33%, para un promedio (F1 Score) del 50%, de todas las predicciones se acertó en un 60%. Finalmente el modelo identificará el 33% de personas sin tumor de forma automática, con un error del 4% en personas identificadas sin tumor. Analizando la sensibilidad del modelo que es muy baja para predecir correctamente quien no tiene realmente un tumor, esta situación está compensada en el umbral del 80% de personas a las cuales se les asignará una cita con el sistema actual del hospital, por lo que el dejar ir personas que no tienen tumor en realidad se sabe que tendrán otra oportunidad con el radiólogo para descartar el tumor. Los que fueron clasificados con no tumor cuando realmente lo tienen, al correo llegará información sobre si vuelve a presentar síntomas relacionados a la enfermedad vuelva a realizarse otra radiografía.

Para la clasificación con tumor, de los que se predijo que tenían tumor se acertó en el 93%, de las personas que tienen tumor se identificó el 57%, para un promedio (F1 Score) del 78%, de todas las predicciones se acertó en un 80%. Finalmente el modelo identificará el 57% de las personas con tumor de forma automática, con un error del 7% en personas identificadas con tumor. Observando los resultados la sensibilidad del modelo es regular, es decir deja ir personas que realmente tienen el tumor como si no lo tuvieran, sin embargo esta situación está compensada en que estas personas volverán a tener de nuevo otra cita con un radiólogo para confirmar el tumor. Los clasificados con tumor cuando realmente no lo tienen, en su cita posterior con especialistas y cirujanos, serán confirmadas las imágenes junto a otros exámenes y su historial médico, en donde se espera que se descarte el tumor.

El modelo identifica tres grupos de pacientes:

20% de personas detectadas sin tumor cerebral

54% de personas seguirán el sistema actual de revisión

25% de personas detectadas con tumor cerebral.

Para ejemplificar, si suponemos que el costo de un radiólogo por diagnóstico de una radiografía de una persona cueste 20.000 COP, si en un día normal ingresan 100 radiografías para ser diagnosticadas el hospital se está ahorrando el 45% de los costos, es decir 900.000 COP diarios, a comparación de los 2.000.000 que costarían todas, suponiendo además la disponibilidad diaria de radiólogos para esta demanda. el 25% de las personas obtendrán una cita máximo en tres días para ser evaluados por especialistas, además el 54% de personas en situación de revisión obtendrán su cita en un menor tiempo.

Despliegue del modelo

Todos los días en la mañana se evaluarán las resonancias de los pacientes por medio del modelo, cuando se obtengan los resultados habrá una persona encargada de enviar un correo con las estrategias planteadas en el diseño según la clasificación de los pacientes:

paciente	clas
image(1)	revision
image(10)	no tumor
image(100)	no tumor
image(11)	revision
image(12)	revision
image(13)	tumor
image(14)	revision
image(15)	revision
image(16)	revision

Tabla 3: Resultados del despliegue.

De esta manera cada paciente por medio de su identificación (en este caso las imágenes no tiene la cédula, pero se realizó la suposición) conocerá si su estado es ‘tumor’, ‘no tumor’ o ‘revisión’, luego esperarán su debida citación según sea el correspondiente caso.