

**UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
ESCUELA DE POSTGRADO**



**“Diseño de un modelo de auditoría genérico e integrado
de Teleinformes basado en imágenes médicas”.**

CLAUDIO ANDRÉS CATALÁN CERDA

**ACTIVIDAD FORMATIVA EQUIVALENTE PARA OPTAR AL GRADO DE
MAGISTER EN INFORMÁTICA MÉDICA.**

Director de Tesis: Prof. Dr. Víctor Castañeda Z.

2022

INDICE

INDICE.....	3
INDICE DE FIGURAS	5
INDICE DE TABLAS.....	6
ABREVIATURAS	8
RESUMEN.....	11
ABSTRACT	13
1.- INTRODUCCIÓN	14
1.2.- ANTECEDENTES	14
1.2.1.- Telemedicina: Definiciones y situación de Teleinformes en el país.....	14
1.2.2.- Calidad en salud, en contextos telemédicos.....	18
1.2.3.- Herramientas para el aseguramiento de la calidad y mejora continua.....	22
1.2.4.- Problema/Oportunidad.....	24
2.- OBJETIVOS	27
OBJETIVO GENERAL.....	27
OBJETIVOS ESPECÍFICOS	27
3.- MARCO TEORICO.....	28
3.1.- Revisión Bibliográfica.....	28
3.1.1.- Métodos de medición de la calidad de la imagen.....	28
3.1.2.- Revisión por pares.....	37

3.1.3.- Interoperabilidad	43
3.1.4.- Estándares	45
3.1.5.- Herramientas Informáticas y mensajería	47
4.- MATERIALES Y MÉTODOS	58
4.1.- Levantamiento de procesos	58
4.2.- Implementación	61
4.4.- Limitaciones de la prueba.....	68
5.- RESULTADOS.....	69
5.1.- Levantamiento de procesos de Teleinformes	69
5.2.- Resultado subproceso Evaluación de la calidad de imagen	73
5.3.- Resultado subproceso Revisión por pares. Evaluación de Discrepancia.....	74
5.4.- Integración de procesos de auditoría con sistemas de teleinformes	75
5.5.- Arquitectura de Sistema planteada	77
5.7.- Resultados de evaluación de la segmentación del Caso de Uso.....	83
6.- DISCUSIÓN	97
7.- CONCLUSIÓN	102
ANEXOS	104
ANEXO 1 Plataforma web para la evaluación de imágenes.	104
ANEXO 2 Encuesta de opinión usuaria.....	108
ANEXO 3 Respuestas de Encuesta de Opinión Usuaría.	111

REFERENCIAS	115
-------------------	-----

INDICE DE FIGURAS

Figura 1. Esquema de las dimensiones de la calidad en salud.....	19
Figura 2. Representación de actual proceso de evaluación de reportes RADPEER	43
Figura 3. Representación esquemática de los niveles de interoperabilidad en salud	45
Figura 4. Representación relacional entre API, en particular REST API, y cliente	48
Figura 5. Relación entre dos sistemas fuente y destino mediante motor integrador	49
Figura 6. Imagen en donde ilustra las etiquetas o Tags de archivos DICOM	55
Figura 7. Relación Solicitud/Respuesta SCU y SCP.....	56
Figura 8. Proceso de revisión de la calidad de la segmentación.....	67
Figura 9. Levantamiento de proceso de Tele dermatología.....	70
Figura 10. Levantamiento de proceso de Tele oftalmología.....	71
Figura 11. Levantamiento de proceso de Teleradiología.....	72
Figura 12. Secuencia del subproceso de evaluación de la calidad de imagen.	73
Figura 13. Secuencia de subproceso de evaluación de informes.	74
Figura 14. Integración de sistema de telereportes con sistema de auditoría	76
Figura 15. Arquitectura sistema informático.	77
Figura 16. Configuración canal Mirth.....	79
Figura 17. Tags Estudios.....	80

Figura 18.	Esquema UML de Funcionamiento del Front End.....	82
Figura 20.	Puntuación general de la segmentación.....	84
Figura 21.	Diagrama de cajas de valores de afectación por lóbulos	90
Figura 22.	Diagrama de caja realizado con el puntaje total de afectación.....	91
Figura 23.	Diagrama de cajas con la distribución segmentacion vidrio esmerilado.....	93
Figura 24.	Diagrama de cajas con la distribución de consolidación	93
Figura 25.	Diagrama de cajas con la distribución de pulmón.....	94
Figura 26.	Diagrama de cajas con la distribución de los valores de puntaje total	94

INDICE DE TABLAS

Tabla 1.	Ejemplo de Criterios de calidad de imagen	34
Tabla 2.	Tercer y último esquema de puntuación de evaluación RADPEER.....	40
Tabla 3.	Descripción de los criterios asociados a la puntuación en Escala de Likert.	64
Tabla 4.	Puntaje de severidad según porcentaje de afectación	65
Tabla 5.	Evaluación específica de la segmentación según	66
Tabla 6.	Resumen de los resultados obtenidos en la primera dimensión de evaluación..	83
Tabla 7.	Puntaje de severidad de lóbulos y compromiso pulmonares segmentado.	85
Tabla 8.	Tabla Matriz de concordancia entre resultados de evaluadores con IA TRad. ..	87
Tabla 9.	Representación de porcentajes de severidad y compromiso	88
Tabla 10.	Distribución de frecuencias de valor de afectación por lóbulo pulmonar.....	90

Tabla 11. Resultados de evaluación por criterios en segmentación en vidrio esmerilado 92

Tabla 12. Resultado de evaluación por criterios en segmentación consolidación 92

Tabla 13. Resultado de evaluación por criterios en segmentación pulmón 92

ABREVIATURAS

OMS:	Organización mundial de la salud.
UAPO:	Unidad de Atención Oftalmológica Primaria
IHEE:	Institute of Electrical and Electronics Engineers
ISO:	International Standardization Organization
CEN:	Committee for Standardization
BSI:	British Standards Institution
ANSI:	American National Standards Institute
SDOs:	Standards Developing Organizations
HL7:	Health Level Seven
JIC:	Joint Initiative Council
DICOM:	Digital Imaging and Communications in Medicine
PACS:	Picture Archiving and Communication Systems
BPMN:	Business Model and Notation
APS:	Atención Primaria en Salud
ROC:	Receiver Operating Characteristics
VGA:	Visual Grading Analysis
VGC:	Visual Grading Characteristics

IOM:	Institute Of Medicine
ACR:	American College of Radiology
API:	Application Programming Interface
HTTP:	Hypertext Transfer Protocol
HIS:	Hospital Information System
RIS:	Radiology Information System
EMR:	Electronical Medical Record
REST:	Representational State Transfer
HTML:	HyperText Markup Language
SMTP:	Simple Mail Transfer Protocol
FTP:	File Transfer Protocol
URL:	Uniform Resource Locator
DNS:	Domain Name System
TCP/IP:	Transmission Control Protocol/Internet Protocol
SGBD:	Sistema Gestor de Bases de Datos
FHIR:	Fast Healthcare Interoperability Resources
SOAP:	Simple Object Access Protocol
SCU:	Service Class User

SCP: Service Class Provider

AET: Application Entity Title

RESUMEN

Desde que la telemedicina comenzó a ser implementada en nuestro país, ha mostrado un crecimiento sostenido, sustentado ampliamente en el constante desarrollo y mejora de las tecnologías de la información.

Los teleinformes o telereportes, en particular, han contribuido a disminuir brechas de accesibilidad a la población, por lo que han tenido una fuerte alza en su uso desde que han sido implementados en el ámbito de la salud. Prueba de ello, ha sido la aparición de diversas empresas proveedoras de estos servicios, a las que el estado recurre a través de licitaciones públicas. Cabe señalar que, inherente a todo proceso médico, se encuentra el aseguramiento de la calidad en la ejecución y entrega de estos servicios, constituyendo una parte esencial para lograr los objetivos para los cuales han sido implementados. Sin embargo, en nuestro país no existe un mecanismo que permita determinar y asegurar de forma objetiva la calidad de los teleinformes.

Con el propósito de elaborar un modelo de auditoría de imágenes y reportes telemédicos, se llevó a cabo una revisión bibliográfica de propuestas similares publicadas en la literatura. A partir de un levantamiento de los procesos actuales de telereportes, en esta AFE se propone un modelo que permite su integración dentro de estos procesos, mediante el uso de herramientas de interoperabilidad. Esto, con el fin de evaluar la calidad de imágenes e identificar fuentes de discrepancias de telereportes. Finalmente, se evalúa su utilización en un Caso de Uso.

Mediante la realización de este trabajo se logró obtener un modelo de auditoría genérico e integrado, que se logra adaptar a los procesos de telereportes y permite determinar de manera

simple y confiable la existencia de fuentes de discrepancias en la calidad. Lo anterior con el propósito de extraer extracción información relevante como insumo para la implementación de políticas de mejora continua.

Si bien este es un primer acercamiento a una herramienta de estas características, se espera que sea un punto de inicio que permita abordar la mejora de los procesos telemédicos, en particular de los teleinformes, que en los actuales contextos de atención, junto al desarrollo de las herramientas de las tecnologías de la información se seguirán incrementando los años venideros.

ABSTRACT

Since it began its implementation in our country, telemedicine has had a sustained growth, especially since the development and improvement of the information technologies. In particular, telereports have contributed to reduce accessibility gaps for the population, and have seen a strong increase in their use since they were implemented. Proof of this has been the appearance on the market of several companies that provide these services, which the state uses through public bids. However, as is inherent in any medical process, quality assurance in the execution and delivery of these services is an essential part of achieving the objectives for which they have been implemented. In our country, there is no such mechanism to objectively determine and assure the quality of these reports.

In order to elaborate an audit model for telemedical images and reports, a bibliographic review of similar proposals in the literature was carried out. A survey of the current telereporting processes was carried out and then a model was proposed that allows its integration within these processes by means of interoperability tools. This in order to evaluate the quality of images and identify possible sources of discrepancies in the telereports. Finally, this equivalent formative activity evaluates the proposed model use in a use case.

By means of this work, a generic and integrated audit model was developed that can be adapted to telereport processes and thus determine, in a simple and reliable way, sources of discrepancies in quality, whose purpose is the extraction of relevant information as input for the implementation of policies for continual improvements.

Although this is a first approach to a tool of these characteristics, it is expected to be a starting point as a tool to address the improvement of telemedical processes, in particular tele-reporting, which in the current contexts of health care together with the development of information technology tools, will increase year after year.

1.- INTRODUCCIÓN

1.2.- ANTECEDENTES

1.2.1.- Telemedicina: Definiciones y situación de Teleinformes en el país.

La Telemedicina puede ser ampliamente definida como toda actividad de salud llevada a cabo a distancia. Según la OMS “es el suministro de servicios de atención sanitaria, en que la distancia constituye un factor crítico, por profesionales que apelan a las tecnologías de la información y de la comunicación con objeto de intercambiar datos para hacer diagnósticos, preconizar tratamientos y prevenir enfermedades o accidentes”. Además, “aplica a la formación permanente de los profesionales de atención de salud y en actividades de investigación y evaluación, con el fin de mejorar la salud de las personas y de las comunidades en que viven” (1). Esto permite una reducción en los tiempos y costos de traslado y a su vez mejora el acceso a especialistas (2). En los últimos años, el uso de la telemedicina ha ido en aumento de la mano de la disponibilidad y mejora en los servicios y equipamiento de telecomunicaciones.

La Telemedicina se puede llevar a cabo en dos modalidades:

i) En tiempo real o modo Sincrónico: Se basa en la interacción en tiempo real y en vivo entre los actores del evento médico.

ii) En tiempo diferido o modo Asincrónico (Store and Forward), permite almacenamiento y transferencia de datos en tiempo diferido, para luego ser enviados a un especialista, junto a una historia clínica del paciente, con el propósito que dicho especialista emita una orientación diagnóstica y terapéutica.

En Chile, la Telemedicina comienza a desarrollarse alrededor del año 1993. Inicialmente fue utilizada para llegar a espacios poblacionales de difícil acceso o geográficamente aislados y, de este modo, mejorar la accesibilidad. Hoy se la considera una herramienta que permite mejorar no solo la accesibilidad y la oportunidad de la atención, sino que también mejorar la eficiencia y la eficacia de los servicios de salud que hacen uso de la telemedicina (2).

En nuestro país, entre los años 2009 y 2013 se fueron implementando estrategias importantes en el ámbito de los Procesos de Apoyo Diagnóstico – Terapéuticos a través de la telemedicina, como son los Teleinformes de exámenes específicos (Teleelectrocardiograma, Teleradiología y Teleoftalmología) y Teleconsultas en atención ambulatoria (Teledermatología). Todas estas estrategias tienen como objetivo común el acceso oportuno a informes de especialidad desde los centros de atención primaria menos especializados o en aquellos donde la disponibilidad de especialistas está limitada debido a escasez de capital humano o porque se requieren fuera de horario convencional. A partir del año 2018 se comienza a implementar la estrategia denominada Hospital Digital con el propósito de expandir la atención de salud a través del uso de tecnologías digitales, con el fin de reducir las listas de esperas por medio de una plataforma virtual inclusiva, sustentable, disponible 24/7, atemporal y sin barreras geográficas, en donde el uso de los recursos telemáticos ha resultado ser esencial (2). A continuación, se describe brevemente cada estrategia y su estado actual en el país.

Teleradiología

Se inicia en Chile el año 2012 y surge ante la necesidad de cubrir brechas existentes entre las horas Radiólogos contratadas y la demanda de informes radiológicos, necesarios en la atención entregada por Unidades de Emergencia y Unidades de Paciente Crítico. De esta forma se hizo posible otorgar cobertura en horario no institucional (posterior a las 14 o 16 horas, según establecimiento) en días hábiles y 24 horas tanto en fin de semana como en días festivos. Esta estrategia requiere la compra de servicios de informes radiológicos y del servicio de Interoperabilidad y Transmisión de Imágenes a través de Licitación Pública, y ha ido incorporando en forma progresiva establecimientos de distintas regiones del país. Es importante señalar que, desde su implementación hasta diciembre de 2017 ya se habían realizado 737.498 informes radiológicos, y 34 establecimientos habían incorporado esta modalidad.

Teleoftalmología

Se inicia el año 2013, con su implementación en las Unidades de Atención Primaria Oftalmológicas (UAPOs). Consiste en la realización de un screening para detectar Retinopatía Diabética a través del examen de Fondo de Ojo, que es informado por un Oftalmólogo, mediante el uso de una plataforma. Funciona en 29 Servicios de Salud. En el año 2013 se implementó teleoftalmología en 14 UAPOs, las que fueron incrementando paulatinamente en forma anual, sumándose 19 en 2014, 24 en 2015, 58 en 2016 y 9 en el 2017. Actualmente, a nivel nacional el servicio de teleoftalmología se encuentra funcionando en un total de 124 UAPOs. Desde su implementación al 31 de diciembre de 2017 ya se habían

realizado más de 300.000 teleinformes de Fondo de Ojo. Por otra parte, sólo durante el año 2017 se realizaron más de 107.516 informes de este tipo.

Teledermatología

Su inicio se remonta al año 2009, como una estrategia que permitiría facilitar el acceso a atención dermatológica procedente de centros de Atención Primaria de Salud. Consiste en el envío de antecedentes clínicos y fotografías por parte de un médico de atención primaria a un especialista dermatólogo, quien evalúa antecedentes y entrega orientación diagnóstica e indicaciones. La Teledermatología se encuentra implementada en 29 Servicios de Salud del país y desde el año 2010 a diciembre de 2017 se han realizado un total de 79.801 teleconsultas.

En este punto, corresponde mencionar que muchos de estos servicios de informes telemáticos son comprados a terceros por el Estado, mediante licitación pública. En la actualidad existen varias empresas que entregan estos servicios en las diversas áreas mencionadas, con un mayor énfasis en Radiología, dada su multiplicidad de modalidades y en menor grado en Oftalmología y Dermatología. En el futuro los Servicios de Salud proyectan comprar y agregar más servicios externos de esta modalidad de exámenes.

1.2.2.- Calidad en salud, en contextos telemédicos

En todo proceso médico que se mantiene de forma periódica, y del cual los servicios telemédicos no están exentos, uno de los aspectos más relevantes tiene relación a llevarlos a cabo en un marco de calidad. La percepción de calidad dentro de cualquier proceso de atención va a ser diferente dependiendo de quien sea el actor involucrado, es decir, la calidad percibida por el consumidor puede estar centrada en aspectos que para quien provee dicha atención puede ser irrelevante. Por lo tanto, existe un componente subjetivo respecto a la apreciación de la calidad.

La norma ISO 9000-2015 (3) define la calidad como “el grado en que un conjunto de características de un objeto cumple con los requerimientos”, entendiéndose como característica a un rasgo distintivo y a requerimiento a una necesidad o expectativa declarada, generalmente implícita u obligatoria.

Esta norma identifica once características básicas de calidad que son relevantes en las organizaciones de salud. La evaluación del cumplimiento de los requerimientos de calidad de la organización necesita la especificación de las características que se debe tener para cumplir estos requerimientos. Además de una serie de aspectos que la organización considere importante para determinar sus requerimientos de calidad para los resultados de sus procesos clínicos. Los aspectos de calidad identificados desde esta perspectiva se pueden resumir en las siguientes 6 dimensiones (Figura 1):



Figura 1. Esquema de las dimensiones de la calidad en salud.

- ***Efectividad:*** Prestar una asistencia basada en evidencia y que permita mejorar los resultados de salud de las personas y las comunidades en función de las necesidades.
- ***Eficiencia:*** Prestar asistencia de forma que se maximice el uso de los recursos y se evite el desperdicio.
- ***Accesibilidad:*** Prestar una asistencia sanitaria oportuna, geográficamente razonable y en un entorno en el que las competencias y los recursos sean adecuados a las necesidades médicas.

- ***Centrado en el paciente:*** Prestar una asistencia que tenga en cuenta las preferencias y aspiraciones de los usuarios individuales de los servicios y las culturas de sus comunidades.
- ***Equidad:*** Prestación de una asistencia que no varíe en calidad debido a características personales como el sexo, raza, etnia, ubicación geográfica o situación socioeconómica
- ***Seguridad:*** Prestar una asistencia que minimice los riesgos y los daños a los usuarios.

En Chile existen programas de acreditación y certificación de las instituciones de salud, que han permitido elevar de forma importante los estándares de calidad con el fin de resguardar la seguridad en la atención del paciente. En particular, respecto a los procesos y provisión de servicios telemédicos, a partir del año 2019 y en el contexto pandémico COVID-19, son publicados dos textos guías con recomendaciones (4) y lineamientos para el buen desarrollo de la telemedicina en Chile (5). Dichos textos establecen criterios mínimos para el buen funcionamiento de una consulta telemédica, así como también las responsabilidades legales en la provisión de los servicios telemédicos en los diversos contextos de atención. Establecen, además, la necesidad de contar con políticas y un plan de calidad que tome en consideración todas las etapas y ciclo de vida del servicio, las que van desde la identificación de las necesidades, diseño y planificación, hasta el desarrollo, implementación y ejecución, incluyendo el monitoreo, medición y evaluación continuos con el propósito de alcanzar los mejores resultados posibles.

En el contexto de la atención mediante telereportes, uno de los aspectos importantes y relevantes de evaluar de forma constante, en función de los aspectos de calidad mencionados, es la efectividad en alcanzar la tarea de reportar de forma correcta cumpliendo con los

requerimientos de calidad. Para esto es necesario definir o establecer un mecanismo de evaluación que permita medir la calidad obtenida para diferenciarla de la mejor calidad posible. Para ello es necesario contextualizar y precisar respecto a lo que se entiende con error en la acción de interpretar en base la observación de imágenes médicas. De acuerdo a Brady et al (6), el error en un informe radiológico se refiere a cualquier discrepancia de interpretación que se desvíe sustancialmente del consenso de los pares. Sin embargo, la palabra error, según el autor, se presta para interpretaciones absolutas y considerando además que en medicina las opiniones no son binarias sino más bien opiniones y apreciaciones subjetivas, es que el término que se propone emplear es el de discrepancia. “Cualquier discrepancia de interpretación que se desvíe sustancialmente del consenso de los pares es una definición razonable y comúnmente aceptada de error radiológico interpretativo” cita textual de Brady et al (6). No obstante, es necesario enfatizar que esto no quiere decir que los errores no ocurran, pero sí que hay que diferenciar entre lo que puede ser una discrepancia, ante la falibilidad inherente a toda actividad humana en ambientes de alto riesgo como los procesos médicos, y los errores inexcusables o negligencias propiamente tales.

Las fuentes de discrepancias más frecuentes en informes médicos son las que tienen que ver con el tipo de percepción, tales como sub-reportar hallazgos clínicos y/o características clínicas significativas o, por el contrario, sobre-reportar características de la imagen que pudieran tener significancia. Y esto es algo que no solamente ocurre en radiología sino que también se observa en otras especialidades (4,5). En Chile hay estudios que abordan esta temática y coinciden en que más del 30% de las discrepancias, se deben a problemas de percepción y en menor medida de juicio/criterio (7).

Desde el punto de vista técnico, la captura de la imagen también tiene un importante grado de relevancia, ya que una mala técnica y/o mala calidad de imagen, debido ya sea a mala praxis, equipamiento defectuoso, sobrecarga de trabajo, etc., puede llegar a afectar significativamente el juicio del profesional que realiza el informe y por ende su resultado (6,8).

Actualmente en nuestro país, dentro de los procesos de teleinformes lo único que existe como referencia por parte de entes proveedores como indicador de calidad en función de una cierta discrepancia, es el porcentaje de *addendum* clínicos realizados en un periodo de tiempo (9). Aun no existen estudios que evalúen servicios de teleinformes y que midan su calidad considerando aspectos de discrepancia en reportes y cuáles serían sus implicancias.

1.2.3.- Herramientas para el aseguramiento de la calidad y mejora continua.

El aseguramiento de la calidad en su forma amplia se puede definir como el conjunto de acciones que toman las instituciones con el objetivo de poder entregar a los consumidores bienes o servicios con el nivel de calidad esperada. Para poder determinar de objetivamente la calidad otorgada, se deben llevar a cabo acciones específicas para aplicar controles de calidad que sean universalmente aceptados, así como las técnicas y actividades de carácter operativo utilizadas en la verificación de los requisitos relativos a determinar si la calidad de producción corresponde a la calidad del diseño. Estas acciones deben permitir medir la Calidad Real, compararla con lo esperado y actuar sobre las diferencias.

Las auditorías médicas nacen ante esta necesidad, y junto a ello debido a la incurrancia de errores que ha llevado a la judicialización de la medicina, aumentando con ello de manera considerable costos y gastos. *Rencoret et al.* (10) las describe como un proceso que no se aparta de las auditorías que se realizan en otros campos del quehacer humano en cuanto a proceso, objetivos, evaluación e informe, constituyéndose en un instrumento de perfeccionamiento y educación continua, que permite tomar resguardos en aspectos técnicos, administrativos, éticos y médico-legales, en búsqueda de la excelencia.

Una auditoría consiste en una evaluación sistemática, realizada por médicos, que observa y compara las características o calidad de la atención brindada con la calidad ideal y deseada, acorde a criterios y normas preestablecidas. Se concibe como una instancia educadora más que punitiva y sancionadora, donde se trata de aprovechar las experiencias, tanto las negativas, a través del análisis de incidentes, como de las positivas para ir introduciendo en forma gradual correcciones y mejoramiento en cada uno de los procesos de atención, para satisfacer de mejor manera los requerimientos de los pacientes. La auditoría de la atención en salud puede ser realizada bajo dos tipos de modalidades:

Auditoría Interna: Realizada por la misma institución que presta los servicios clínicos.

Auditoría Externa: Responsabilidad a cargo de una organización independiente de la auditada, la cual posee una ventaja comparativa ya que dada su independencia esta auditoría puede ser de calificación, recalificación, de mantenimiento o periódica de seguimiento (11).

Revisión por pares como herramienta de auditoría

La revisión por pares se define como una reflexión continua, sistemática y crítica por parte de varios proveedores de atención, sobre su propia actuación y la de sus colegas, utilizando procedimientos estructurados, con el objetivo de lograr una mejora continua de la calidad de la atención. La revisión por pares también puede ser interna o externa (12).

1.2.4.- Problema/Oportunidad

En Chile, desde el año 2012 en adelante ha habido una importante expansión en la incorporación y uso de recursos telemédicos, sobre todo en lo concerniente a informes médicos. Esto, en respuesta a los buenos resultados que, desde su implementación, se han obtenido, tanto en cobertura y resolutivez así como en tiempos de respuesta (2). Una consecuencia importante de la incorporación de los servicios de telemedicina ha sido el disminuir en una importante medida el traslado de pacientes desde la atención primaria hacia centros de mayor complejidad. Lo anterior ha contribuido a disminuir la saturación de estos establecimientos, evitando así el aumento de las largas listas de espera actualmente existentes.

Si bien, dentro de los lineamientos para el desarrollo de la telesalud, los entes proveedores de servicios telemédicos contemplan la gestión de calidad de sus procesos (5), a la fecha el ente regulador contratante no se ha pronunciado sobre herramientas o modelos de aseguramiento de la calidad que permitan auditar la entrega de los diferentes servicios. Este punto cobra particular relevancia en aquellos servicios que tienen relación con el uso de imágenes médicas.

Una revisión de la literatura en el área demuestra que se han descrito propuestas y medidas para el control y aseguramiento de la calidad, en diversas especialidades (10,13). Estas se basan principalmente en la modalidad Store and Forward, mediante modelos de auditoría de estos procesos, con el fin de implementar mejoras de forma sistemática y constante, que apunten a reducir discrepancias que puedan redundar en un daño en la salud de las personas (14–16).

En la presente AFE se abordó el diseño de un modelo de auditoría que reúne tres importantes condiciones: Primeramente, que sea genérico, es decir, un modelo común y adaptable a cualquier modalidad de teleinformes basados en imágenes médicas. En segundo lugar, que sea integrado, es decir, capaz de reunir en un solo sistema diferentes herramientas informáticas que permitan interactuar con los otros sistemas para auditar y por último como condición necesaria y habilitante para las dos anteriores, que esté basado en estándares de interoperabilidad.

El trabajo abordó de manera particular los teleinformes basados en imágenes médicas, considerando en primer lugar el sostenido crecimiento que han experimentado desde que fueron implementados y tomando además en consideración la oportunidad que presenta la falta de una herramienta que permita controlar de manera objetiva su calidad, con miras a detectar falencias que permitan emprender acciones de mejora continua, en forma pertinente y oportuna. Todo lo anterior en consonancia con las exigencias previamente mencionadas orientadas a la gestión de calidad y gestión de riesgos.

A través de la realización de este trabajo se espera aportar al área de la salud con una herramienta de auditoría que contribuya a la gestión de calidad de aquellos procesos telemédicos que involucren el uso remoto de imágenes clínicas. Recoger esta oportunidad

significa abordar el desafío de proponer enfoques innovadores alineados a las exigencias actuales de la sociedad, incorporando las herramientas y ventajas que aportan las tecnologías de la información.

2.- OBJETIVOS

OBJETIVO GENERAL

Desarrollar, implementar, testear y analizar un modelo genérico e integrado de auditoría de informes telemédicos, realizados por servicios externos, basados en imágenes médicas.

OBJETIVOS ESPECÍFICOS

- 1.- Identificar y seleccionar estrategias para la elaboración de un modelo de auditoría de reportes basados en imágenes clínicas.
- 2.- Modelar los procesos existentes y el que se propondrá, con herramientas de modelado de procesos Business Process Model and Notation (BPMN).
- 3.- Definir las métricas a utilizar para la medición objetiva de la concordancia o nivel de discrepancia de los reportes a auditar, estableciendo los posibles orígenes de las discrepancias.
- 4.- Establecer un mecanismo de integración en los flujos de trabajo de las distintas modalidades de teleinformes.
- 5.- Aplicar el modelo elaborado como una herramienta de caso uso de reportes basado en imágenes médicas y analizar los resultados obtenidos.

3.- MARCO TEORICO

3.1.- Revisión Bibliográfica

Para una adecuada comprensión de lo realizado en este trabajo AFE, se realizará una breve descripción de los diferentes aspectos revisados en la literatura, junto a conceptos y elementos necesarios para llevar a cabo lo que se plantea en los objetivos de este trabajo.

3.1.1.- Métodos de medición de la calidad de la imagen.

Desde el punto de vista fotográfico, la calidad de una imagen se puede definir como la capacidad que tiene dicha imagen para representar el objeto original; es decir, constituye una medida de semejanza. Visto desde un punto de vista clínico, la calidad de una imagen se definiría en base a aquellas características que la hagan capaces de proporcionar una alta sensibilidad para detectar una patología y, a su vez, una alta especificidad para identificar estructuras anatómicas normales (17).

La evaluación de la calidad de la imagen puede clasificarse en seis niveles (18):

- 1.- Capacidad técnica.
- 2.- Precisión diagnóstica.
- 3.- Impacto diagnóstico.
- 4.- Impacto terapéutico.
- 5.- Resultado para el paciente.
- 6.- Uso óptimo (coste-beneficio, costo-eficacia).

Los estudios de calidad que comprenden los niveles de 3 al 6 son escasos en la literatura. La mayoría de los estudios existentes se limitan a abordar ya sea propiedades físicas de la imagen o el rendimiento de los observadores, es decir, los niveles 1 y 2.

Un punto de vista razonable sería que, sea cual sea la medida de la calidad de la imagen, esta debería incluir al menos una correlación con la opinión de los evaluadores sobre su calidad (19–21); en otras palabras, la calidad de una imagen debe ser tal que permita que el evaluador sea capaz de:

- i. distinguir entre varios estados de enfermedad y salud,
- ii. informar con precisión de las estructuras y características relevantes para el diagnóstico,
- iii. clasificar con precisión los distintos tipos de anomalías, o
- iv. detectar de forma confiable estructuras relevantes en la imagen.

En la imagen médica, la información clínica del paciente y sus posibles anomalías es transferida al evaluador principalmente en dos pasos, i) mediante la adquisición de los datos y formación de la imagen; y ii) mediante el procesamiento y visualización de esta. El primer paso depende principalmente de las características técnicas y físicas del equipo que se utiliza para la captura de imagen, en tanto que el segundo incluye la vital importancia del evaluador y su interpretación de las estructuras observables en la imagen. Por lo tanto, en el presente trabajo la calidad del proceso de adquisición de imágenes ha sido evaluada considerando estos dos pasos.

Existen diferentes metodologías para evaluar la calidad de una imagen, las que pueden ser agrupadas en métodos físicos, psicofísicos y de observación. Cada uno de ellos presenta

ventajas y desventajas, de acuerdo a los diversos aspectos que concierne la evaluación de una imagen (21–23). Los métodos físicos y psicofísicos son métodos de evaluación no clínicos y de costosa implementación, por lo que escapan a los objetivos de este trabajo. No obstante, se entrega a continuación una breve descripción de ambos.

Métodos Físicos

Se basan en la medición de la resolución, ruido y cálculos entre la relación de señal y ruido. Son metodologías complejas de evaluar en un contexto de auditoría, que evalúan básicamente la performance de los detectores propios de los equipos de imagen. Si bien son metodologías importantes para el análisis en la formación de imágenes, esta se limita específicamente sólo a dicho paso.

Métodos Psicofísicos

Este tipo de métodos contempla la evaluación de un observador frente a la presentación de ciertos estímulos visuales gruesos. Si bien existen estudios que demuestran una correlación de sus resultados con aspectos clínicos de evaluación, su implementación es costosa y sólo se aplica con fantomas o maniqués de prueba.

Con respecto a los métodos de observación, estos tienen una mayor relevancia para el presente trabajo, ya que consideran de forma subjetiva la valoración en base a la observación por evaluadores de los atributos de la imagen, tomando en consideración aspectos clínicos en la observación de estructuras anatómicas (20,21,23–26). A continuación, se entrega una descripción detallada de esta metodología, incluyendo sus principales herramientas.

Métodos de Observación

En los métodos de observación la evaluación visual de la calidad de la imagen juega un rol esencial. Los dos principales enfoques de este método recaen en la capacidad de detección de la patología, por una parte y en la evaluación de la visibilidad de estructuras anatómicas, por otra. Para evaluar el desempeño del observador en la distinción de patologías, el análisis típicamente utilizado es el Receiver Operating Characteristics (ROC); mientras que para la visualización de estructuras anatómicas se utiliza el Análisis de Graduación Visual (VGA por sus siglas en inglés).

Análisis ROC

El análisis ROC es considerado uno de los métodos gold standard para la evaluación de calidad de imagen, ya que permite evaluar la imagen en términos de su capacidad para demostrar la existencia de anormalidades. No obstante, presenta algunos inconvenientes ya que es un método que consume tiempo, ya que implica un conocimiento previo de la existencia de cierta anormalidad, y a la vez requiere de un número importante de imágenes para darle mayor robustez al estudio (20). Por lo tanto, en este trabajo se concluye que como herramienta de estudio para el sistema de auditoría, el análisis ROC no revertiría una utilidad sustancial, debido a los inconvenientes previamente descritos, además de lo binario de su resultado, ya que no consideraría la visualización de otras estructuras y se limitaría a la presencia o ausencia de una señal en particular.

Análisis de Graduación Visual (VGA)

Este método se basa en la premisa que el grado de visibilidad de la anatomía normal se correlaciona fuertemente con la capacidad de detectar estructuras patológicas (21). Como metodología de medición de calidad de imágenes es más simple e intuitiva que las dos anteriores y por ende más rápida y costo efectiva. En términos generales, consiste en la graduación de la visualización de estructuras en las imágenes, realizadas por observadores utilizando escalas de puntuación absolutas o relativas. La facilidad de implementación de este tipo de metodologías la hacen viable para su uso en estudios en ambientes clínicos.

En este punto es necesario señalar que la correlación entre los métodos físicos y de observación ha sido estudiada a lo largo de los años (19,20). Se ha encontrado una fuerte correlación en la calidad de imagen entre las pruebas físicas y las de observación. Este hecho permitiría concluir que un análisis de calidad o de mejora de la calidad tendría un impacto similar si es evaluado con uno u otro método (19). Esto es muy relevante, ya que dada la simpleza del método VGA este puede ser implementado con relativa facilidad dentro de ambientes clínicos.

Cabe hacer notar que existen dos tipos de VGA, absolutos y relativos. En los absolutos, las imágenes se evalúan de forma aislada sin que se requiera una imagen comparativa de referencia. Su escala de valoración por lo general se responde en puntajes de 1-3 o de 1-5; en donde 1 significa “no se reproduce” o “la estructura no se puede discernir” y 5 (ó 3) significa “muy bien reproducida” o “la estructura se distingue completamente en sus formas”. Una ventaja de este método es que tiene posibilidades estadísticas, sin embargo, al no tener una imagen de referencia puede dar lugar a una alta variabilidad tanto intra como inter observador.

En el análisis VGA relativo o comparativo, la evaluación se realiza graduando la visibilidad de las estructuras en una imagen experimental, en comparación con las mismas estructuras de una imagen de referencia. El observador puede graduar la visibilidad de cada estructura utilizando una escala en donde “0” puede implicar igual visibilidad, mientras que un valor negativo implica menor visibilidad, y a la inversa con números positivos. La ventaja de este método es su utilidad para estudios de optimización de dosis en estudios radiológicos. Su desventaja radica en las escasas posibilidades estadísticas de sus datos y el hecho de que toda la puntuación dependa de una imagen de referencia (21,23–26).

VGA basado en criterios de Imagen

A partir del año 1996 un grupo de radiólogos y físicos médicos desarrollaron criterios para el diagnóstico por imágenes, los que posteriormente pasaron a formar parte de la European Guidelines for Image Quality (24). Estos criterios fueron desarrollados como guías para asegurar la calidad de la imagen en cada una de las evaluaciones específicas, tanto para radiografías como para tomografías computarizadas. A partir de estos criterios de calidad de imágenes este método de evaluación de imágenes se ha sustentado, ya que se basan en el trabajo clínico lo que le otorga validez para ser comparado con otros estudios VGA basados en los mismos criterios (23).

El análisis VGA basado en criterios de imagen, se basa en el grado de cumplimiento de cada uno de dichos criterios. Esta graduación puede ser elaborada según lo que se pretenda responder en cada estudio, pero por lo general va en una escala de 1 a 5, en donde además se puede agregar una pregunta final “si la imagen evaluada es posible ser usada para propósitos

diagnósticos”. De esta manera, cada imagen será evaluada en cada uno de sus criterios predefinidos en dicha escala (ver Tabla n°1 como ejemplo). Bajo este esquema de evaluación, es posible establecer un método para evaluar cualquier imagen médica, determinando previamente los criterios validados de imagen para evaluar y graduar la completitud de dichos criterios.

Tabla 1. Ejemplo de Criterios de calidad de imagen de Rx de columna AP basados en European guidelines on quality criteria for diagnostic radiographic images.

Criterios para columna lumbar AP	
C1	Reproducción visualmente nítida de las superficies de la placa superior e inferior
C2	Reproducción visualmente nítida de los pedículos.
C3	Reproducción visualmente nítida de la corteza y estructuras trabeculares.
C4	El ruido de la imagen no interfiere con mi evaluación clínica.
C5	La calidad general de la imagen es suficiente para el diagnóstico

Fuente: Månsson LG. Methods for the evaluation of image quality: A review (18)

La puntuación final del análisis VGA se obtiene a partir de la suma del grado de cumplimiento de cada criterio, para cada imagen, por cada observador (VGA Score o VGAS) como en la siguiente formula (25):

$$VGAS = \frac{\sum OI S_c}{N_i N_o}$$

S_c : Para observadores (O) e imágenes (I)

N_i Número total de imágenes

N_o Número total de observadores

La utilización estadística de los resultados del VGA Score tiene una validez cuestionable, ya que se basa en el análisis de datos ordinales cuya distribución es desconocida. No obstante, como herramienta ha sido de gran utilidad para el análisis de correlación con otras metodologías o para la validación de nuevas técnicas en imagenología (19,25,27).

Cabe señalar que el número de imágenes necesarias para realizar un estudio con la metodología VGA depende del tipo de estudio y lo que se pretenda estudiar (20). Así, por ejemplo, en estudios de calidad de imágenes con fantomas o cadáveres la variabilidad entre una imagen y otra es pequeña, por lo que no es necesario contar con una gran cantidad de imágenes. Sin embargo, en estudios realizados con pacientes hay más factores que podrían implicar la necesidad de contar con una mayor cantidad de imágenes. En este caso, los factores más relevantes podrían corresponder a diferencias anatómicas entre pacientes, o a diferencias en la técnica de adquisición (20).

Otro de los aspectos relevantes tiene que ver con la cantidad de observadores participantes en el estudio, lo que a su vez depende de lo que se pretenda estudiar y de la experiencia de los evaluadores. Basándose en lo reportado en diferentes publicaciones, la cantidad ideal de observadores es del orden de 10 personas. No obstante, el autor recomienda que debido a la realidad de los servicios el mínimo debería ser 3, aunque recomienda 5. El mismo autor también recomienda considerar el factor variabilidad intra e interobservadores como un control de calidad del estudio (20).

Para la evaluación de las imágenes y su graduación, existen plataformas informáticas desarrolladas con el fin de agilizar la elaboración de estudios. En su funcionamiento se contemplan estrategias para la eliminación de eventuales sesgos que se podrían generar durante la evaluación por los observadores. Una de las estrategias más utilizadas para este propósito es ViewDex, desarrollada por la Universidad de Gotemburgo, Suecia (26).

En Oftalmología, específicamente, para la evaluación de la calidad de imágenes de fondo de ojo de pacientes con retinopatía diabética, se desarrollaron estrategias y criterios de calidad que permitieron evaluar si la calidad constituía un criterio suficiente para la valoración del estado de salud retinal (28,29). Mediante el estudio de las falencias que presentaban algunas imágenes fue posible establecer pautas de mejora para la obtención de imágenes por parte de los operadores de equipos (30). Con el tiempo, ante el surgimiento del procesamiento automatizado de imágenes mediante inteligencia artificial (IA), la evaluación de la calidad ha migrado del uso de la valoración humana de imágenes, al uso de metodologías automatizadas de la calidad basadas tanto en *Machine Learning* como en *Deep Learning*. (31–33). Esto ha permitido establecer, para esta área en particular, una metodología automatizada para la evaluación de la calidad. Considerando el desarrollo vertiginoso de los avances en el campo de la IA, resulta razonable suponer que similares enfoques se lleguen a aplicar para evaluar la calidad de imágenes de otras especialidades médicas, en concordancia con el actual desarrollo de herramientas de soporte a la decisión clínica.

Para el estudio de la calidad en otras especialidades, como por ejemplo Dermatología, se hace necesario el desarrollo de criterios de calidad de imágenes que sean consensuados y validados entre especialistas expertos. En su defecto, ante la ausencia de dichos criterios, una forma de medir su calidad sería con herramientas de graduación visual relativa, tomando imágenes de

referencias para medir su calidad, y estableciendo una graduación en función de la visualización de ciertas características particulares de la imagen de referencia (20).

3.1.2.- Revisión por pares

Dentro de los diferentes tipos de revisiones por pares existen variadas metodologías que han podido ser cotejadas en la literatura, de las principales y más frecuentes usadas en la práctica habitual están:

La doble lectura, es un método prospectivo en el que se asigna a un segundo radiólogo la revisión de un estudio de imagen recién realizado para identificar rápidamente errores clínicamente significativos y mejorar la precisión (34,35).

La observación incidental, este método aprovecha las conferencias interdisciplinarias y las consultas del equipo clínico registrando todos los casos discrepantes encontrados durante el flujo de trabajo diario, en la consulta clínica con colegas, en conferencias de revisión de casos clínicos y en conferencias interdisciplinarias de análisis de bases de datos de revisión por pares (36).

La auditoría profesional, es un proceso retrospectivo en el que la interpretación de las imágenes se compara con un estándar de referencia, como los resultados de la patología, los hallazgos operativos o el seguimiento clínico, y puede reducir la subjetividad de la revisión por pares.

La revisión centrada en la práctica, esta metodología incorpora los resultados comunicados por los pacientes y otros profesionales de la salud, tiene la ventaja de utilizar información clínicamente relevante para mejorar la práctica.

Todas estas metodologías suponen un trabajo exhaustivo ya que al incorporar elementos que involucra distintos estamentos la hace más costosa y laboriosa de implementar

RADPEER

Luego de la publicación del informe del Institute Of Medicine (IOM por sus siglas en inglés) de Estados Unidos “To err is human”, en marzo del 2000 (37) y al desarrollo de programas de acreditación médica por parte de los entes reguladores en dicha nación, la preocupación por mantener y certificar las competencias médicas en aumento. Para el mantenimiento de la certificación, definieron cuatro aspectos fundamentales, los cuales son (14):

- i. El prestigio profesional.
- ii. El compromiso con el aprendizaje permanente y la autoevaluación periódica.
- iii. La experiencia cognitiva.
- iv. La evaluación del rendimiento en la práctica.

La American College of Radiology (ACR) analizó diversos programas de evaluaciones de pares y llegó a la conclusión que, para cumplir el cuarto requisito de mantenimiento de las competencias, se debería llevar a cabo un programa de revisión por pares a nivel nacional, uniforme en estructura y función en todas las prácticas, preciso, fácil, no punitivo y capaz de integrarse a los programas de calidad de un centro. Esta fue la motivación que impulsó el desarrollo del programa de revisión por pares RADPEER.

El sistema de revisión RADPEER se basa en la premisa que cada vez que se interpreta un nuevo estudio de imágenes para compararlo con uno más antiguo, esto constituye una instancia óptima para la revisión por pares de la exactitud de la interpretación del estudio

antiguo. De esta manera, si se evalúa basado en un sistema de puntuación, es posible tabular los resultados de estas evaluaciones contando de esta manera con un sistema que cumpla con los objetivos de la revisión por pares.

Para el análisis de la información se considera la tasa de discrepancia, cociente en donde en el numerador se consigna la cantidad de evaluaciones con cada una de las calificaciones y en el denominador el total de informes evaluados. Estos indicadores permiten evaluar por modalidad, especialidad, radiólogo, etc.

Cambios en el tiempo

Desde su implementación en el año 2005, el sistema de revisión por pares comenzó a ser cada vez más utilizado. En el año 2007 se sumó la Joint Commission en donde los objetivos de evaluación continua de la práctica profesional se acoplaron con los resultados del RADPEER (31). Con esto, la expansión del sistema de revisión por pares fue cada vez más rápida ya que adopta nuevas implicancias.

Transcurridos los 5 primeros años de uso, en el año 2009, se llevó a revisión el sistema de evaluación por pares (38), ya que hubo controversia respecto al significado de las categorías y a que éstas se prestaban para confusión. Así, el sistema fue adoptando enfoques de otros sistemas de revisión por pares en medicina, específicamente en radiología, en donde se simplificó el significado de las categorías y se agregó el de la significancia clínica ya que tiene más impacto en la seguridad del paciente.

En el año 2015 en el marco de revisión por seguimiento de los reportes de errores de diagnóstico por el IOM, dicho organismo citó en su informe que estos errores contribuyen en un 10% de las muertes de pacientes (39). Como objetivo declarado en ese informe estaba

"desarrollar y desplegar enfoques para identificar, aprender y reducir los errores de diagnóstico y los cuasi errores en la práctica clínica". Bajo este contexto, y ante la revisión periódica del sistema de revisión por pares, es que se le han ido realizando nuevas actualizaciones, con el afán de estar en línea con el objetivo planteado por el IOM (40).

Dentro de los cambios principales están:

Cambio en el sistema de puntuación. En el caso de hallazgos discrepantes con puntuaciones de 2b o superiores, los comités de revisión de la institución se reúnen para mantener o modificar la puntuación final. Con el propósito de darle agilidad a la resolución de estos casos y considerando que no aportan en el aprendizaje adicional, se decide dejar la puntuación 3 y 4 en una sola puntuación discrepante. De esta manera, pasa de ser un sistema de 4 puntos a uno de 3, manteniendo la significancia clínica en los puntos 2 y 3 (Ver Tabla n°2).

Tabla 2. Tercer y último esquema de puntuación de evaluación de desempeño interpretativo de reportes RADPEER.

Puntuación	Significado	Opcional
1	Coincide con la interpretación	
2	Discrepancia en la interpretación/no se espera que se haga normalmente (fallo comprensible)	a. Es poco probable que sea clínicamente significativo b. Es probable que sea clínicamente significativo
3	Discrepancia en la interpretación/debería hacerse la mayoría de las veces	a. Es poco probable que sea clínicamente significativo b. Es probable que sea clínicamente significativo

Fuente: Goldberg-Stein S, Frigini LA, Long S, Metwalli Z, Nguyen XV, Parker M, et al. ACR RADPEER Committee White Paper with 2016 Updates: Revised Scoring System, New Classifications, Self-Review, and Subspecialized Reports. J Am Coll Radiol. (34)

Estratificación de casos por edad, tipo de examen y tipo de discrepancia. Para proporcionar más información sobre las discrepancias identificadas en la interpretación, se propuso clasificarlas. Las categorías propuestas, basadas en orden de frecuencia son: i) error de percepción, ii) error de interpretación y iii) error de comunicación. Esta clasificación la realiza el revisor o bien el comité revisor, dando como resultado un informe más específico, esto basado en que estas categorías son las responsables de la mayor cantidad de demandas entabladas contra médicos radiólogos (40).

i) Error de percepción se refiere a un error cognitivo, que se produce cuando el médico no identifica o no describe una anomalía en la imagen diagnóstica. Son identificados de forma retrospectiva y representan la mayor parte de las discrepancias clínicamente significativas en los informes.

ii) Error de interpretación es también un error cognitivo que se produce cuando el médico identifica y describe correctamente la anomalía, pero no identifica su importancia o impacto clínico. Estos errores se deben más bien a una valoración o juicio incorrecto y según Renfrew et al. (41) representan un 26% en una revisión realizada en base a una serie de 182 errores.

iii) Error de comunicación es un fallo en la comunicación de un resultado de un estudio. Puede deberse a un informe poco claro o ambiguo y/o que no comunica de forma clara las recomendaciones del médico evaluador. También pueden deberse a que hallazgos de carácter urgente no se comuniquen adecuadamente, de forma efectiva o en un plazo de tiempo adecuado. Representa entre un 4 a un 10% de ocurrencia, según distintos autores (6,41,42).

La estratificación de los casos, no sólo por modalidad, sino también por tipo de examinación, edad del paciente y tipo de discrepancia, permitiría a los participantes contar con más

información, con el afán de focalizar los esfuerzos en la mejora de aquellas áreas con debilidades identificadas. La clasificación de discrepancias en combinación con otros modificadores puede ayudar a la disminución de discrepancias, así como a resolver los problemas de comunicación antes mencionados.

Enfoque centrado en el aprendizaje más que un acto punitivo.

A partir del año 2016 el enfoque del sistema está en migrar de una metodología tradicional de puntuación a un modelo más colaborativo basado en el aprendizaje. El temor a la acción punitiva obstaculiza la participación plena, auténtica y sincera. Se estudian metodologías de cambio en la cultura de percepción de la herramienta para que sea percibida como una oportunidad de mejora constante. Esto para que esté en consonancia con las recomendaciones del informe reciente del IOM, que aconseja desarrollar y desplegar enfoques para reducir los errores de diagnóstico y los cuasi errores (39).

Diagrama del proceso final del sistema de revisión por pares RADPEER

El proceso se inicia con la recepción de nueva imagen para informar por especialista con examen y reporte previo. Ante esto el revisor actual decide antes de informar revisar o no el reporte previo. Si el resultado de su evaluación es 1 o 2a el resultado de dicha evaluación se aloja en una base de datos. De lo contrario va a comité para reevaluación en donde confirma o recalifica la evaluación. Si se confirma la evaluación el comité envía un reporte de retroalimentación a un especialista evaluador (ver Figura n° 2).

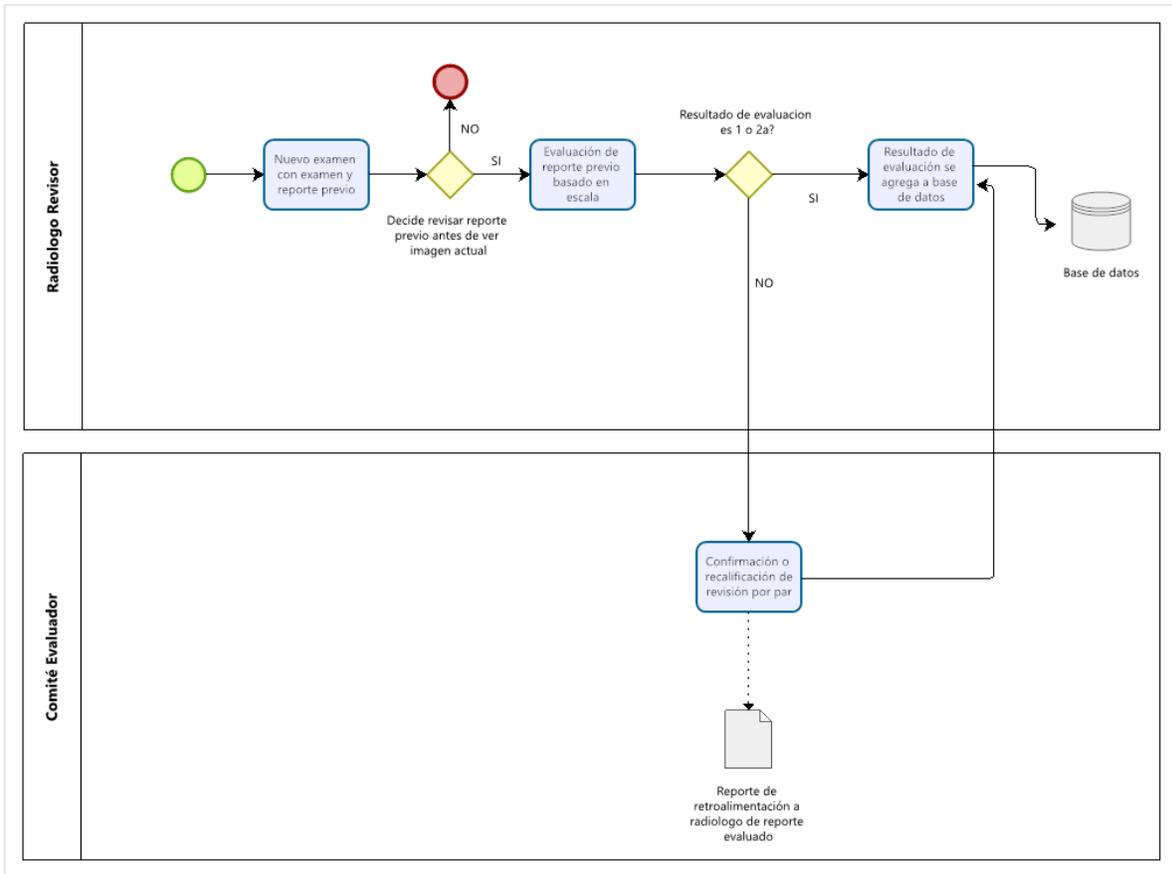


Figura 2. Representación de actual proceso de evaluación de reportes RADPEER. Se representa el diagrama del proceso final del actual sistema de revisión por pares RADPEER. Fuente: Elaboración propia basado en la última actualización del sistema.

3.1.3.- Interoperabilidad

Dentro de los lineamientos de la telesalud, es importante considerar que toda estrategia que se desee implementar debe considerar la interoperabilidad con otros sistemas de información, como por ejemplo, los registros clínicos electrónicos y el acto de remitir datos a sistemas de repositorios de información. Es, por lo tanto, necesario precisar y definir el concepto de

interoperabilidad, así como sus implicancias y el rol de los estándares para lograr el objetivo antes mencionado.

El concepto más ampliamente usado de interoperabilidad ha sido definido por el *Institute of Electrical and Electronics Engineers* (IEEE), que señala que es la habilidad de dos o más sistemas o componentes, de intercambiar información y usar la información que ha sido intercambiada. Cabe hacer hincapié que esta definición incluye por una parte el intercambio de información, es decir, la interoperabilidad técnica y por otra, la capacidad de quien recibe de usar dicha información, es decir, la interoperabilidad semántica. Un tercer concepto que pertenece al uso de la información es la interoperabilidad de procesos, ámbito en el cual se puede incluir la interoperabilidad clínica. A continuación se describen brevemente estos tres ámbitos de la interoperabilidad:

Interoperabilidad Técnica: Se refiere al movimiento de datos desde un punto A hacia un punto B, neutralizando los efectos de la distancia, es decir obtener información 100% confiable de un canal completamente ruidoso. Es la capa tecnológica, por lo tanto es independiente del dominio y no requiere conocer el significado de lo que se intercambia.

Interoperabilidad Semántica: Hace referencia a que la entidad que envía y la que recibe la información necesitan entenderla de la misma forma. Permite a los computadores compartir, entender, interpretar y usar la información sin ningún tipo de ambigüedad. La interoperabilidad semántica es específica al dominio y contexto por lo que requiere del uso códigos e identificadores libres de ambigüedades.

Interoperabilidad de procesos: Es alcanzada cuando grupos humanos comparten un entendimiento común a través de una red, o sistema de negocios. En el ámbito de la salud, el

foco está centrado en la interoperabilidad clínica, que es un subtipo de la interoperabilidad de procesos. Esta se define como la habilidad de dos o más clínicos en diferentes contextos de transferir información o datos clínicos de pacientes y proveerles el mismo cuidado (43).

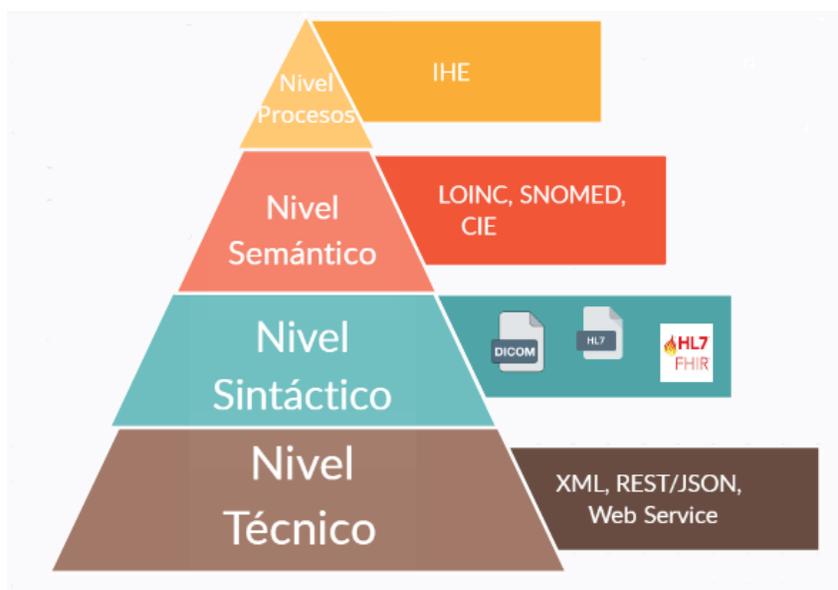


Figura 3. Representación esquemática de los niveles de interoperabilidad en salud con los respectivos estándares más característicos basado en la definición de European Telecommunication Standards Institute (ETSI). Fuente: https://iris.paho.org/bitstream/handle/10665.2/28188/9789275318812_spa.pdf

3.1.4.- Estándares

Con el propósito de lograr la interoperabilidad y que esta se lleve a cabo de manera efectiva se requiere la utilización de estándares. A continuación, se describe brevemente qué es y de dónde viene un estándar y porqué su utilización es tan necesaria, sobre todo en el campo de la salud. La *International Standardization Organization* (ISO) define una norma o estándar

como un documento establecido por consenso y aprobado por un organismo reconocido, cuyo propósito es proporcionar reglas, directrices o características para el uso común y repetido de actividades o resultados, con el fin de alcanzar el grado óptimo de orden en un contexto determinado (44).

Dos de los términos claves en dicha definición son “consenso” y “organismo reconocido”. El consenso es un acuerdo general, caracterizado, en primer lugar, por la ausencia de oposición sostenida a cuestiones sustanciales por una parte importante de los intereses afectados. El consenso se caracteriza, además, por ser un proceso que implica el tratar de considerar las opiniones de todas las partes interesadas y conciliar los argumentos conflictivos. Es necesario enfatizar que el consenso no implica necesariamente unanimidad.

Un organismo reconocido es una organización de desarrollo de normas reconocida a nivel internacional, como lo son ISO, CEN, BSI, ANSI y sus SDO acreditadas, incluida HL7.

La organización de los desarrolladores de estándares de informática médica es compleja y está sujeta a frecuentes cambios. En general todas derivan de la *International Standardization Organization* (ISO) y nacen alineadas a sus normas. *American National Standards Institute* (ANSI), *British Standards Institute* (BSI). Todas ellas han establecido comités más especializados los cuales han derivado y se han agrupado para armonizar el desarrollo de estándares en la *Joint Initiative Council* (JIC) las cuales incluyen 8 desarrolladores, dentro de los cuales se encuentra *Health Level Seven* (HL7), FHIR y *Digital Imaging and Communications in Medicine* (DICOM), tres de los estándares principales en los cuales se centrará este trabajo.

3.1.5.- Herramientas Informáticas y mensajería

La integración de datos dentro del área de salud es algo relativamente nuevo dentro de nuestro país. No obstante, llevando el tema a horizontes vecinos se puede ver que la integración de los datos no solo puede ser llevada a cabo por medio de un motor de integración, sino que también se puede lograr por medio del uso de una Interfaz de Programación de Aplicaciones (API del inglés Application Programming Interface) de forma combinada, como también lo permite la mezcla de varios Web Services, según corresponda a cada sistema (45).

Interfaz de Programación de Aplicaciones

Una API consiste en un conjunto de rutinas que provee acceso a funciones de un determinado software. Se trata de un conjunto de llamadas a determinadas bibliotecas, cuya función es ofrecer ciertos servicios a partir de los procesos; dicho de otra manera, la API representa una función para conseguir abstracción en la programación. Uno de los principales objetivos de una API consiste en brindar un conjunto de funciones de uso general. De esta manera, los programadores aprovechan las ventajas de la API haciendo uso de su funcionalidad para evitar el trabajo de programar todo desde el principio. Las API asimismo son abstractas: el software que proporciona una cierta API generalmente es llamado la implementación de dicha API.

De la misma manera que una interfaz de usuario permite la interacción y comunicación entre un software y una persona, una API facilita la relación entre dos aplicaciones durante el proceso de intercambio de mensajes o datos. Esto se consigue mediante un conjunto de funciones y procedimientos que ofrece una biblioteca para que otro software la utilice como

capa de abstracción, así como un espacio de acceso e intercambio de información adicional en la parte superior. De este modo, una aplicación se sirve de la información de la otra, sin dejar de ser independientes (46).

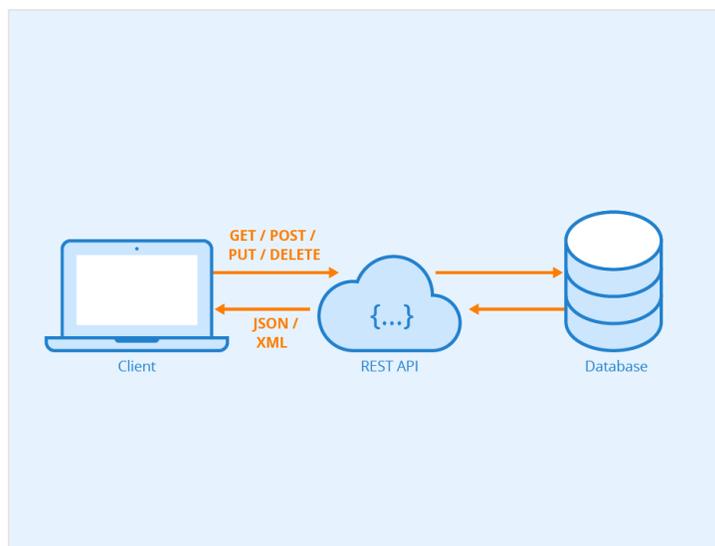


Figura 4. Representación relacional entre API, en particular REST API, y cliente para el intercambio de información. Fuente: <https://www.seobility.net/en/wiki/images/thumb/f/f1/Rest-API.png/675px-Rest-API.png> (47)

Motor de Integración

Un motor de integración es un middleware que permite la conexión y transmisión de información de un sistema a otro. Puede recibir datos provenientes de una variedad de fuentes y tomar acciones poderosas respecto de dichos datos, incluyendo el envío de éstos a múltiples sistemas externos. También puede transformar los datos de un formato a otro, o extraer

fragmentos de los datos sobre los que se desea actuar o enviarlos al exterior. Las interfaces que se configuran para realizar estas tareas se denominan canales.

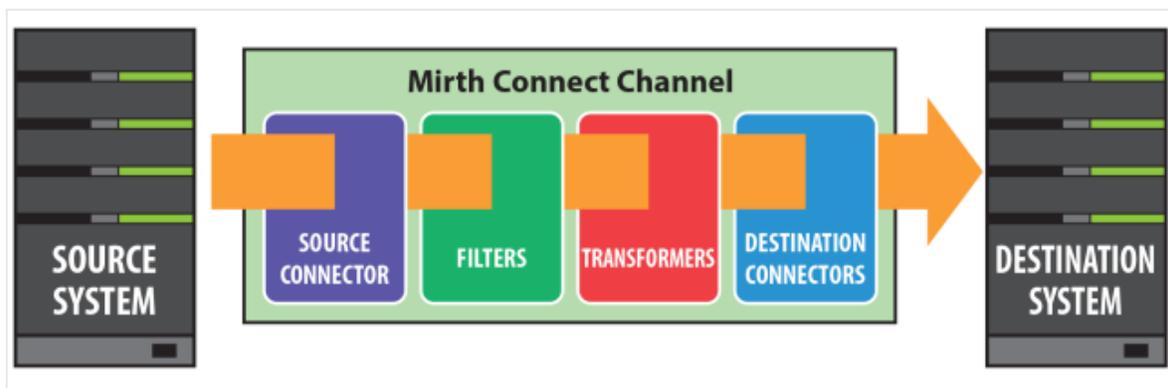


Figura 5. Relación entre dos sistemas fuente y destino mediante canal de motor integrador, en donde el mensaje puede ser filtrado y transformado. Fuente: <https://www.nextgen.com/-/media/files/nextgen-connect/nextgen-connect-39-user-guide.pdf> (48)

Un canal está formado por varios conectores. Un conector es una pieza que realiza el trabajo de introducir datos en el motor de integración (conector de origen), o de enviar datos a un sistema externo (conector de destino). Cada canal tiene exactamente un conector de origen y al menos un conector de destino. Dado que un canal puede tener varios conectores de destino, esta interfaz puede ser considerada como un proceso que obtiene datos de una fuente particular, y los envía a muchos lugares diferentes. Por ejemplo, puede recibir datos a través de HTTP, luego escribir los datos en un archivo en algún lugar, y también insertar partes de los datos en su base de datos personalizada. A este respecto, el motor de integración MIRTH CONNECT es un programa que reúne las características antes mencionadas, es de código libre, multiplataforma y soporta todos los formatos de mensajería necesarios, por lo que ha sido seleccionado para llevar a cabo este trabajo AFE.

Servidor DICOM

Es un servidor, también denominado PACS, cuyo propósito es el almacenamiento de imágenes, las que son archivadas en formato DICOM y usualmente son transferidas a estaciones de trabajo desde donde se puede acceder y revisar. Es de uso habitual en Radiología, ya que soporta las diferentes modalidades de imagen que se generan en dicha especialidad. Estos servidores son capaces de integrarse con los distintos sistemas de información hospitalaria, tales como HIS y RIS, así como con los registros médicos electrónicos (EMR).

Inicialmente, estos servidores se ubican de manera local en los diferentes Centros de Salud. En la actualidad, existe la posibilidad de contar con la misma funcionalidad, a través de servicios en la nube. También existen softwares de código abierto, como ORTHANC, que permiten transformar cualquier computador en un mini servidor en donde también cuentan con API RESTful que permitirían manejar el software desde cualquier lenguaje informático. Del mismo modo, también cuentan con plugins tales como un visor web y otros, que permitirían ampliar sus funciones básicas (49).

Servidor Web

Un servidor web es básicamente un ordenador que ejecuta páginas web. Es un programa informático que distribuye dichas páginas a medida que se solicitan. Su principal objetivo es almacenar, procesar y entregar las páginas web a los usuarios, intercomunicación que se realiza mediante el protocolo de transferencia de hipertexto (HTTP). Estas páginas web son en su mayoría de contenido estático e incluyen documentos HTML, imágenes, hojas de estilo,

test, etc. Además de HTTP, este servidor también soporta los protocolos Simple Mail Transfer Protocol (SMTP) y File Transfer Protocol (FTP), para el envío de correos electrónicos, así como para la transferencia y almacenamiento de archivos.

Un servidor web que no es de uso público pero que se utiliza internamente, se denomina servidor de intranet. Cuando un usuario solicita un sitio web añadiendo la URL o dirección de internet en la barra de direcciones de un navegador (como Chrome o Firefox), dicho navegador envía una solicitud a Internet para desplegar la página web correspondiente a esa dirección. Un servidor de nombres de dominio (DNS) convierte esta URL en una dirección IP (por ejemplo, 172.27.25.47), que a su vez apunta a un servidor web.

Se solicita al Servidor web que presente el sitio de contenido al navegador del usuario. Todos los sitios web de Internet tienen un identificador único en forma de dirección de Protocolo de Internet (IP). Esta IP se utiliza para comunicarse entre diferentes servidores a través de Internet. En la actualidad, el servidor web Apache es el más común disponible en el mercado, la gran mayoría de las aplicaciones basadas en internet lo utilizan como entorno de servidor por defecto, constituyendo de esta manera el software de código abierto que maneja casi el 70% de todos los sitios web disponibles hoy en día.

Base de Datos

Una base de datos es una colección sistemática de datos que permite almacenarlos y manipularlos de forma electrónica, facilitando su gestión. Existen diferentes tipos de bases de datos, por lo que corresponde señalar que en esta AFE se utilizarán las del tipo relacional, donde se definen las relaciones en forma de tablas. También se denomina Sistema de Gestión

de Base de Datos Relacional (SGBD), que es el tipo más popular en el mercado. Algunos ejemplos de sistemas SGBD son MySQL, Oracle y Microsoft SQL Server. Para el manejo más sencillo de estas bases de datos existen aplicaciones web tales como PhpMyAdmin, un software muy popular basado en PHP, que sirve para administrar bases de datos MySQL de forma sencilla y con una interfaz amistosa al usuario.

Servidor HAPI FHIR

Es una implementación de la especificación HL7 FHIR Fast Healthcare Interoperability Resources (pronunciado como fire) para la plataforma Java, que sin embargo no forma parte del estándar HL7 HAPI FHIR, pero que se basa en el mismo principio de practicidad de FHIR. Este servidor está centrado en la creación de una biblioteca con recursos, a través de la cual se puede acceder a distintos campos de información clínica, datos que a su vez son factibles de ser consultados y expuestos por medio de la API de FHIR.

Formatos de Mensajería

HL7

Es un conjunto de estándares establecidos con el propósito de facilitar el intercambio electrónico de información clínica. Utiliza una notación formal del lenguaje unificado de modelado (Unified Modeling Language, UML) y un metalenguaje extensible de marcado con etiquetas (Extensible Markup Language, XML).

- Mensajería HL7 Versión 2: Estándar de mensajería para el intercambio electrónico de datos de salud. Cabe mencionar que HL7 v2 es el formato de intercambio sanitario más utilizado en la actualidad. Los mensajes se envían en respuesta a eventos desencadenantes y constan de un conjunto de segmentos, definidos mediante una tabla de sintaxis de mensajes abstractos. Los segmentos contienen campos y los campos contienen componentes que pueden a su vez contener subcomponentes, separados por delimitadores. Los segmentos, denotados con la letra Z, se definen localmente y se especifican mediante tablas de definición. Los tipos de datos, algunos simples y otros complejos, constituyen los componentes básicos de cada elemento. Los códigos se definen mediante tablas.
- Mensajería HL7 Versión 3: Estándar de mensajería para el intercambio electrónico de datos de salud basada en el Modelo Referencial de Información (Reference Information Mode, RIM).
- CDA HL7: (Clinical Document Architecture) Estándar de arquitectura de documentos clínicos electrónicos.
- FHIR HL7: Se trata del último estándar desarrollado y promovido por la organización internacional HL7 (Health Level Seven), responsable de algunos de los protocolos de comunicaciones más utilizados hoy en día en el ámbito sanitario. FHIR trata de combinar lo mejor de cada uno.

DICOM

DICOM (Digital Imaging and Communications in Medicine) es una norma o estándar internacional para el manejo y gestión de imágenes médicas e información relacionada.

Constituye el formato universal para el almacenamiento y transferencia de imágenes de los sistemas de comunicación y archivo de imágenes (PACS), los que comprenden dispositivos de adquisición de imágenes digitales, archivos de imágenes y estaciones de trabajo.

El estándar DICOM define los formatos de las imágenes médicas que pueden intercambiarse, con la información anexa y la calidad necesaria para su uso clínico. DICOM se utiliza para la mayoría de las modalidades de imagen, como la radiografía, la resonancia magnética, la medicina nuclear, la ecografía, la tomografía computarizada, la ecocardiografía, la radiografía, la ecografía y otras modalidades utilizadas en radiología, cardiología, radioterapia, oftalmología y odontología. Esta norma es utilizada por los principales fabricantes y está incorporada en la mayoría de los equipamientos médicos.

El estándar DICOM se puede dividir en dos partes, una que especifica el formato de archivo y la otra que especifica el protocolo de red.

El formato de archivo DICOM es similar a formatos conocidos de imágenes como JPEG, PNG o TIFF. Sin embargo, además de los llamados "datos de píxeles" que codifican la propia imagen médica, un archivo DICOM también incorpora información médica codificada. Esta se denomina conjunto de datos y toma la forma de una matriz asociativa clave-valor. Cada valor puede ser en sí mismo una lista de conjuntos de datos (llamada secuencia), lo que lleva a una estructura de datos jerárquica que se parece mucho a un archivo XML o JSON.

En la terminología DICOM, cada tecla se denomina tag o etiqueta DICOM. La lista de las etiquetas DICOM estándar está normalizada por un diccionario oficial, donde cada etiqueta se identifica con dos números hexadecimales de 16 bits. Por ejemplo, la fecha de nacimiento de un paciente está asociada con la etiqueta DICOM. Para una mejor legibilidad, también es

común apodar estas etiquetas DICOM con un nombre en inglés de mayúsculas y minúsculas (como "PatientName " o " StudyDescription"). El estándar asocia cada etiqueta DICOM con un tipo de datos (una cadena, una fecha, un número de punto flotante, etc), que se conoce como su representación de valor (ver Figura n°6).

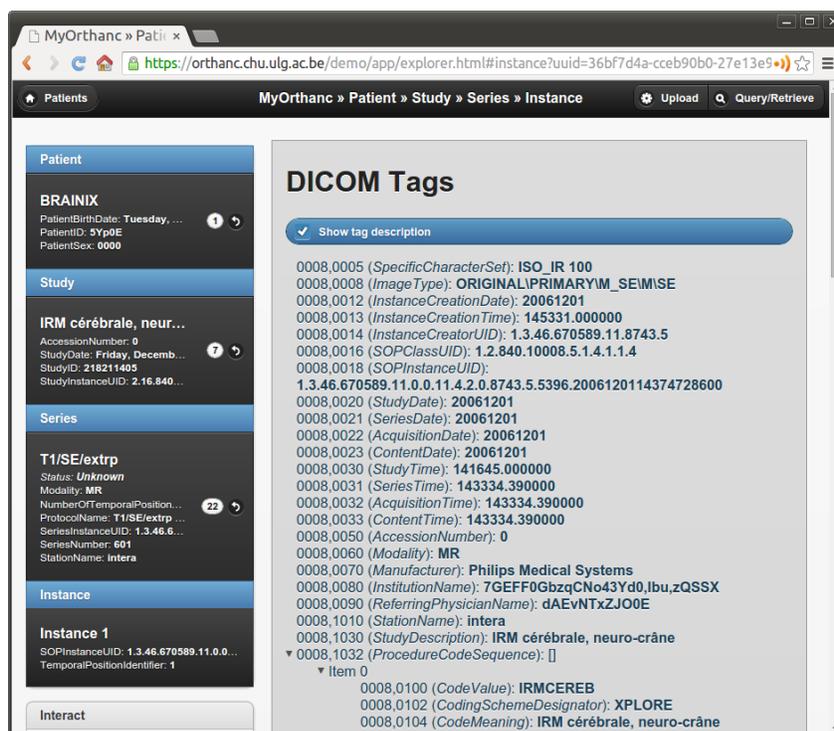


Figura 6. Imagen en donde ilustra las etiquetas o Tags de archivos DICOM. Fuente: Fuente: Understanding DICOM with Orthanc - Orthanc Book documentation (50)

El protocolo DICOM es uno de los primeros ejemplos de servicios web, implementado mucho antes de la introducción de SOAP o REST. Permite:

- Probar la conexión entre dos dispositivos (C-Echo).

- Enviar imágenes desde el dispositivo de imagen local a un dispositivo remoto (C-Store).
- Buscar el contenido de un dispositivo remoto (C-Find).
- Recuperar imágenes desde un dispositivo remoto (C-Move o C-Get).

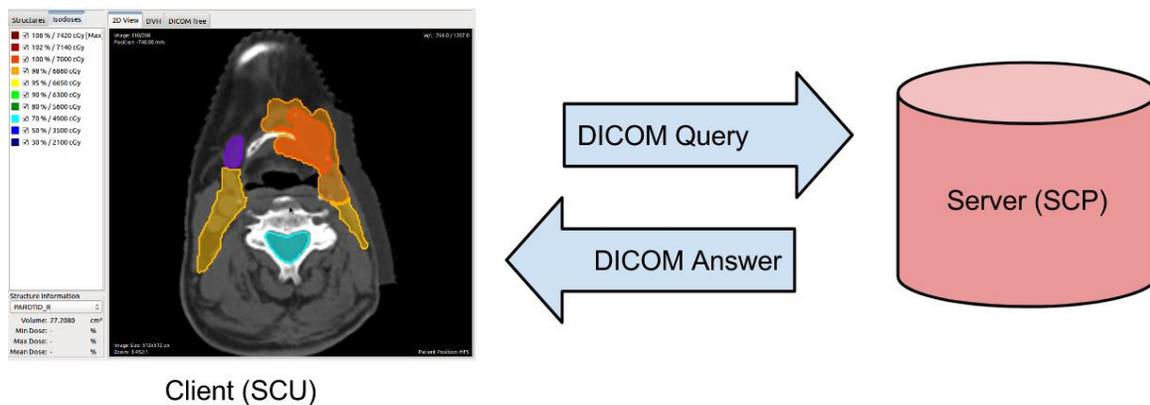


Figura 7. Relación Solicitud/Respuesta entre cliente (SCU) y Proveedor de clase de servicio (SCP).
Fuente: Understanding DICOM with Orthanc - Orthanc Book documentation (44)

En la terminología DICOM, el cliente de uno de sus servicios se denomina ‘Usuario de Clase de Servicio’ (SCU) y el servidor que maneja las solicitudes se denomina ‘Proveedor de Clase de Servicio’ (SCP). El cliente envía una solicitud que está codificada como un archivo DICOM y el servidor responde con un archivo DICOM.

La conexión entre un SCU y un SCP se denomina asociación. Dicha asociación se inicia con un acuerdo entre las partes detallando los comandos que se pueden intercambiar, así como las sintaxis de transferencia que se admiten. El resultado de esta negociación se denomina

‘Contexto de Presentación’. Una vez que se negocia la asociación, este canal de comunicación se puede utilizar para enviar sucesivamente múltiples comandos independientes.

Parámetros de un servidor DICOM

Históricamente, el protocolo DICOM se diseñó para funcionar en enlaces punto a punto. Hoy en día, el mencionado protocolo se utiliza sobre el estándar de comunicación TCP/IP. Esto implica que un servidor DICOM puede identificarse especificando los parámetros de su socket de red, su dirección IP (o, de manera equivalente, su nombre de host DNS simbólico) y su puerto TCP (el puerto DICOM estándar es 104).

Por otra parte, cada dispositivo de imagen (ya sea un cliente o un servidor) debe estar asociado con un nombre simbólico, el que se denomina ‘Título de Entidad de Aplicación’ (AET). El AET es único dentro de la Intranet de un determinado hospital. Para una mejor compatibilidad entre proveedores, el AET debe estar compuesto únicamente por caracteres alfanuméricos en mayúsculas (más los caracteres “-” y “_”), y su longitud debe ser inferior a 16 caracteres. En conjunto, la dirección IP, el puerto TCP y el AET describen todos los parámetros de un servidor DICOM.

4.- MATERIALES Y MÉTODOS

4.1.- Levantamiento de procesos

La realización de este trabajo contempló una amplia revisión de la literatura existente en el campo de los sistemas de auditorías médicas en las diferentes especialidades de teleinformes que han sido abordadas a la fecha. Se analizó de manera específica el estado del arte respecto a la evaluación de imágenes médicas y reportes, ponderando aquellas metodologías que se ajustaban de mejor manera a los objetivos específicos del presente trabajo. La búsqueda bibliográfica, realizada principalmente en las plataformas PUBMED, SCIELO y COCHRANE, se centró en terminologías relacionadas a la gestión de la calidad en telemedicina y particularmente a sistemas de evaluación de imágenes y auditoría de reportes. Los términos específicos contemplados en la búsqueda fueron “*Quality assurance*”, “*Quality Improvement*”, “*Telemedicine*”, “*Teleradiology*”, “*Audit*”, “*Peer Review*”, “*Image Quality*”.

Para el levantamiento de procesos de telereportes, se revisó la literatura nacional respecto a las diferentes modalidades, tomando en consideración principalmente lo detallado en el Programa Nacional de Telesalud (MINSAL). Dichos procesos, así como los nuevos que se proponen en el presente trabajo, fueron realizados mediante una herramienta informática de modelamiento BPMN, *Bizagi Process Modeler*, software gratuito en sus versiones más básicas, que permite extraer los procesos modelados en diferentes formatos.

Para cumplir con el tercer objetivo y la definición de métricas, para la evaluación de imágenes se consideró la metodología *Visual Grading Analysis*. La razón para adoptar dicha metodología, fue que este análisis se ajusta de mejor manera a los flujos de trabajo clínicos, debido a que se basa principalmente en la observación de distintas estructuras anatómicas,

que es en definitiva el aspecto que se evalúa en una revisión de rutina de cualquier imagen médica. Además, la incorporación de criterios de imagen basados en guías clínicas, otorgan validez suficiente al análisis para permitir que a partir de sus resultados se logren obtener mediciones de calidad. En la selección del mencionado análisis también se consideró la evidencia aportada por diversos estudios que demuestran la existencia de una fuerte correlación con el análisis simultáneo con métodos físicos de calidad de imagen (19,27). Por otro lado, tanto la obtención como el manejo de las mediciones obtenidas suponen un trabajo relativamente rápido con el cual se logra obtener conclusiones de igual forma, ya que es posible desagregar de manera simple aquellos criterios en los cuales la calidad esté mermada. Esto permitiría tomar las medidas necesarias para mejorar aquellos aspectos que muestran ser mejorables y mantener al mismo tiempo aquellos que están bien, generando así un ciclo virtuoso de mejora.

Respecto a la revisión de informes por pares, la metodología *RADPEER* es la que mejor se ajusta a lo que sería una revisión rápida de la calidad de un reporte. Esto, básicamente, a través de la medición del grado de discrepancia de dicho reporte ante una segunda lectura. Sobre la base de una escala que valora el grado de discrepancia, es posible recolectar y tabular fácilmente la información para obtener conclusiones. Como se expuso anteriormente, este sistema experimentó diversos cambios que le fueron dando mayor significancia clínica a las discrepancias, permitiendo un análisis más detallado de sus distintos grados, no centrándose solo en las discrepancias gruesas, ya que éstas, en la práctica, han demostrado ser poco frecuentes. Por lo tanto, aunque el nivel de discrepancia que exista en la revisión de un reporte sea bajo, lo relevante a considerar es el umbral permitido para la existencia de significancia clínica. Entre los elementos que se incorporaron con el propósito de evitar sesgos está la

doble lectura ciega, ya que eso permitiría al revisor no hacer un juicio previo del reporte que está analizando. De esta manera, se hace posible recrear de manera íntegra el mismo contexto de revisión, aspecto que es fuertemente criticado en el sistema *RADPEER* original.

En cuanto a la aleatorización o no aleatorización de los casos, en esta AFE se considerará que en un contexto de auditoría externa esta debe ser necesariamente de carácter aleatorio. Cabe señalar que, aunque la no aleatorización de casos discrepantes ha demostrado tener mucho valor, esto sólo es válido en el contexto de revisión por pares interna con fines de aprendizaje. No se aplica a una auditoría externa, debido a que es costosa de recolectar y demanda mucho tiempo, lo cual la hace inviable en este contexto.

Respecto al tamaño muestral, no existen estudios que validen un tamaño muestral determinado. Las muestras utilizadas en distintos estudios están en el rango de un 3 a 5% de los reportes, y han sido recolectados en periodos semestrales o anuales, de forma retrospectiva (8,51). Existe un estudio (51) en donde sugiere que para obtener información valiosa se requiere tomar una muestra de al menos el 2.5% de los reportes de cada radiólogo, con un máximo de 300 en el periodo de un año. Para darle validez estadística a la muestra se definirá un tamaño muestral estadísticamente significativo en periodos semestrales utilizando las herramientas metodológicas de muestreo para la evaluación de la calidad de la superintendencia de salud, ya que es un método conocido y utilizado en la práctica diaria en cualquier establecimiento de salud (52).

Es importante hacer notar que la muestra debe estar previamente anonimizada o pseudoanonimizada, tanto con respecto a la información del paciente como a la del especialista que reporta. La anonimización de los datos del paciente se sustenta ante el cumplimiento del manejo de datos sensibles en el marco de la ley N° 19.628 de protección

de la vida privada, así como de la ley N° 20.584 sobre derechos y deberes de las personas en salud. Por su parte, el resguardo de la información del especialista se debe a que escapa a los objetivos del desarrollo de cualquier sistema de auditoría que, como se mencionó anteriormente, constituye un instrumento de perfeccionamiento y mejora continua. El revelar dicha información podría ser percibido como un acto que pudiera implicar acciones punitivas. De acuerdo con los argumentos presentados, en esta AFE se considerará dentro de sus procesos el trabajar con datos anonimizados, tanto para los pacientes como para el personal clínico.

Otro aspecto fundamental para llevar a cabo una adecuada revisión por pares es que cada reporte debe adjuntar la misma información de contexto con la que contaría el evaluador original. Cabe señalar que se ha reportado como un factor de error y a causal de discrepancia la falta de dicha información de contexto en la elaboración de reportes (7). Los análisis se realizan por modalidad, esto es, basado en parte del cuerpo, si es pediátrico o adulto, por establecimiento, etc. Lo anterior, debido a que los contextos son propios de cada lugar, por lo que, con el propósito de obtener un mejor análisis, la información debe estar desagregada.

4.2.- Implementación

La implementación del sistema propuesto en este trabajo se llevará a cabo mediante la instalación de los siguientes servidores:

- Motor de Integración Mirth Connect
- Servidor PACS Orthanc para el almacenamiento de estudios y despliegue de imágenes.

- Servidor HAPI FHIR para el almacenamiento de reportes.
- Servidor Web Apache y base de datos MySQL para el alojamiento y elaboración del Front End y almacenamiento de resultados de evaluación. El Front End o plataforma web será programada en PHP, para la creación y administración de usuarios y además la creación de los formularios y su secuenciación. Para el despliegue de los estudios se utilizarán las capacidades de la API de Orthanc para embeber los estudios de manera que se desplieguen en las evaluaciones respectivas.

4.3.- Caso uso

Para la aplicación de esta herramienta de medición se trabajó en base al proyecto IA TRad Chile, el que se lleva a cabo en el Servicio de Imagenología del Hospital Clínico de la Universidad de Chile, en conjunto con diversas unidades dependientes de la Facultad de Medicina de la misma casa de estudios. El proyecto IA TRad nació en el contexto pandémico COVID, como una herramienta que provee soluciones IA validadas para optimizar la gestión de la presión asistencial a través del apoyo automatizado de reportes radiológicos. Dicho proyecto también abarca la predicción de eventos clínicos adversos, mediante el análisis híbrido de hallazgos imagenológicos con datos clínicos multidimensionales (53).

En particular, mediante la segmentación de imágenes de CT de pulmón, específicamente de lesiones asociadas a infección por COVID (opacidad en vidrio esmerilado, consolidación y pulmón), fue posible obtener, a través de la reconstrucción tridimensional de los *slices* del examen, una representación volumétrica que corresponde al porcentaje de afectación, en relación con el volumen total del pulmón.

Para su incorporación dentro del contexto asistencial, la calidad de la segmentación resulta ser un elemento esencial al momento de validar herramientas de segmentación de imágenes basados en IA. Para lograr esta validación existen diversos métodos que permiten evaluar la calidad de la segmentación mediante pruebas subjetivas y objetivas. Las pruebas subjetivas son las más utilizadas para evaluar el desempeño de los algoritmos de segmentación. Estas se basan en su mayoría en estudios de observación en donde grupos de observadores expertos evalúan la calidad de la segmentación (54).

Para cumplir con este propósito en la presente AFE, se propone a la Comisión Médica del proyecto Ia Trad un enfoque que permita abordar la calidad de la segmentación aplicando una estrategia similar a lo propuesto para la auditoría de teleinformes e imágenes médicas. Esto es, se propone realizar dicha evaluación en distintas dimensiones, basado en enfoques similares descritos en la literatura. A continuación, se detalla el proceso propuesto en este trabajo.

Una primera dimensión será evaluar, en una escala de Likert de 1 a 5, la calidad de segmentación general en función de la calidad de segmentación basado en la presencia de errores y el tiempo que tomaría corregirlos (55). En la escala, 1 corresponde a ‘Muy pobre’ y 5 corresponde a ‘Excelente’. Los criterios correspondientes a cada puntuación de la escala se ilustran en la Tabla n°3.

Tabla 3. Descripción de los criterios asociados a la puntuación en Escala de Likert.

Puntuación	Criterio
1	Muy pobre: segmentación con muchos errores que requieren un esfuerzo excesivo del lector para corregirlos
2	Pobre: segmentación con errores que requieren correcciones considerables y/o que requieren mucho tiempo
3	Aceptable: segmentación con inexactitudes que requieren correcciones limitadas y/o breves
4	Buena: segmentación con pequeñas imperfecciones insignificantes para el lector
5	Excelente: segmentación que corresponde al resultado ideal para el lector

Fuente: Berta L, Rizzetto F, De Mattia C, Lizio D, Felisi M, Colombo PE, et al. Automatic lung segmentation in COVID-19 patients: Impact on quantitative computed tomography analysis. *Phys Med.* julio de 2021;87:115-22. (55)

Una segunda dimensión del enfoque propuesto en esta AFE es la evaluación indirecta de la calidad de la segmentación, lo que en este caso corresponde al grado de severidad de daño pulmonar (56,57) aspecto que será evaluado mediante la suma de puntajes de afectación por lóbulos pulmonares. Esto, con el propósito de comparar posteriormente este parámetro con el informe de afectación volumétrica de cada estudio. Un puntaje alto de afectación pulmonar debería concordar con un alto volumen de compromiso pulmonar segmentado. De esta manera se podrá evaluar la concordancia entre el volumen de compromiso reportado por el algoritmo y la evaluación subjetiva de severidad realizada por el médico evaluador en cada estudio. Para la evaluación del puntaje de severidad se utilizará la metodología basada en el estudio de Li K, Fang et al (57), donde propone un sistema de medición estableciendo el nivel de daño por lóbulo en una escala de puntaje en donde cada valor corresponde a un intervalo en porcentaje de afectación por la enfermedad.

Tabla 4. Puntaje de severidad según porcentaje de afectación valorado por lóbulo pulmonar

Lóbulo	Puntaje Severidad (% de afectación)				
1R	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
2R	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
3R	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
1L	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
2L	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

Puntaje severidad	Porcentaje afectación
0	0%
1	1-25%
2	26-50%
3	51-75%
4	76-100%

Puntaje de severidad total
0-20

Puntaje Total Severidad

Fuente: Li K, Fang Y, Li W, Pan C, Qin P, Zhong Y, et al. CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). Eur Radiol. agosto de 2020;30(8):4407-16. (57)

Una tercera dimensión de evaluación que se propone en esta AFE, es específica y está basada en criterios de calidad de segmentación. Las clases de evaluación objetiva estas se dividen en evaluaciones supervisadas y no supervisadas. Las primeras son aquellas en donde se evalúan los resultados de algoritmos de segmentación tomando como referencia un *gold standard* el que corresponde a imágenes previamente segmentadas de forma manual. El resultado comparativo de ambas segmentaciones determina la evaluación final. Por el contrario, las evaluaciones no supervisadas no requieren una imagen de referencia, ya que la evaluación de una imagen segmentada se realiza en función de su adecuación a un amplio conjunto de características deseables de las imágenes segmentadas (54). Estas características se resumen en cuatro criterios de calidad de segmentación (58), los que se describen a continuación:

- I. Las regiones deben ser uniformes y homogéneas con respecto a alguna(s) característica(s) (Ej: color, textura).

II. Las regiones adyacentes deben tener diferencias significativas con respecto a la(s) característica(s) en la(s) que son uniformes.

III. Los interiores de las regiones deben ser simples y no deben contener agujeros.

IV. Los límites de las regiones deben ser simples, no irregulares, y ser espacialmente precisos (No aplica a imágenes médicas).

De esta forma, los tres primeros criterios de calidad de segmentación se aplicarán para un análisis de graduación visual en función de cada una de las características a evaluar, graduando el nivel de cumplimiento de criterio en una escala Likert, que son, como ya se mencionaron, opacidad en vidrio esmerilado, consolidación y pulmón. La descripción de los criterios se ilustra en la Tabla n° 5.

Tabla 5. Evaluación específica de la segmentación según valoración de criterio de calidad.

OPACIDAD EN VIDRIO ESMERILADO/CONSOLIDACION/PULMON					
Criterio	Nivel de cumplimiento de criterio				
I	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
II	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
III	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Puntaje total VGA					

1.- Confianza clara de que criterio no se cumple.
 2.- Algo seguro de que criterio no se cumple.
 3.- Indeciso si criterio se cumple.
 4.- Algo seguro de que criterio se cumple.
 5.- Confianza clara de que criterio se cumple.

Proceso para la evaluación de la calidad de la segmentación.

Para el proceso de evaluación se contó con la participación de 3 médicos residentes de la Beca de Radiología del Hospital Clínico de la Universidad de Chile, además de un radiólogo *Senior*, quienes contaron con un acceso de usuario individualizado a la plataforma *web* donde accedieron a listas de trabajo para cada uno de los estudios de imágenes anonimizados. Desplegados en cada uno de estos estudios estuvieron los formularios de evaluación en cada una de las dimensiones anteriormente evaluación general, severidad y específica basada en criterios. Como filtro de evaluación inicial, se tomó en consideración la primera dimensión de evaluación, que es la más general, con el propósito de establecer si la evaluación realizada por el residente debe pasar a revisión por el comité (que en este caso sólo estará compuesto por el radiólogo *senior*). En este sentido, toda evaluación que sea menor a 4 será considerada para que el comité confirme o recalifique la evaluación previa, tal como se ilustra en la Figura 8 que muestra las etapas del proceso.

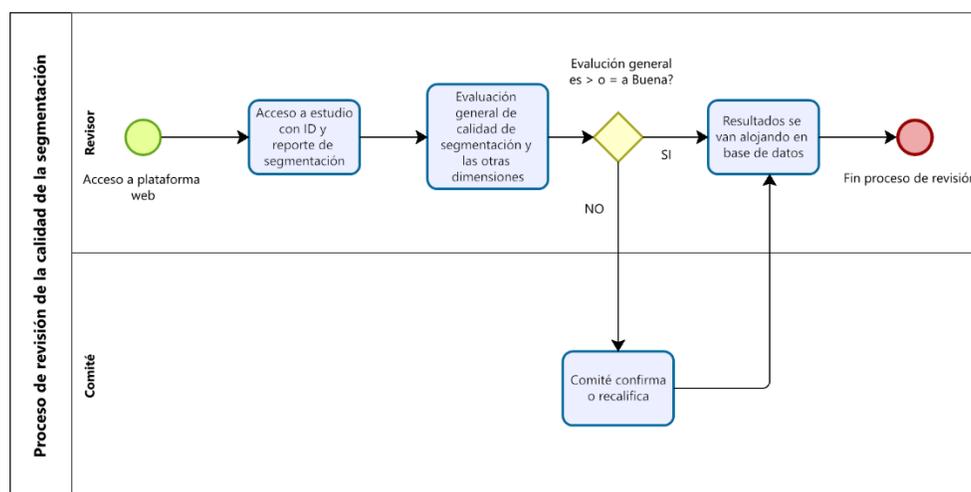


Figura 8. Se detalla el proceso de revisión de la calidad de la segmentación en las tres dimensiones mencionadas.

Para el análisis de los datos obtenidos de la evaluación se utilizó R Studio, entorno de *software* libre para cálculo estadístico y gráficos. También se utilizó Excel para la confección de gráficos y diseño de tablas complementarias.

Finalmente, se realizó una encuesta de opinión usuaria (Ver anexo n°2) con el objetivo de contar con una retroalimentación útil para implementar mejoras y cambios. Dicha encuesta fue elaborada mediante el uso de *Google Forms*, cuya plataforma permite obtener de manera rápida y remota los resultados junto a su análisis respectivo.

4.4.- Limitaciones de la prueba.

Por motivos de restricción de tiempo no se pudo llevar a cabo una evaluación amplia con un tamaño muestral que sea representativo para poder obtener resultados que permitieran concluir de forma confiable la calidad de segmentación del algoritmo, aunque dicho objetivo escapa al alcance de este trabajo.

Lo que corresponde a la evaluación de calidad de la imagen y de reporte en esta prueba de concepto se tuvo que modificar respecto a lo que se planteó como la opción de preferencia. La evaluación de la imagen se modificó a la de segmentación, por lo tanto, se consideraron criterios ajustados a ese propósito. Respecto al reporte de volumen y porcentaje de afectación que entrega el algoritmo, se buscó una metodología de evaluación subjetiva que permitiera acercarse en similitud, dado que constituyen un valor numérico y no una apreciación interpretativa, como es el caso en los reportes convencionales basados en imágenes.

5.- RESULTADOS

Para la realización del presente trabajo, el primer objetivo fue encontrar una metodología de revisión, tanto de reportes como de imágenes médicas, que fuera de rápida ejecución y que entregara información cuantitativa útil para detectar áreas de las distintas disciplinas involucradas en los telereportes donde la calidad pueda ser mejorada.

Para situarse en el contexto de lo que sería un sistema de auditoría de telereportes basados en imágenes médicas, hay que entender primeramente cómo éstos operan, en otras palabras, conocer a cabalidad los principales atributos de estos procesos de atención a distancia.

Como ya fue mencionado en la sección Introducción, las principales estrategias o modelos actuales para la realización de teleinformes están en las especialidades de Teledermatología, Teleoftalmología y Teleradiología. A continuación, se describe, de manera sucinta, la secuencia de procesos que se realizan en cada una de ellas.

5.1.- Levantamiento de procesos de Teleinformes

Teledermatología

En las unidades de atención primaria, ante la necesidad de una evaluación dermatológica, el médico solicitante realiza la recolección de información clínica del paciente, junto a la o las imágenes competentes del objeto de interés. Esto es enviado mediante la misma plataforma web donde está alojada la información para que sea revisada por el médico especialista, quien accede de forma remota para entregar tanto su informe como la prescripción. Una vez realizado el informe, este queda alojado en la plataforma para que el médico solicitante

acceda a sus resultados y proceda a notificarlos posteriormente al paciente (59). (Ver Figura n° 9).

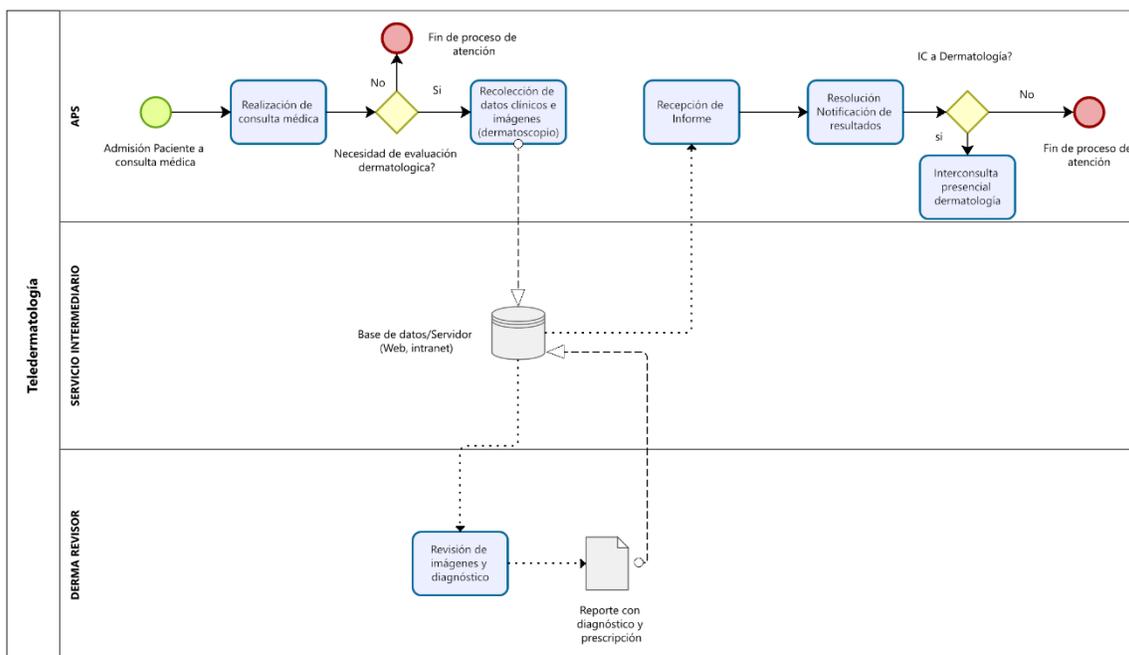


Figura 9. Levantamiento de proceso de Teledermatología. Fuente: Elaboración propia.

Teleoftalmología

En las UAPOS, mediante interconsulta bajo el programa de salud cardiovascular (60) de pacientes diabéticos, el tecnólogo médico realiza el examen de fondo de ojo. El *set* de fotos resultante es enviado vía plataforma *web* con la información demográfica del paciente. El médico especialista accede a la plataforma de forma remota para informar los hallazgos clínicos según corresponda. El reporte es descargado en APS para la resolución y notificación de resultados.

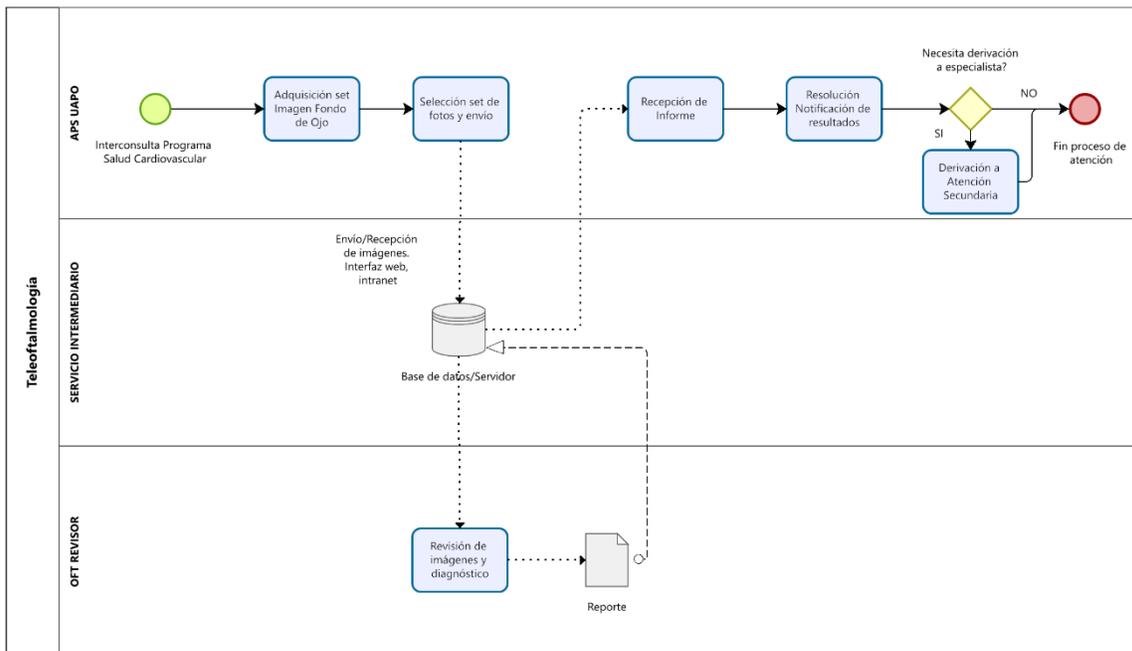


Figura 10. Levantamiento de proceso de Teleoftalmología. Fuente: Elaboración propia.

Teleradiología

En diversos tipos de centros de salud se adquieren imágenes en cualquiera de sus diversas modalidades, las que son enviadas a una plataforma web intermediaria, en conjunto con información de contexto. A dicha plataforma accede de forma remota el médico revisor especialista con el propósito de emitir su informe. Luego dicho informe es descargado en el Centro de Salud para la posterior resolución y notificación de cada caso (61).

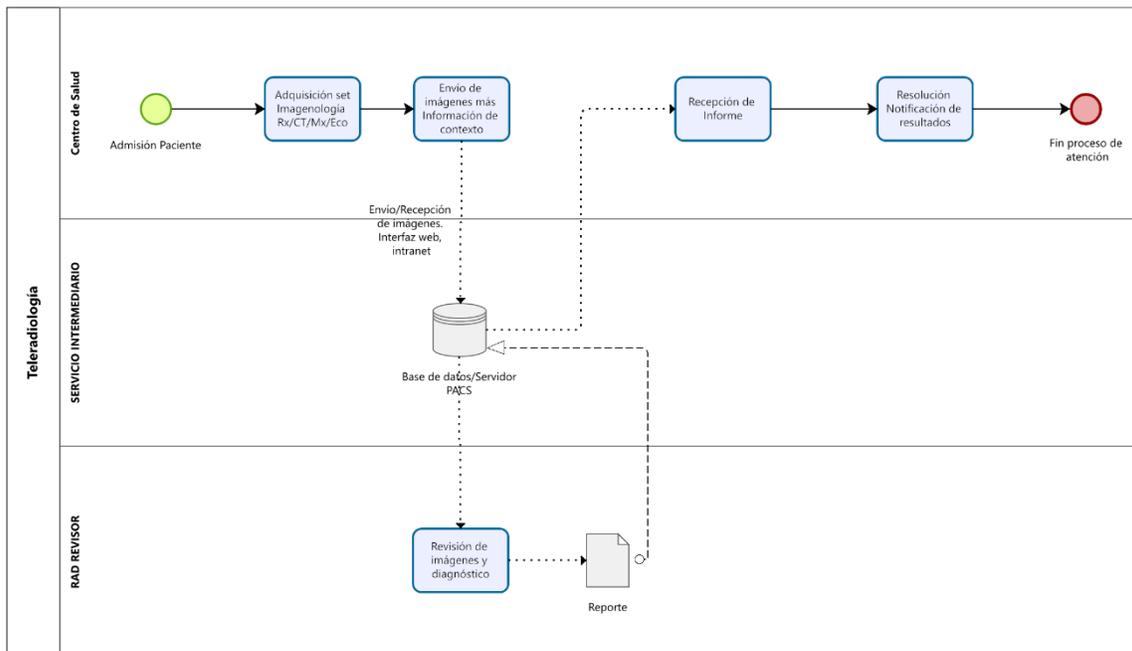


Figura 11. Levantamiento de proceso de Teleradiología. Fuente: Elaboración propia.

El análisis de estos tres modelos o estrategias de provisión de salud telemática permite concluir que son dos los principales elementos en los cuales sería posible llevar a cabo una revisión remota y sobre los cuales es factible obtener conclusiones que permitan llevar a efectuar actos de mejora. Los dos elementos claves en este tipo de proceso son los entregables que confluyen en un nodo común, con los cuales se construye la interacción remota. Dichos elementos son primeramente las imágenes médicas, a partir de las cuales surge el segundo elemento, que es el reporte. Por lo tanto, se concluye que la evaluación de estos dos elementos - ambos *outputs* de subprocesos que en conjunto constituyen el proceso total de la atención - permitirá configurar un sistema que permita auditar su calidad y por consiguiente obtener información con miras a la mejora continua.

5.2.- Resultado subproceso Evaluación de la calidad de imagen

Este subproceso se inicia con la recepción de la imagen anonimizada, la que debe estar asociada a un número de identificación único. A continuación, el revisor evalúa cada criterio de la imagen en la escala de Likert de 1 a 5. Una vez evaluada la imagen, si esta no cumple el puntaje con umbral mínimo de calidad asignado previamente por consenso (Ej.: puntaje umbral 75% del total), pasa a una segunda evaluación, esta vez realizada por un comité, donde se confirma o recalifica la imagen, asignando un nuevo puntaje en el segundo caso. Finalmente, los resultados son alojados en base de datos para posterior análisis (ver Figura n°12).

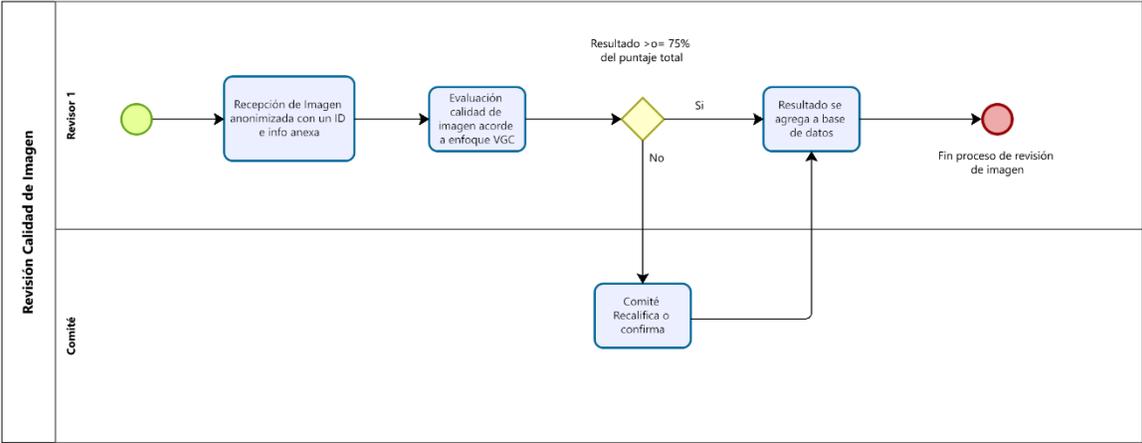


Figura 12. Secuencia del subproceso de evaluación de la calidad de imagen. Una vez recepcionada la imagen anonimizada y con su respectivo ID se procede a la evaluación de su calidad mediante el enfoque basado en criterios de imagen. Según al resultado y al umbral previamente definido de calidad se decide si pasa o no a comité para reevaluación. Fuente: Elaboración propia.

5.3.- Resultado subproceso Revisión por pares. Evaluación de Discrepancia del Reporte

Este subproceso se inicia con la recepción por un primer revisor (Revisor 1) tanto de la imagen como de la información anexa de contexto, necesaria para elaboración de su reporte. En paralelo, un segundo revisor (Revisor 2) recibe el reporte original y posteriormente el nuevo reporte elaborado por el Revisor 1. El Revisor 2 evalúa el grado de discrepancia entre ambos reportes, lo que puede dar lugar a dos opciones: i) Si el resultado es igual a 1 o a 2 con grado de discrepancia sin significancia clínica, dicho resultado se aloja en la base de datos; ii) Si el resultado es igual a 2, pero con significancia clínica, ambos reportes, más la imagen e información de contexto, es recibida por un comité evaluador, que confirma la evaluación previa o la recalifica (ver Figura n° 13).

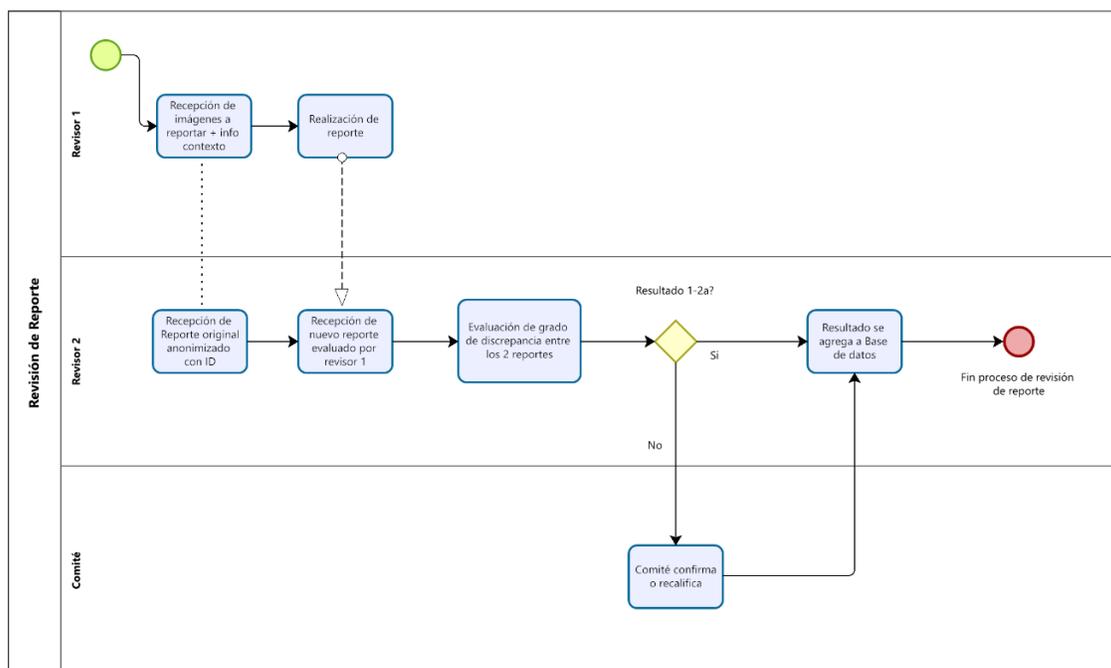


Figura 13. Secuencia de subproceso de evaluación de informes. Se incorpora evaluación doble ciego por Revisor 1, finalmente Revisor 2 determina el grado discrepancia. De acuerdo al resultado pasa o no a comité. Fuente: Elaboración propia.

5.4.- Integración de procesos de auditoría con sistemas de teleinformes

Para la integración de ambos subprocesos, esto es la evaluación de calidad de imagen y la revisión por pares, con los modelos de teleinformes, hay que considerar el supuesto de que tanto las imágenes como los reportes elaborados deben ser alojados en un repositorio central, según como fue descrito en el levantamiento de procesos de las diferentes estrategias de teleinformes, mencionadas previamente.

El acceso mediante herramientas informáticas basadas en estándares de interoperabilidad permitirá interactuar con dichos repositorios para acceder a reportes e imágenes y alojarlos en un repositorio propio para ser sometidos posteriormente a evaluación. El nivel de interoperabilidad corresponde a un nivel técnico ya que se encarga sólo del movimiento de información a un nivel de reconocimiento sintáctico.

Según esto, el proceso general de auditoría daría cuenta de la interacción entre estos dos sistemas, tal como se muestra en la Figura n°14. El proceso comienza con la adquisición de la imagen en un Centro de Salud; estas imágenes son luego enviadas a un repositorio (PACS) donde son sometidas a la evaluación de rutina del proceso de teleinforme. La interacción de herramientas de integración (motor de integración) permitiría la comunicación con dicho repositorio, para acceder a una muestra de imágenes y reportes las que serían alojadas en un servidor propio. Posteriormente, revisores auditores podrían acceder a esta información para ser sometida a evaluación y sus resultados ser alojados en el mismo servidor para su posterior análisis.

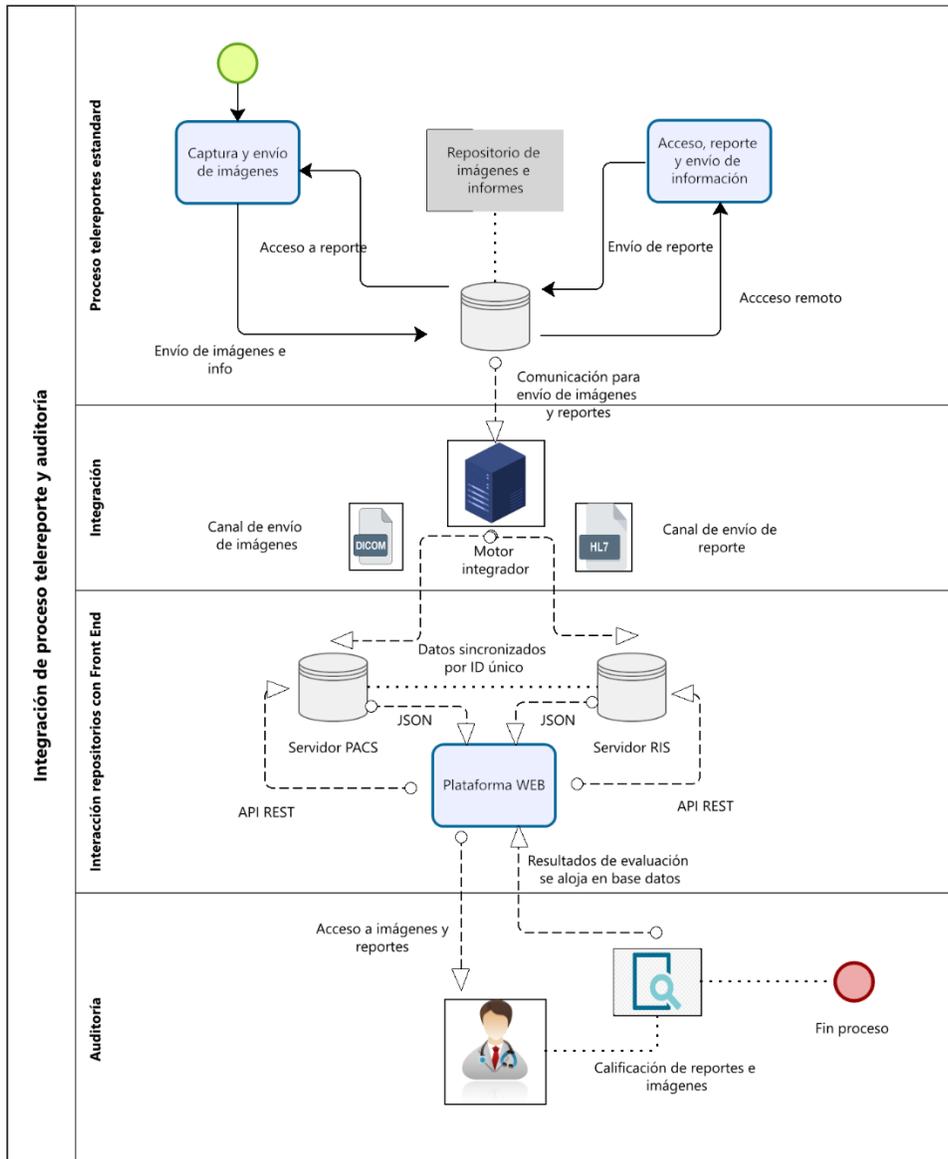


Figura 14. Integración de sistema de telereportes con sistema de auditoría. El flujo comienza con la obtención de la imagen para posteriormente ser enviada a un repositorio en donde accede el revisor para elaborar el reporte. Ambos resultados quedan alojados en el repositorio. Ante un contexto de auditoría estos serían enviados por intermedio de un motor integrador que mediante sus canales transmite ambos mensajes a los servidores que los alojarán respectivamente. Mediante las funcionalidades de las API de cada servidor estos se pueden desplegar en una plataforma web para su posterior revisión mediante acceso remoto. El resultado de dicha evaluación se aloja posteriormente en la base de datos del servidor web, constituyendo el fin del proceso.

5.5.- Arquitectura de Sistema planteada

La arquitectura propuesta para el desarrollo de la presente AFE se ilustra en la Figura n° 15 que presenta el flujo a través de la arquitectura del sistema completo.

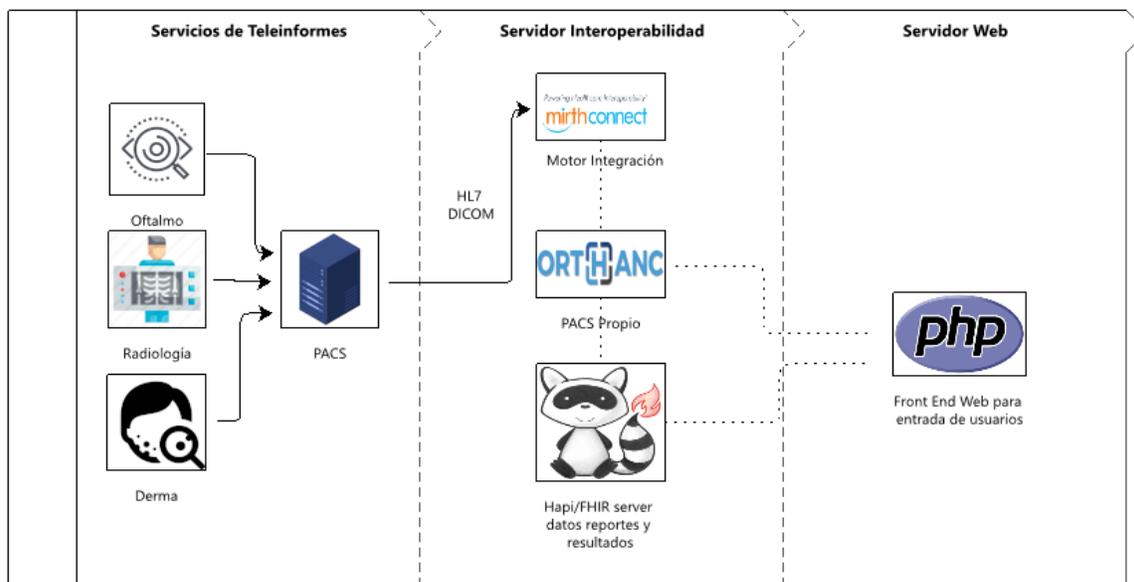


Figura 15. Arquitectura sistema informático. En concordancia a la figura 14 en esta figura se especifican las principales herramientas informáticas que se utilizaron en la implementación del sistema
Fuente: Elaboración propia

La información recolectada en los PACS de los servicios de teleinformes es enviada, como ya se mencionó previamente, en los formatos de mensajería interoperable HL7 y DICOM. Esta información es escuchada y transmitida por medio de los canales respectivos, para luego ser enviada a los servidores correspondientes: Orthanc que es el PACS propio para los estudios y HAPI FHIR para los reportes. Mediante el acceso a estos dos servidores se despliega el *Front End* para la elaboración de las listas de trabajo y visualización de estudios y reportes.

Implementación para elaboración de caso uso

La implementación informática del modelo de auditoría que se propone en este trabajo comienza mediante la instalación de los respectivos servicios o servidores que permitirán llevar a cabo el proceso. Estos son Mirth Connect como motor Integrador, Orthanc PACS como repositorio de imágenes y HAPI FHIR para alojar los respectivos reportes. Dichos servidores se implementaron en una plataforma en forma de contenedores llamada *Docker*. *Docker* es un proyecto de código abierto que en palabras simples es un contenedor de aplicaciones las cuales se pueden llevar a cabo de manera independiente y eficiente, en un entorno de seguridad.

Configuración *Mirth Connect*

Para la configuración de *Mirth* como motor integrador, y de acuerdo a las necesidades puntuales del caso uso, sólo fue necesario implementar un canal conector que solamente escuchara archivos DICOM, dado que cada estudio incorpora dentro de la información en uno de sus TAG el reporte de segmentación del algoritmo (Ver detalle más adelante en configuración de *Orthanc*). La implementación de dicho conector consistió en la configuración de un canal conector específico de escucha para archivos DICOM, los que son recibidos previo envío a su puerto TCP. Una vez recibidos los mensajes, estos pueden ser transformados a otros formatos o extraer información relevante de ellos. En el caso de esta AFE, esto no fue necesario por lo que el mensaje viaja íntegro, previa configuración de otro canal conector que envía el mensaje. Para el envío fue necesario configurar el conector de salida con la IP del Host, puerto TCP y AET cuyo nombre es AUDIT_CIMT.

The screenshot displays the configuration interface for Orthanc PACS, divided into several sections:

- Connector Type:** Set to "DICOM Listener".
- Listener Settings:**
 - Local Address: All interfaces Specific interface: 0.0.0.0
 - Local Port:
- Source Settings:**
 - Source Queue: OFF (Respond after processing)
 - Queue Buffer Size:
 - Response:
 - Process Batch: Yes No
 - Batch Response: First Last
 - Max Processing Threads:
- DICOM Listener Settings:**
 - Application Entity:
 - Max Async operations:
- Destination Settings:**
 - Connector Type: Wait for previous destination
 - Queue Messages: Never On Failure Always
 - Advanced Queue Settings: Retries
 - Validate Response: Yes No
 - Reattach Attachments: Yes No
- DICOM Sender Settings:**
 - Remote Host: Lc
 - Remote Port: Lc
 - Remote Application Entity: Local Applicati
 - Max Async operations:

Figura 16. En esta figura se muestra la configuración respectiva, tanto del canal que “escucha” el mensaje recibido, como del que envía al servidor en específico mediante la configuración del Host, puerto y AET.

Configuración Orthanc PACS

De acuerdo a lo mencionado anteriormente la implementación del PACS fue suficiente para el manejo tanto de las imágenes como de su respectivo reporte. Esto permitió la simplificación del trabajo de implementación ya que cada estudio tenía adosada en cada una de sus instancias como TAG número 0008,103e *SeriesDescription* la información respecto al volumen pulmonar segmentado, porcentajes de afectación en relación al volumen total del pulmón y los porcentajes específicos de afectación en vidrio esmerilado y consolidación. Cada estudio cuenta con un ID numérico de identificación en el TAG 0008,0050 *AccessionNumber*. Este número es único y relevante para los comandos de interacción entre el *Front End* y el PACS para el despliegue de los estudios y la interacción relacional de la base de datos. Toda información respecto a datos específicos del paciente está anonimizada.

```
0008,0008 (ImageType): DERIVED\SECONDARY
0008,0012 (InstanceCreationDate): 20211116
0008,0013 (InstanceCreationTime): 172249
0008,0016 (SOPClassUID): 1.2.840.10008.5.1.4.1.1.2
0008,0018 (SOPInstanceUID): 1.1.3.12.2.1107.5.1.4.83885.20211116172249.1
0008,0020 (StudyDate): 20211116
0008,0021 (SeriesDate): 20211116
0008,0030 (StudyTime): 172249.604899
0008,0031 (SeriesTime): 172249
0008,0050 (AccessionNumber): 2717245
0008,0060 (Modality): CT
0008,0090 (ReferringPhysicianName):
0008,103e (SeriesDescription): Volumen Pulmonar:2289.17mL;Volumen Infeccion:429.01mL;Compromiso Pulmonar (CP):18.74%(CP VE:12.74%,CP COND:6.0%).Composicion infeccion:67.98%VE;32.0%COND.
0010,0010 (PatientName):
0010,0020 (PatientID): 121
```

Figura 17. En esta figura se pueden ver las principales etiquetas o TAG de un estudio. El ID del estudio se puede apreciar en el tag correspondiente a AccessionNumber y el reporte de la segmentación del estudio en el tag que corresponde a SeriesDescription.

Implementación *Front End*

Para este propósito se realizó una evaluación de diversas herramientas de diseño *web*, creación y administración de usuarios, búsqueda de aplicación o *plugin* de elaboración de formularios. Para el despliegue de las imágenes en la plataforma, se utilizaron las capacidades de la API de Orthanc para embeber dentro de la página las imágenes asociadas a un ID.

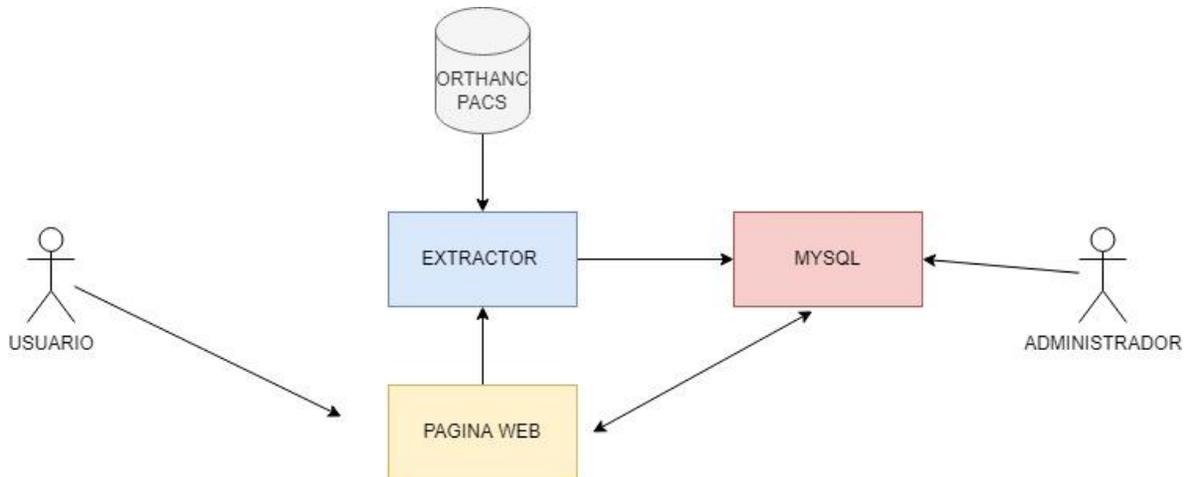


Figura 18. Esquema UML de Funcionamiento del Front End, ilustrando la relación que existe entre sus componentes y usuario y administrador. Fuente: Elaboración propia.

La relación entre los componentes del *Front End* se ilustra en la Figura n° 18. En líneas generales, el PACS propio levantado con Orthanc alberga el almacenamiento de los estudios en curso. La figura “Extractor” representa el traspaso de la información desde el PACS, para luego alojarla en la base de datos MySQL. Dicha extracción se realiza de manera automática, cada vez que se inicia la página *web*. Mediante una función *cURL* (intérprete de comandos orientada a la transferencia de archivos) se obtiene un archivo JSON con la información de cada uno de los estudios. Posteriormente, estos son cargados a la base de datos, en una conexión en PHP mediante la extensión MySQLi. Finalmente, la página *web* presenta los datos extraídos desde el PACS que han sido almacenados en la base de datos. Los estudios son presentados mediante un visor *web* de la API REST de Orthanc, embebido en el desarrollo. A través de este visor, el usuario tiene la posibilidad de revisar cada estudio y sus imágenes, disponiendo de las mismas funcionalidades que le otorga una estación de trabajo estándar. De esta manera, para cada estudio se desplegarán las tablas de graduación de la calidad de segmentación. Además, la página tiene control de usuarios y perfiles, existiendo

un perfil “súper usuario revisor”, quien confirma o recalifica las evaluaciones de los otros usuarios en el caso de los estudios calificados con un valor menor a 4 en la tabla de evaluación general.

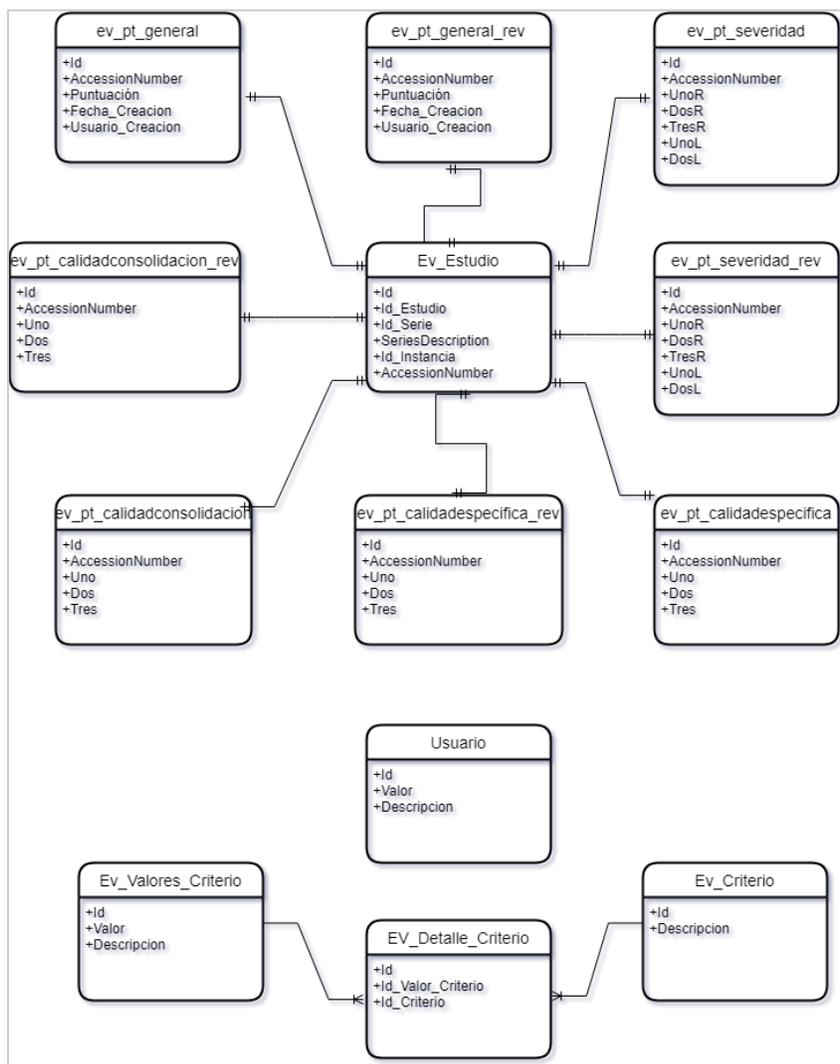


Figura 19. En este esquema se grafica el Diagrama de entidad relación de la base de datos.

5.7.- Resultados de evaluación de la segmentación del Caso de Uso.

En total se evaluaron 6 estudios de segmentación de pulmón de pacientes con Covid-19 por 3 evaluadores, evaluando dos estudios cada uno. Posteriormente, dichos evaluadores respondieron una encuesta de opinión usuaria (ver Anexo 2) en donde se hacen preguntas relativas al funcionamiento de la plataforma, facilidad de uso, calidad de las imágenes, etc., incluyendo además preguntas abiertas para sugerencias y opiniones respecto a su uso (62).

5.7.1.- Resultado Primera dimensión de evaluación:

Una primera dimensión de evaluación de la calidad de la segmentación, como se mencionó anteriormente, tiene relación con la valoración general de la calidad de la segmentación de las imágenes, en función del tiempo que se necesitaría para corregir dicha segmentación. El resultado obtenido para cada estudio con su respectiva puntuación se presenta en la Tabla 6, donde además se muestra la información relativa al reporte de segmentación del algoritmo de IA TRad. El cual entrega el volumen pulmonar total, volumen de infección, porcentaje de compromiso pulmonar de la infección y la proporción que corresponde a patrón en vidrio esmerilado y condensación, para cada uno de los estudios.

Tabla 6. Resumen de los resultados obtenidos en la primera dimensión de evaluación

ID	AccessionNumber	Puntuacion	Vol Pulmonar	Volumen Infección	Compromiso Pulmonar
1	2834089	3	2409.69	1121.71	46.55%(CP VE 29.99%,CP COND 16.53%)
2	2714765	4	3526.57	647.56	18.36%(CP VE 9.77%,CP COND 8.58%)
3	2881720	4	2081.55	768.51	36.92%(CP VE 15.64%,CP COND 21.24%)
4	2838365	4	4792.36	132.78	2.77%(CP VE 2.25%,CP COND 0.52%)
5	2717245	4	2289.17	429.01	18.74%(CP VE 12.74%,CP COND 6.0%)
6	2714254	2	3039.56	1464.37	48.18%(CP VE 35.16%,CP COND 12.99%)

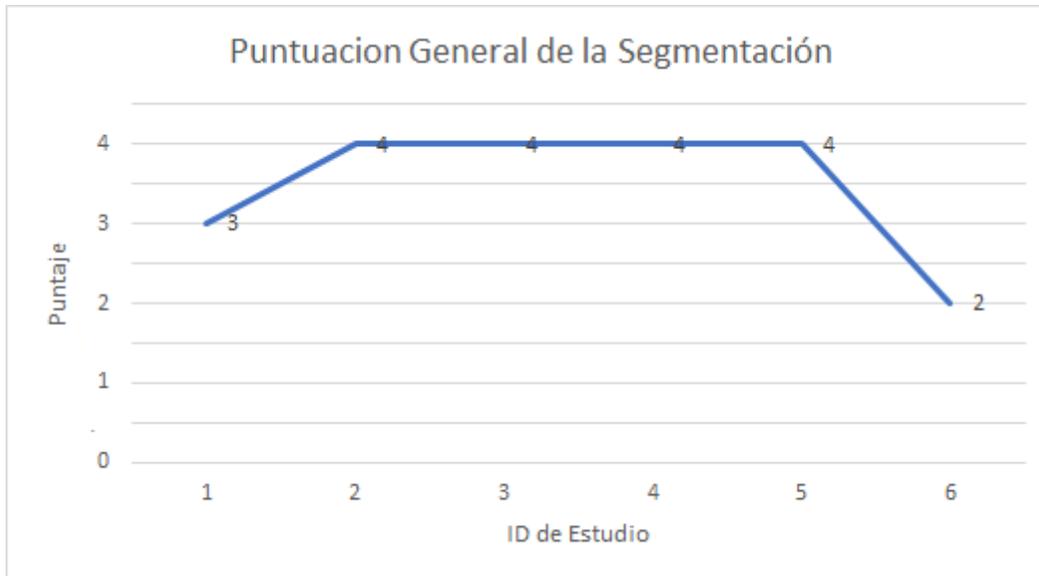


Figura 20. Puntuación general de la segmentación. Sólo dos estudios tuvieron un valor igual o inferior a tres. Fuente: Elaboración propia.

De los estudios evaluados, dos de ellos fueron calificados con un valor menor o igual a 3 y ambos fueron ratificados por comité integrado por radiólogo *senior*. De acuerdo a lo que muestra la tabla, ambos estudios coinciden en que son los que presentan mayor compromiso pulmonar, por lo tanto coinciden en presentar un mayor volumen de pulmón segmentado por el algoritmo de IA TRad.

5.7.2.- Resultados Segunda dimensión de evaluación:

La segunda dimensión de evaluación de la calidad considerada en esta AFE tenía relación con una graduación del nivel de severidad de afectación del pulmón por lóbulo comprometido. Por lo tanto, se evaluó el porcentaje de severidad de la infección en los lóbulos correspondientes al pulmón derecho e izquierdo. Los resultados para estos parámetros se pueden resumir en la siguiente Tabla 7.

Tabla 7. Puntaje de severidad de lóbulos y compromiso pulmonares segmentado.

Id	UnoR	DosR	TresR	UnoL	DosL	Puntaje Total	Severidad (a)	Compromiso (b)	(a-b)
1	1	2	3	2	4	12	48%	46,55%	1,45%
2	1	1	1	2	1	6	24%	18,36%	5,64%
3	2	2	4	1	3	12	48%	36,92%	11,08%
4	1	1	1	1	1	5	20%	2,77%	17,23%
5	1	1	2	1	2	7	28%	18,74%	9,26%
6	1	3	4	2	4	14	56%	48,18%	7,82%

Dichos datos fueron contrastados con los de la segmentación, específicamente con el porcentaje de compromiso o afectación pulmonar. Para este propósito, el puntaje total de severidad se transformó en porcentaje, asumiendo como valor de 25 el 100% de severidad. De esta manera, se pudo establecer una diferencia neta entre los porcentajes de severidad y compromiso pulmonar de cada estudio permitiendo identificar los estudios con mayores diferencias. En orden decreciente los estudios con mayores diferencias fueron 4, 3, 5, 6, 2, 1. Cabe hacer notar que la diferencia entre ambos porcentajes es siempre a favor de la valoración de severidad.

El estudio 4 es el de mayor diferencia entre las dos evaluaciones, y es a su vez el de menor porcentaje tanto de severidad como de compromiso pulmonar. El segundo de mayor diferencia, el estudio 3, presenta un porcentaje mayor de afectación. Ambos estudios no coinciden con los estudios previamente calificados con menor calidad de segmentación.

Existen distintas metodologías para evaluar y medir la concordancia entre los resultados de un evaluador con la del otro. Dependiendo si los porcentajes se ordenan en grupos o categorías ordinales, se podría medir su acuerdo mediante (i) índice de Kappa de Cohen. En

cambio, si estos resultados son evaluados como valores continuos su concordancia en la calificación podría ser evaluada mediante un coeficiente de concordancia intraclase.

Cálculo mediante Kappa de Cohen

La *Kappa* ponderada de Cohen se utiliza ampliamente en la clasificación cruzada como medida del acuerdo entre evaluadores observados. Es un índice de acuerdo adecuado cuando las puntuaciones son escalas nominales sin estructura de orden.

Los datos deben cumplir los siguientes supuestos para calcular el Kappa de Cohen.

1. Tiene dos variables categóricas de resultado, que pueden ser variables ordinales o nominales.
2. Las dos variables de resultado deben tener exactamente las mismas categorías.
3. Tiene observaciones emparejadas; cada tema es clasificado dos veces por dos evaluadores o métodos independientes.
4. Se utilizan los mismos dos evaluadores para todos los participantes o estudios.

En el caso de esta AFE, exceptuando el punto 4, se cumple con las condiciones para hacer el cálculo del índice *Kappa*. No obstante, dado que el número de evaluadores y estudios es reducido, se asumirá como la concordancia de medición de severidad de evaluadores versus compromiso pulmonar de algoritmo.

Para realizar el cálculo se agruparon los resultados de ambos métodos en categorías por intervalos porcentuales, tal como se muestra en la Tabla 8:

Tabla 8. Tabla que representa a la matriz de concordancia entre resultados de evaluadores con IA TRad.

		IA Trad					Total
		*	0	1-25%	26-50%	51-75%	
Evaluadores Médicos	0	0	0	0	0	0	0
	1-25%	0	3	0	0	0	3
	26-50%	0	0	2	0	0	2
	51-75%	0	0	1	0	0	1
	76-100%	0	0	0	0	0	0
	Total	0	3	3	0	0	6

*	Value	ASE	z	Pr(> z)
Unweighted	0.7143	0.2166	3.297	9,76E-01
Weighted	0.7500	0.1712	4.382	1,18E-02

*	Lwr	Upr
Unweighted	0.2897153	1
Weighted	0.4145261	1

Kappa	Interpretación
0-0.2	ínfima concordancia
0.2-0.4	escasa concordancia
0.4-0.6	moderada concordancia
0.6-0.8	buna concordancia
0.8-1.0	muy buena concordancia

El índice de kappa ponderado obtenido en la muestra de 6 estudios, según los dos sistemas de evaluación, fue de 0.75 (IC 95%: 0.41-1.00) representando una buena concordancia.

(ii) Cálculo mediante coeficiente de correlación intraclase:

Para la medición de los datos continuos se utilizó un coeficiente de correlación intraclase para determinar si los elementos o sujetos pueden ser calificados de manera confiable por diferentes evaluadores (cita).

El valor de un ICC puede variar de 0 a 1, donde 0 indica que no hay confiabilidad entre los evaluadores y 1 indica una confiabilidad perfecta.

Tabla 9. Representación de porcentajes de severidad medido por evaluadores y porcentaje de compromiso por el algoritmo

Id	Severidad (Med)	Compromiso (IA Trad)
1	48%	46,55%
2	24%	18,36%
3	48%	36,92%
4	20%	2,77%
5	28%	18,74%
6	56%	48,18%

Single Score Intraclass Correlation

Model: oneway

Type: agreement

Subjects = 6

Raters = 2

ICC (1) = 0.841

F-Test, H0: $r_0 = 0$; H1: $r_0 > 0$

F (5,6) = 11.6, p = 0.0049

95%-Confidence Interval for ICC Population Values:

0.317 < ICC < 0.975

El coeficiente de correlación intraclase (CCI) resulta ser 0,841.

Según la interpretación de descrita por *Koo & Li et al (63)* , si el valor de un coeficiente de correlación intraclase es:

- Menos de 0,50: poca fiabilidad.
- Entre 0,5 y 0,75: fiabilidad moderada.

- Entre 0,75 y 0,9: buena fiabilidad.
- Mayor que 0,9: excelente fiabilidad.

Por lo tanto, se concluye que un ICC de 0,841 indica una buena concordancia entre la evaluación de la severidad por los evaluadores y el porcentaje de compromiso pulmonar calculado por el algoritmo, aunque con un intervalo de confianza amplio, probablemente debido a lo acotada de la muestra.

Respecto a la valoración de severidad realizada por cada uno de los evaluadores, se observa en general que, de acuerdo a los niveles de afectación, estos se distribuyen en una frecuencia mayor en lóbulos superiores *versus* los inferiores, aunque con un nivel de afectación bajo. En cambio, si bien la distribución de frecuencias en lóbulos inferiores es más baja, estos concentran mayores rangos de niveles de severidad, incluyendo las formas con puntuación más alta. En la Tabla 10 se puede observar en detalle lo planteado, al igual que en el diagrama de cajas que se muestra en la Figura 21. En orden de afectación de menor a mayor estarían el lóbulo superior derecho, luego superior izquierdo, medio y ambos lóbulos inferiores.

Tabla 10. Distribución de frecuencias en porcentaje de valor de afectación por lóbulo pulmonar

Valor/Lóbulo	Lóbulo Uno R	Lóbulo Dos R	Lóbulo Tres R	Lóbulo Uno L	Lóbulo Dos L
0	0	0	0	0	0
1	83%	50%	33%	50%	33%
2	17%	33%	17%	50%	17%
3	0	17%	17%	0	17%
4	0	0	33%	0	33%
5	0	0	0	0	0

R: Right (derecho) L: Left (Izquierdo) Fuente: Elaboración propia

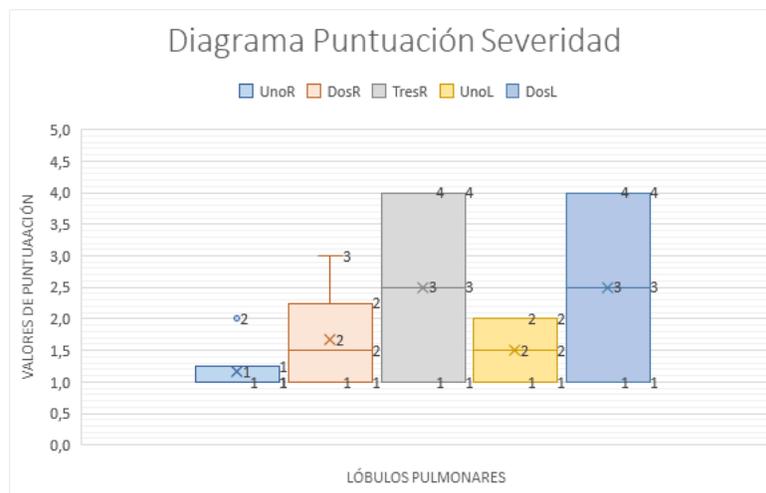


Figura 21. Diagrama de cajas de valores de afectación por lóbulos, donde se aprecia de manera más gráfica aquellos lóbulos perjudicados que alcanzan mayor porcentaje de afectación.

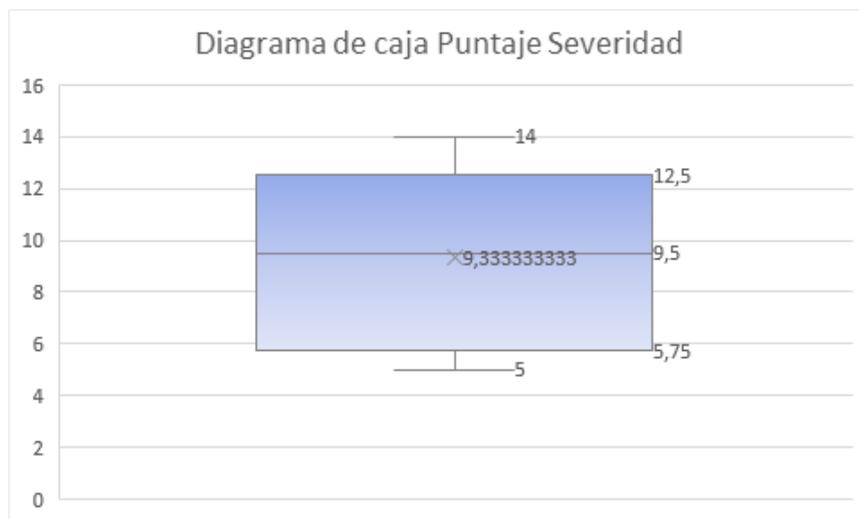


Figura 22. Diagrama de caja realizado con el puntaje total de afectación de los estudios evaluados con una media de 9,3.

De acuerdo a los puntajes totales de severidad obtenidos de los estudios evaluados, estos en conjunto promedian una puntuación de 9,3. En el estudio anteriormente citado (56) se establece el valor 7,5 como puntuación de corte para determinar el grado de severidad grave. De los 6 estudios evaluados en esta AFE, sólo 2 de ellos están bajo dicha puntuación.

5.7.3.- Resultados Tercera dimensión de evaluación:

Por último, la tercera dimensión fue la evaluación de la calidad de la segmentación de manera específica, tomando tres patrones de segmentación (vidrio esmerilado, consolidación y pulmón) para graduar en una escala de Likert de 1 a 5 según grado de cumplimiento de los criterios de calidad de segmentación. A continuación, se muestran en las tablas 11, 12 y 13 resumen de sus resultados y los diagramas de caja de la segmentación separada por patrón y la distribución de cada criterio. Finalmente se muestra el diagrama comparando la distribución de resultados por cada tipo de segmentación.

Tabla 11. Resultados de evaluación por criterios en segmentación en vidrio esmerilado

Puntuación Segmentación en Vidrio Esmerilado				
ID	Criterio Uno	Criterio Dos	Criterio Tres	Total
1	4	3	4	11
2	4	5	5	14
3	3	4	5	12
4	5	4	4	13
5	4	4	5	13
6	4	2	5	11
	24	22	28	

Tabla 12. Resultado de evaluación por criterios en segmentación consolidación

Puntuación Segmentación Consolidación				
ID	Criterio Uno	Criterio Dos	Criterio Tres	Total
1	4	4	4	12
2	5	5	4	14
3	4	4	4	12
4	-	-	-	0
5	-	-	-	0
6	4	4	5	13
	17	17	17	

Tabla 13. Resultado de evaluación por criterios en segmentación pulmón

Puntuación Segmentación Pulmón				
ID	Criterio Uno	Criterio Dos	Criterio Tres	Total
1	4	4	3	11
2	5	4	4	13
3	5	5	5	15
4	4	5	4	13
5	5	5	5	15
6	5	5	5	15
	28	28	26	

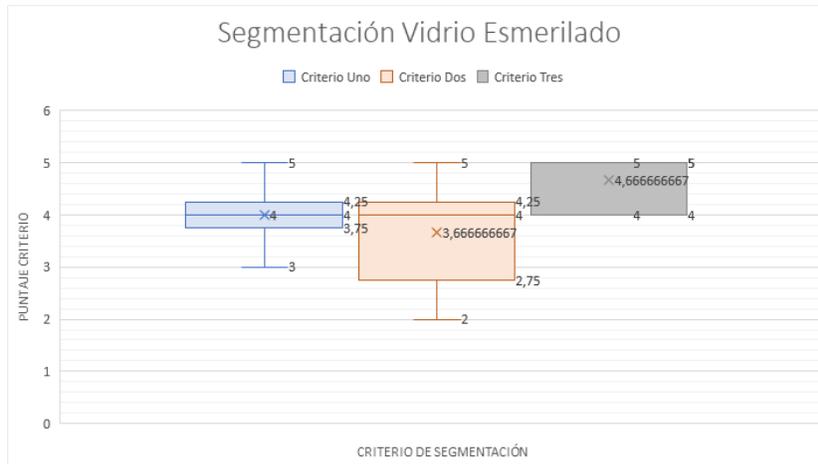


Figura 23. Diagrama de cajas con la distribución de los valores de calificación por criterio, para la segmentación en vidrio esmerilado. El segundo criterio es el que tendría mayor dispersión en los valores en la distribución. En segundo lugar, está el criterio uno con una menor dispersión.

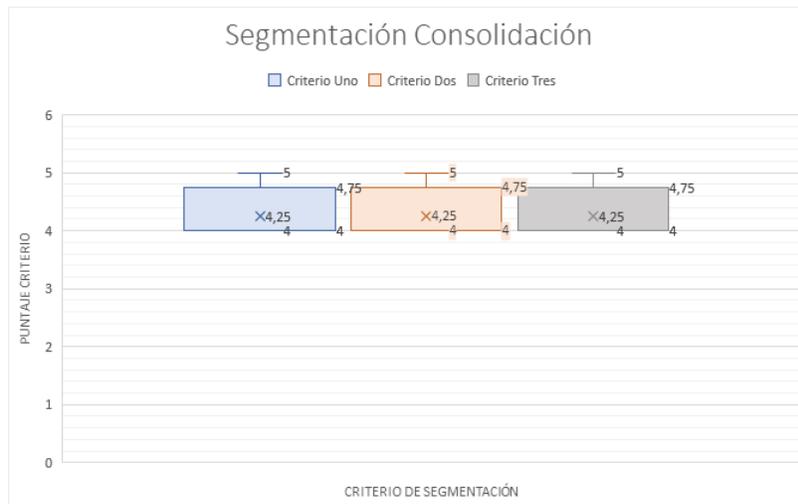


Figura 24. Diagrama de cajas con la distribución de los valores de calificación por criterio, para la segmentación de Consolidación. En general los 3 criterios fueron evaluados de igual forma.

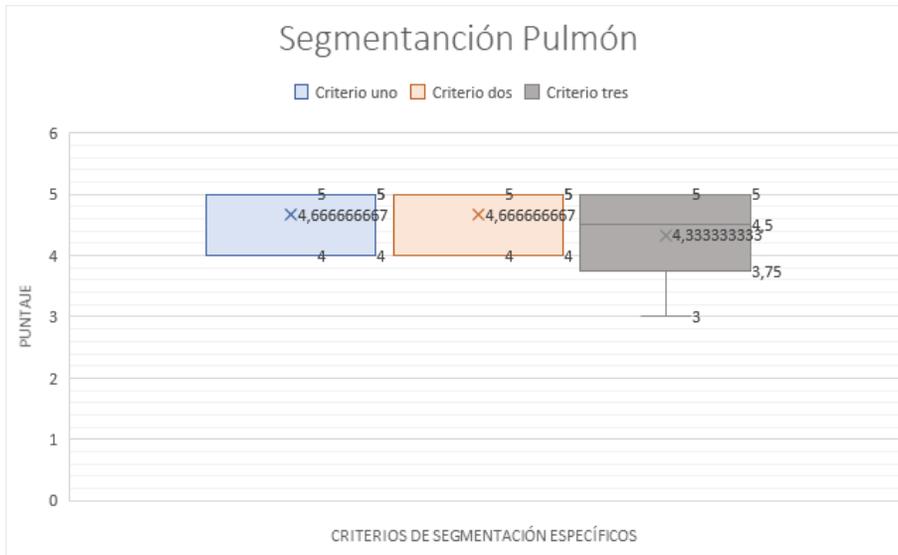


Figura 25. Diagrama de cajas con la distribución de los valores de calificación por criterio, para la segmentación de Pulmón. Es el de mejor valoración de las tres segmentaciones.

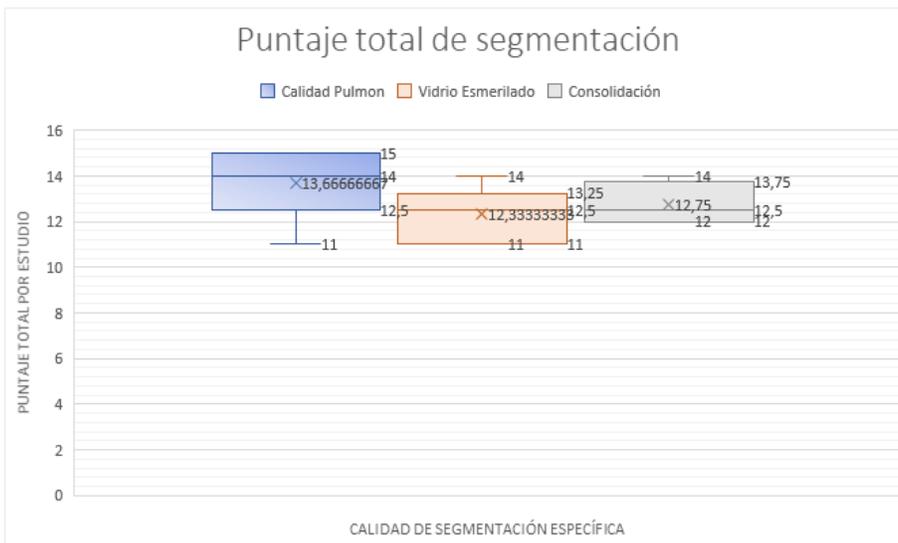


Figura 26. Diagrama de cajas con la distribución de los valores de puntaje total de segmentación específica. En este gráfico se puede ver que el de menor desempeño acorde a la valoración de los evaluadores sería la segmentación en patrón en vidrio esmerilado

Cabe hacer notar que, en lo que respecta a la segmentación de consolidación se tomaron en consideración sólo cuatro estudios, ya que dos de los restantes tenían patrones o la presencia de dicho signo clínico, de acuerdo al criterio de los evaluadores. No obstante, en la segmentación del algoritmo si lo detecta, pero en porcentajes bajos, 0,52% en uno y 6% en el otro, con respecto al total de volumen pulmonar segmentado.

En general, para los seis estudios y para cada una de las segmentaciones evaluadas, según los datos de puntuación, se obtuvieron puntuaciones que fluctuaron entre un 82 y 91% del total.

5.7.4.- Resultados de la encuesta de opinión usuaria.

Los resultados de la encuesta usuaria (ver Anexo 3), establecen que la plataforma es útil para los usuarios, ya que permite realizar la evaluación acorde a los objetivos planteados. Sin embargo, uno de ellos opina que aún necesita un mayor desarrollo para superar ciertas dificultades. En cuanto a la dificultad de uso presentada por la plataforma, hay acuerdo en todos los evaluadores en que no es difícil de manejar, no implica un mayor desafío, aunque requiere de una guía inicial de familiarización. En cuanto al tiempo necesario para realizar la evaluación, este fue considerado entre adecuado a bajo. En términos generales, respecto a la interacción con la plataforma, se evidenció una mayor diversidad de opinión, pero sólo un usuario la consideró compleja y que demanda de apoyo constante. Respecto a la calidad de las imágenes desplegadas en la plataforma, la encuesta muestra que hay acuerdo en que éstas son buenas, pero que en algunos casos se consideró necesario mejorar el tamaño. La velocidad de carga es muy dependiente del tipo de conexión que tenga cada usuario, ya que determina la velocidad de carga de las imágenes. Ante la pregunta de si en un contexto de

evaluación de reportes a terceros la plataforma sería útil para tales efectos, hubo unanimidad en que esto sería posible. En una escala de evaluación de 1 a 7, dos evaluadores la calificaron con nota 6 y otro con nota 5.

6.- DISCUSIÓN

Dentro de los procesos telemédicos, tal como ocurre en cualquier proceso clínico convencional, la calidad es uno de los aspectos importantes que deben estar sujetos a una constante evaluación. Dicha evaluación debiera realizarse mediante herramientas que no sólo permitan asegurar dicho propósito, sino que además permitan mediante, un ciclo virtuoso, ir mejorando dicho aspecto, en consonancia con los propósitos de buenas prácticas y lineamientos estratégicos establecidos en el desarrollo de la telesalud en nuestro país.

Un enfoque de auditoria como el propuesto en este trabajo, que contemple la evaluación simultanea de imágenes y reportes, en un sistema que incorpore herramientas de interoperabilidad, no ha sido encontrado en la literatura hasta la fecha. Existe la revisión externa de reportes, realizada por pares servicio que es ofrecido por empresas de teleradiología como una forma de liberar de esa obligación al control interno de revisión de centros de salud en los EE. UU. (64); sin embargo, no existen trabajos que evalúen y ponderen sus ventajas o desventajas ni entreguen detalles de cómo estas se implementan.

La inclusión de la evaluación de las imágenes dentro del proceso de auditoría se justificó por ser parte y pieza importante del resultado de la atención médica, y como tal fuente de posibles mejoras a la práctica de obtención de dichas imágenes.

En este trabajo de AFE se tomaron en consideración los cambios en el tiempo experimentados por las principales metodologías de evaluación de reportes, comenzando por ajustar los procesos de evaluación con el fin de evitar sesgos en la revisión, así como la incorporación de la revisión doble ciego dentro del diseño. De igual modo, se consideró la incorporación de comités de recalificación en casos discrepantes con el fin de reducir la

subjetividad, ya que esta es una de las principales falencias que presentan los métodos tradicionalmente más utilizados.

Con las metodologías así definidas, se modelaron los procesos finales con el fin de adaptarlos a un entorno de prueba. La fase de implementación fue la etapa que requirió la mayor inversión de tiempo, ya que requirió de la instalación de la infraestructura y herramientas informáticas en un entorno de seguridad necesaria para que los repositorios de imágenes logaran interoperar. Consecutivamente a lo anterior, la programación de la plataforma *web* y base de datos, para que fuera el canal por el cual los evaluadores accedieran a las imágenes y evaluación posterior, significó el proceso más largo ya que implicó el ensayo y error con distintas herramientas informáticas de programación de páginas *web* para finalmente tomar la decisión de programar en *PHP*.

Respecto al caso uso, fue necesario adaptar la metodología de evaluación de imágenes y reportes. El cambio sustancial para efectos del presente trabajo radicó en encontrar una metodología que permitiera evaluar la segmentación, sin modificar de manera dramática ni la estructura ni los procesos de evaluación planteados. Con el fin de obtener la mayor cantidad de información posible acerca de la calidad de la segmentación del algoritmo, se realizó una extensa revisión de la literatura, lo que permitió encontrar métodos para evaluar y comparar la calidad de la segmentación, así como su concordancia con el reporte. Esto último constituyó el mayor cambio implementado, ya que se tuvo que adaptar un método de evaluación de reporte interpretativo a uno en que mediante la interpretación y valoración de severidad se comparara a una medición objetiva de segmentación.

Con respecto a los resultados de la evaluación, sólo fue posible la evaluación de 6 estudios, la que fue llevada a cabo por 3 becados de la especialidad de radiología, evaluando 2 estudios

cada uno. No obstante, aún con este bajo número de estudios, se logró obtener una adecuada cantidad de información para los propósitos de esta AFE.

En la primera dimensión de evaluación se analizó la calidad de la segmentación en función al tiempo necesario para rectificarla en escala Likert. De los 6 estudios considerados, 2 fueron evaluados como deficientes, es decir, bajo el umbral definido previamente. Coincidentemente estudios son los de mayor porcentaje de afectación valorado por ambos métodos. En este punto se podría hipotetizar que mientras mayor sea el volumen de compromiso la cantidad de posibles errores de segmentación es mayor, y por lo tanto, será también mayor el tiempo en corregirlo, lo que podría explicar la calificación del evaluador.

La segunda dimensión de evaluación consistió en asignar puntajes de severidad de afectación por lóbulos en cada uno de los estudios. En esta evaluación se midió el puntaje total y se llevó a porcentaje de severidad para compararlo con el nivel de compromiso pulmonar obtenido de la segmentación. En general el método fue capaz de cuantificar el nivel de severidad manteniendo el mismo orden, es decir, tanto los valores mayores como los menores coincidían. Sin embargo, en la diferencia neta entre porcentajes se encontró una diferencia mayor, esto debido posiblemente a que los intervalos utilizados eran de rangos muy amplios. Para evaluar la concordancia se utilizaron dos métodos, si bien ambos entregaron buena concordancia, el que mejor resultado entregó fue el coeficiente de correlación intraclase, aunque con intervalos de confianza muy amplios dado lo limitado de la muestra.

La tercera y última dimensión de evaluación entregó en general un buen cumplimiento en los criterios de segmentación, siendo el de mejor desempeño la segmentación de pulmón seguida por la de consolidación. No obstante, es necesario señalar que los evaluadores no detectaron patrón en consolidación en dos de los estudios, pero el algoritmo si lo hizo, por lo que queda abierta la pregunta si lo que segmenta el algoritmo en ciertos casos constituye un falso

positivo de la segmentación o constituye un falso negativo por parte de los evaluadores. Se requieren evaluaciones con un mayor número, tanto de estudios como de evaluadores para tratar de dilucidar este punto.

La segmentación de vidrio esmerilado es la que tuvo un menor desempeño respecto a las otras, en particular en lo que respecta a su segundo criterio. Los estudios con menor desempeño en este criterio coinciden en que son los de mayor porcentaje de compromiso pulmonar. Se ha descrito en la literatura (55) que en evaluaciones llevadas a cabo con diferentes algoritmos de segmentación, estos tienden a disminuir su capacidad de segmentación mientras mayor sea el nivel de severidad de la enfermedad. Sin embargo, lo limitado de la muestra solo permite plantear conclusiones hipotéticas.

Respecto a los resultados de la encuesta usuaria, en general los usuarios opinaron que la plataforma podría ser útil en un contexto de auditoría pero que aún necesita un mayor desarrollo para superar ciertas dificultades, en particular en el manejo del tamaño de las imágenes. En cuanto a la dificultad de uso de la plataforma, hay acuerdo en que esta no fue difícil de manejar, aunque sería mejor contar con una guía inicial. Su manejo fue rápido, ya que no hubo que dedicarle mucho tiempo a la revisión y la calidad de las imágenes fue adecuada para llevar a cabo la tarea. La velocidad de carga es muy dependiente del tipo de conexión ya que esta determina la velocidad de carga de las imágenes. En general la plataforma tuvo una evaluación satisfactoria.

Sería interesante y queda pendiente para estudios futuros, constatar mediante el uso de lo propuesto en este trabajo, medir cual sería el impacto que tendría sobre la calidad el hecho de llevar a cabo una estrategia de auditoría en un servicio de telereportes. En un estudio prospectivo realizado en Rusia mediante el uso centralizado de su sistema de información,

se utilizó un método de auditoría similar con herramientas de telemedicina (8) y que se llevó a cabo durante un año, mostró que en función de sus resultados se realizaron acciones de mejora tales como capacitaciones remotas para el personal involucrado. Posterior a esto, se llevó a cabo una nueva auditoría en donde se logró constatar una disminución de las “discrepancias” y un aumento en informes “sin discrepancia”, ambos en porcentajes estadísticamente significativos.

7.- CONCLUSIÓN

Mediante la realización de esta AFE se pudo establecer y proponer un nuevo modelo para la auditoría remota de telereportes, basado en imágenes clínicas y su integración a dichos procesos mediante la utilización de herramientas de interoperabilidad. El modelo propuesto se pudo someter a prueba mediante un Caso de Uso el cual fue testeado y evaluado por usuarios revisores.

A partir de esto se puede concluir que se logró determinar un método de evaluación de imágenes y reportes que se logre ajustar a las necesidades de un servicio de telereportes.

Se modelaron los procesos de telereportes y como estos se podrían acoplar a la estructura de un modelo de auditoría mediante el uso de herramientas informáticas y estándares de interoperabilidad.

La implementación del Caso de Uso mediante una plataforma *web* permitió establecer la utilidad del sistema en un ambiente de auditoría tanto de imágenes como de reportes.

Si bien la prueba no se realizó en un Caso de Uso real de teleinforme, el marco de trabajo propuesto para dichos fines en esta AFE se adaptó adecuadamente a las necesidades de evaluación del Caso de Uso alternativo planteado en este trabajo. Esto permite concluir que el modelo propuesto resulta versátil, ya que mostró ser adaptable y a su vez escalable a distintas situaciones y necesidades de evaluación. Junto a ello, se logró obtener información valiosa respecto a la evaluación de la calidad de la segmentación del proyecto IA TRad.

Finalmente, este trabajo entrega un aporte al establecer un punto de partida en el desarrollo de herramientas para el aseguramiento de la calidad y mejora continua en procesos de teleinformes basados en imágenes médicas. Si bien queda abierto el desafío de implementarlo en un Caso de Uso real, la alternativa planteada en este trabajo permitió cumplir con los

objetivos de evaluación. Quedará para estudios futuros la aplicación del modelo en contextos reales de evaluación y el impacto que tendría en la mejora de la calidad de estos procesos telemédicos.

ANEXOS

ANEXO 1 Plataforma web para la evaluación de imágenes.

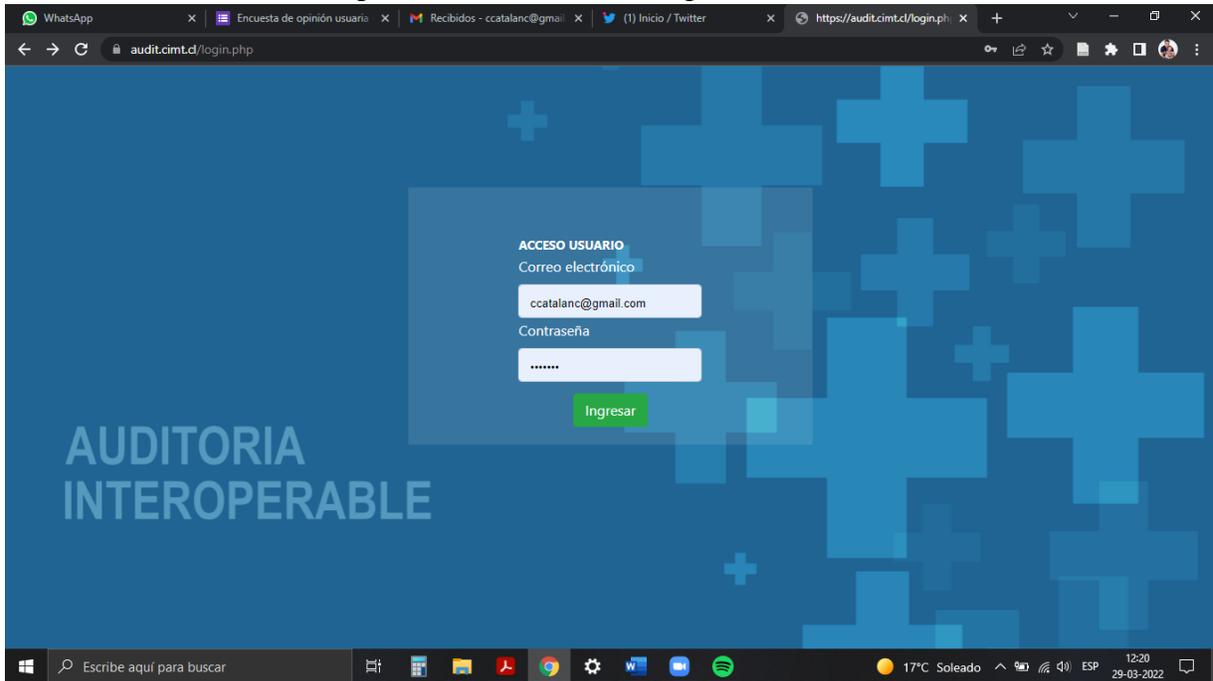


Figura 1 Página de inicio audit.cimt.cl/login.php para el acceso de usuarios evaluadores mediante correo y contraseña. Ejemplo: usuario ccatalanc@gmail.com contraseña claudio

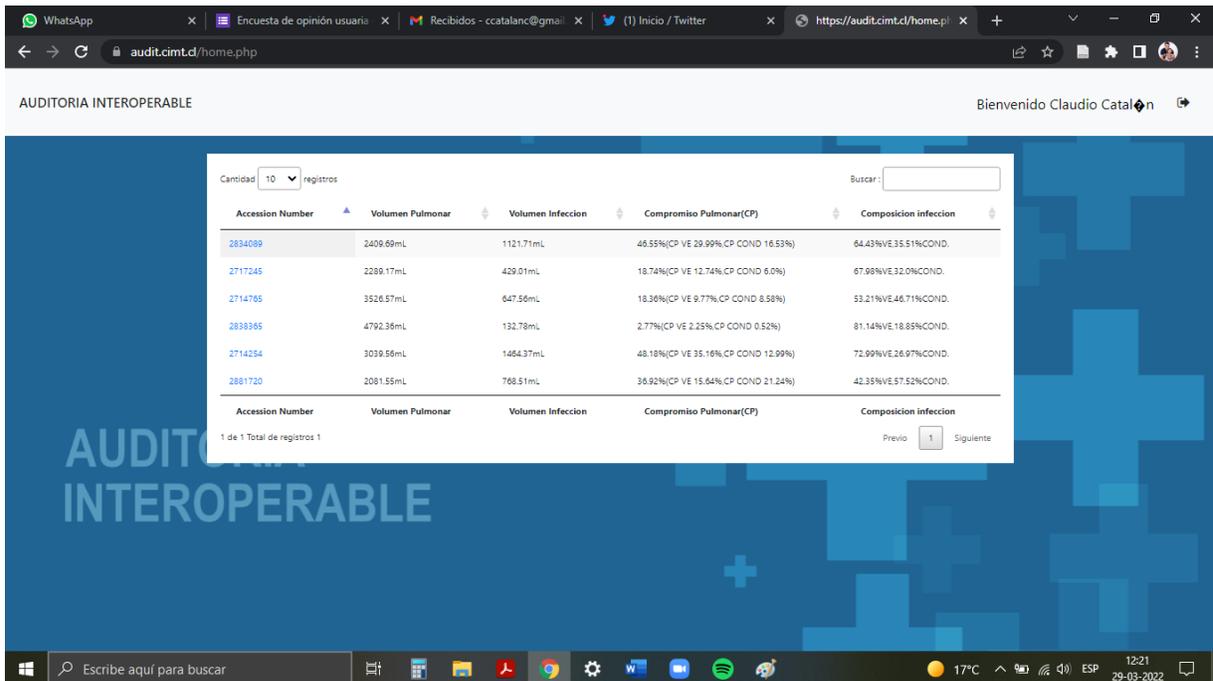


Figura 2 Una vez ingresado, usuario tiene acceso a lista de estudios para evaluar mediante un número de acceso.

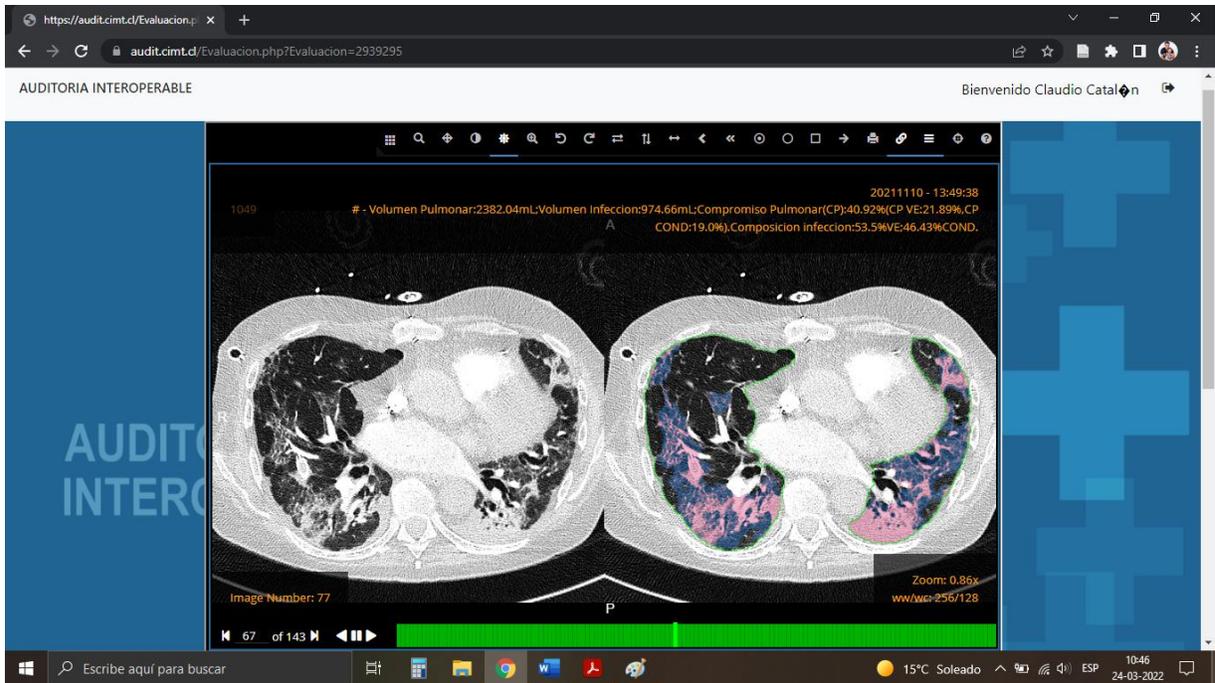


Figura 3 Al pinchar en un estudio se abre la imagen correspondiente mediante un visor web del PACS embebido en la página.

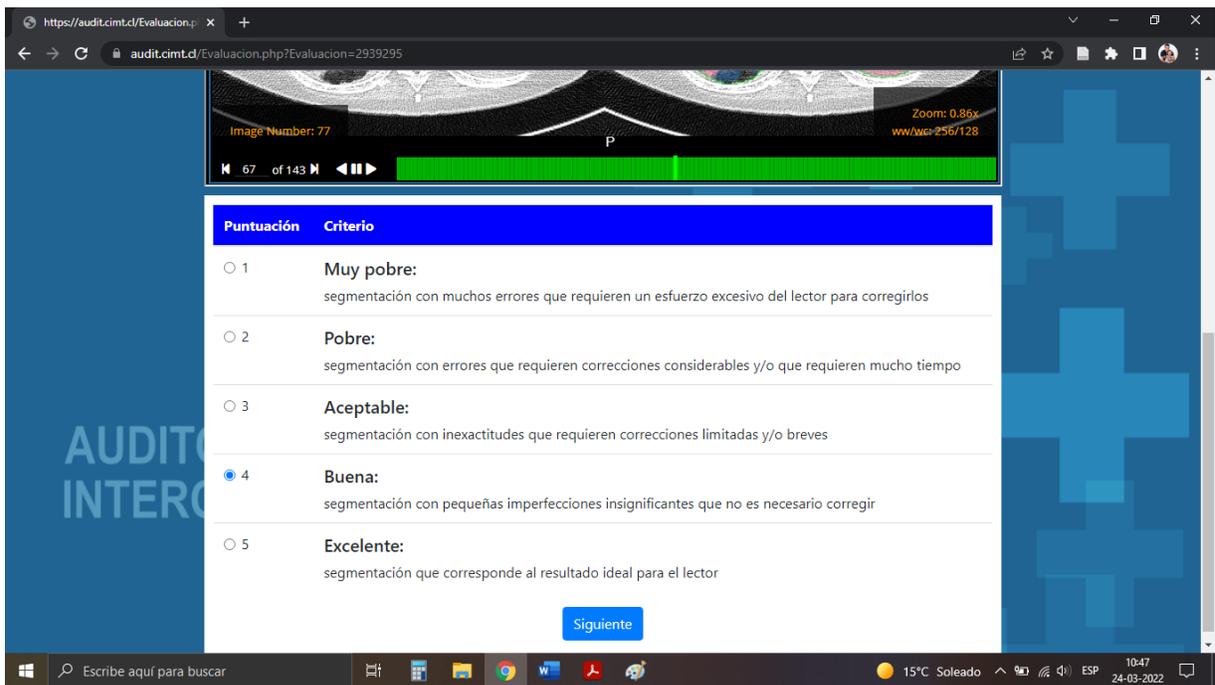


Figura 4 Junto a la imagen en la parte inferior muestra el primer formulario de respuesta que corresponde a la evaluación general.

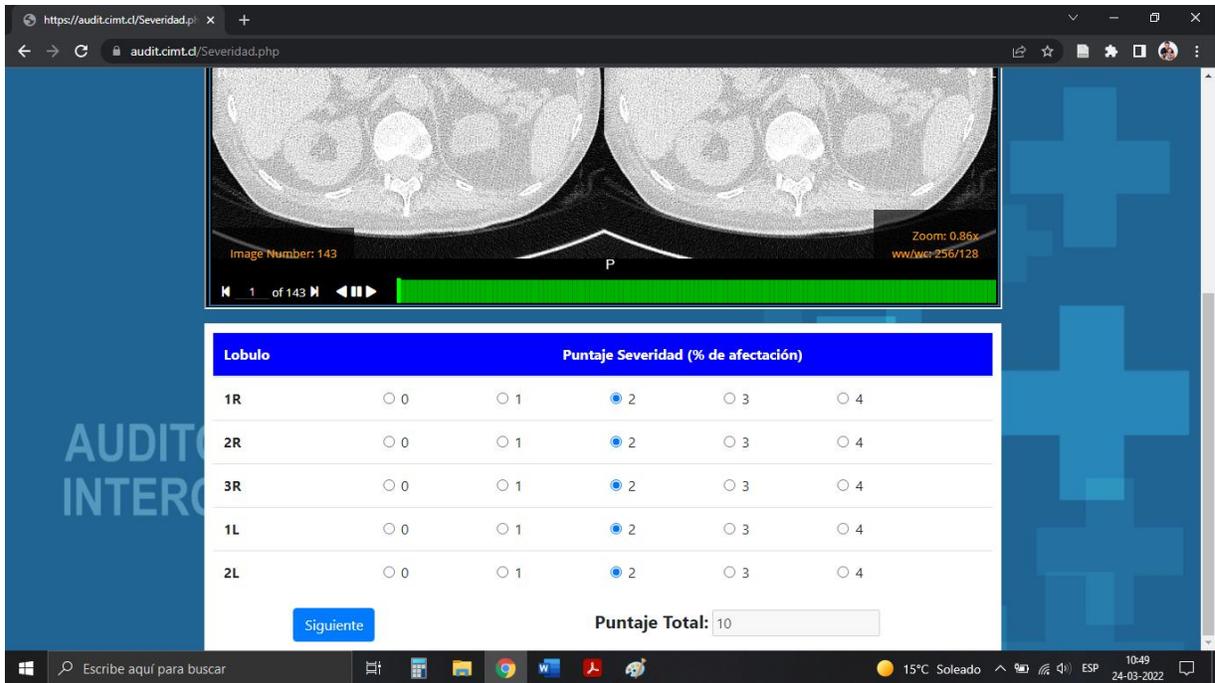


Figura 5 Luego de respondido el primer formulario y dar "siguiente" se despliega el segundo formulario, el de puntaje de severidad.

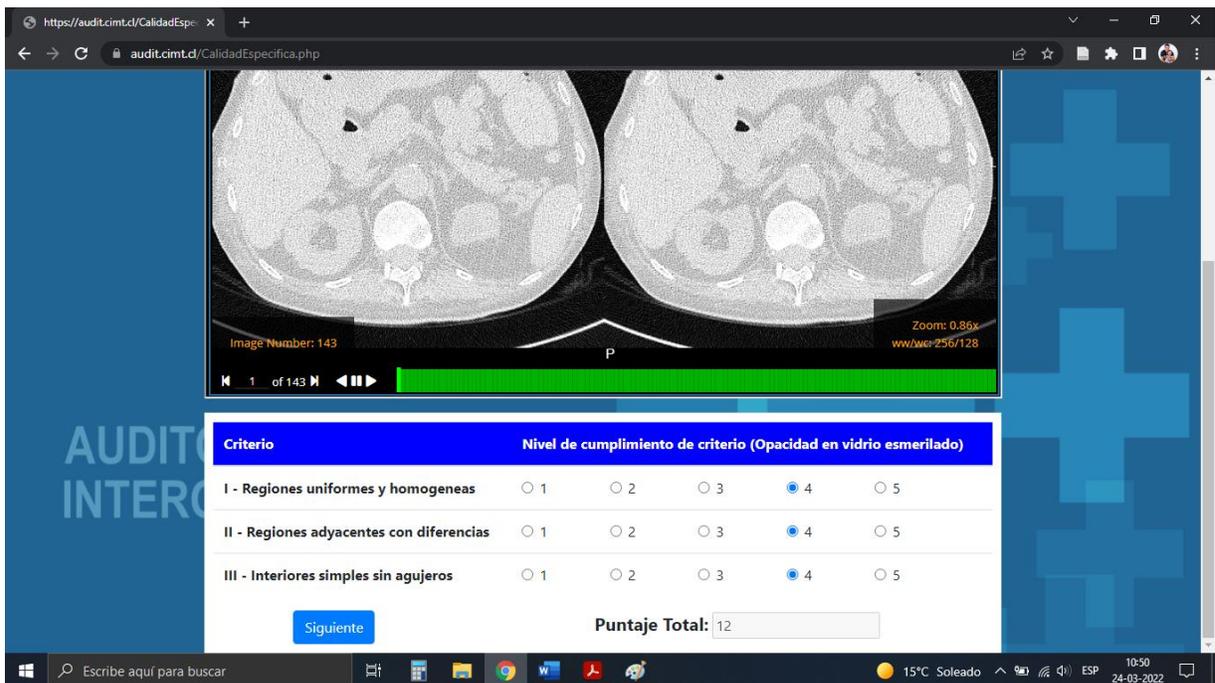


Figura 6 Una vez respondido pasa al siguiente formulario de evaluación específica.

The screenshot shows a web browser window with the URL `https://audit.cimt.d/CalidadEspecificaPulmon.php`. The main content area displays two axial CT scan images of the chest. Below the images is a navigation bar with 'Image Number: 143', a 'P' marker, and a green progress bar. A table below the images lists three criteria for evaluation, each with five radio button options (1-5). The 'Finalizar' button is located below the table, and the 'Puntaje Total' is 15.

Criterio	Nivel de cumplimiento de criterio (Pulmon)				
I - Regiones uniformes y homogéneas	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input checked="" type="radio"/> 5
II - Regiones adyacentes con diferencias	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input checked="" type="radio"/> 5
III - Interiores simples sin agujeros	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input checked="" type="radio"/> 5

Finalizar Puntaje Total: 15

Figura 7 Finalmente al responder el último formulario se da fin a la evaluación y automáticamente vuelve a la página en donde están desplegados el resto de los estudios.

ANEXO 2 Encuesta de opinión usuaria (62).

Encuesta de opinión usuaria

Formulario de preguntas respecto a *front end* y su uso

1. Respecto a la utilidad de la plataforma de revisión de imágenes en un contexto de auditoria

Marca solo un óvalo.

- Es útil permite evaluar las imágenes acorde al objetivo planteado.
- Es potencialmente útil pero requiere mayor desarrollo.
- Poco útil para la evaluación de imágenes en el contexto planteado.

2. Respecto al nivel de dificultad en el uso de la plataforma

Marca solo un óvalo.

- Fácil no implica un mayor desafío.
- Adecuado implica un desafío prudente.
- Difícil implica un desafío mayor.

3. Respecto al tiempo que hay que dedicarle a la revisión de imágenes.

Marca solo un óvalo.

- Bajo su desarrollo implica poco tiempo.
- Adecuado requiere una cantidad prudente de tiempo.
- Excesivo su desarrollo implica mucho tiempo.

4. En términos generales, la interacción con la plataforma y la revisión:

Marca solo un óvalo.

- Es intuitiva y fácil no requiere apoyo adicional.
- Es necesario un apoyo inicial pero luego es sencillo.
- Es compleja y es necesario un apoyo constante para su uso.

5. Respecto a la calidad de las imágenes, se puede decir que:

Marca solo un óvalo.

- Es adecuada posee buen tamaño y aumento lo que permite ver detalles en la imagen.
- Es variable hay algunas buenas y en otras es necesario mejorar tamaño de imágenes.
- Inadecuada baja calidad y tamaño, no permite ver detalles.

6. Respecto a la velocidad de carga del estudio, se puede decir que:

Marca solo un óvalo.

- Es adecuada no dificulta para llevar a cabo revision del estudio.
- No es problema, sin embargo podría ser más rápido.
- Es deficiente se demora y enlentece el proceso.

7. En un contexto de revisión de reportes elaborado por terceros, se puede decir que:

Marca solo un óvalo.

- Sería posible reportar y medir algún grado de discrepancia con el reporte original usando esta plataforma.
- Sería posible solo en algunos casos.
- No sería posible bajo dicho contexto.

8. ¿Cómo calificaría esta plataforma para la evaluación de imágenes, siendo 1 la nota más baja y 7 la más alta?

Marca solo un óvalo.

- | | | | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| <input type="radio"/> |

9. ¿Qué fué lo que le gustó de esta plataforma?

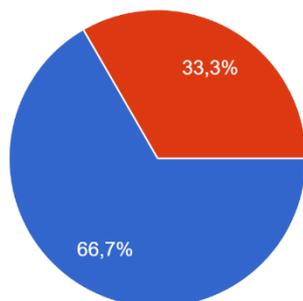
10. ¿Qué mejoras incluiría en la plataforma?

11. Si lo desea, agregue un comentario general de su experiencia utilizando la plataforma.

ANEXO 3 Respuestas de Encuesta de Opinión Usuaría.

Respecto a la utilidad de la plataforma de revisión de imágenes en un contexto de auditoría

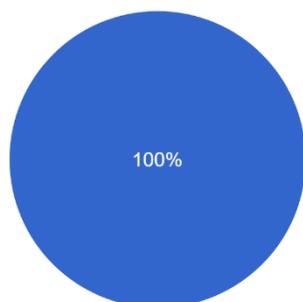
3 respuestas



- Es útil, permite evaluar las imágenes acorde al objetivo planteado.
- Es potencialmente útil, pero requiere mayor desarrollo.
- Poco útil para la evaluación de imágenes en el contexto planteado.

Respecto al nivel de dificultad en el uso de la plataforma

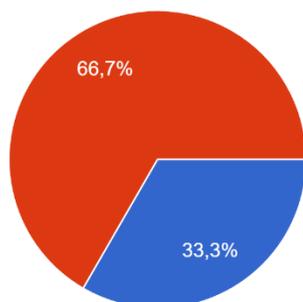
3 respuestas



- Fácil, no implica un mayor desafío.
- Adecuado, implica un desafío prudente.
- Difícil, implica un desafío mayor.

Respecto al tiempo que hay que dedicarle a la revisión de imágenes.

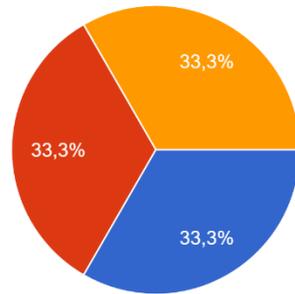
3 respuestas



- Bajo, su desarrollo implica poco tiempo.
- Adecuado, requiere una cantidad prudente de tiempo.
- Excesivo, su desarrollo implica mucho tiempo.

En términos generales, la interacción con la plataforma y la revisión:

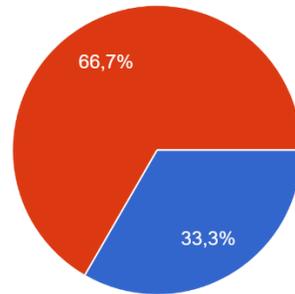
3 respuestas



- Es intuitiva y fácil, no requiere apoyo adicional.
- Es necesario un apoyo inicial, pero luego es sencillo.
- Es compleja y es necesario un apoyo constante para su uso.

Respecto a la calidad de las imágenes, se puede decir que:

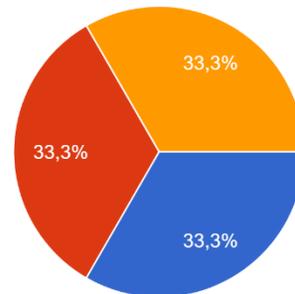
3 respuestas



- Es adecuada, poseen buen tamaño y aumento, lo que permite ver detalles en la imagen.
- Es variable, hay algunas buenas y en otras es necesario mejorar tamaño de imágenes.
- Inadecuada, baja calidad y tamaño no permite ver detalles.

Respecto a la velocidad de carga del estudio, se puede decir que:

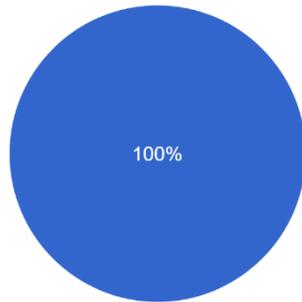
3 respuestas



- Es adecuada, no dificulta para llevar a cabo revisión del estudio.
- No es problema, sin embargo podría ser más rápido.
- Es deficiente, se demora y enlentece el proceso.

En un contexto de revisión de reportes elaborado por terceros, se puede decir que:

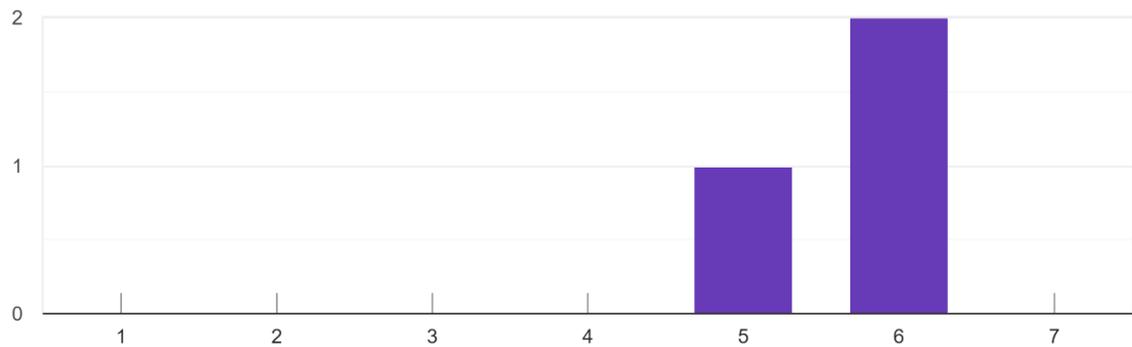
3 respuestas



- Sería posible reportar y medir algún grado de discrepancia con el reporte original usando esta plataforma.
- Sería posible sólo en algunos casos.
- No sería posible bajo dicho contexto.

¿Cómo calificaría esta plataforma para la evaluación de imágenes, siendo 1 la nota más baja y 7 la más alta?

3 respuestas



¿Qué fué lo que le gustó de esta plataforma?

3 respuestas

que mostraba las imagenes con y sin segmentacion lado a lado y enlazadas para navegar entre los slides de manera simultanea a cada lado

Fue muy sencillo comprenderla

La rapidez de carga y la ventaja de disponer de multiplataformas de visualización

¿Qué mejoras incluiría en la plataforma?

2 respuestas

las preguntas no me parecieron claras, y no es tan intuitivo el sistema que permite mostrar la explicación detallada de las preguntas.

Lo otro es que sugeriría que solo las preguntas (no necesariamente también las imágenes) se vayan cargando a medida que se avanza en el cuestionario. Esto permitiría ahorrar tiempo de carga de las imágenes cuando estas son las mismas entre las diferentes preguntas.

Velocidad

Si lo desea, agregue un comentario general de su experiencia utilizando la plataforma.

2 respuestas

En general bien pero el tiempo que toma responderla es muy dependiente de la conexión.

Para realizar la revisión, se deberían incluir en la misma visión del examen una pequeña explicación del objetivo a buscar.

REFERENCIAS

1. World Health Organization. WHO guideline: recommendations on digital interventions for health system strengthening [Internet]. Geneva: World Health Organization; 2019 [citado 18 de enero de 2022]. Disponible en: <https://apps.who.int/iris/handle/10665/311941>
2. Narváez P. Programa Nacional de Telesalud MINSAL. Ministerio de Salud - Subsecretaría de Redes Asistenciales - División de Gestión de la Red Asistencial. Marzo de 2018. <http://biblioteca.digital.gob.cl/handle/123456789/3635>.
3. Quality Management Systems - EN ISO 9001:2015 for healthcare [Internet]. [citado 24 de junio de 2022]. Disponible en: <https://www.sis.se/api/document/preview/8024286/>
4. cens-gua-teleconsulta-en-tiempos-de-pandemia.pdf [Internet]. [citado 10 de enero de 2022]. Disponible en: <https://cens.cl/wp-content/uploads/2020/04/cens-gua-teleconsulta-en-tiempos-de-pandemia.pdf>
5. Proyecto Lineamientos para el desarrollo de la Telemedicina y Telesalud en Chile – Bien Público Estratégico financiado con aportes del Fondo de Innovación para la Competitividad InnovaChile Corfo [Internet]. [citado 30 de marzo de 2022]. Disponible en: <https://lineamientostelesalud.cl/>
6. Brady AP. Error and discrepancy in radiology: inevitable or avoidable? Insights Imaging. febrero de 2017;8(1):171-82.
7. Ortega T D, García M C, Zárate C M. PERSPECTIVAS DE LOS RADIÓLOGOS CON RESPECTO AL ERROR Y A LA TRIBUNALIZACION EN EL EJERCICIO DE LA ESPECIALIDAD EN CHILE. Rev Chil Radiol [Internet]. 2008 [citado 10 de enero de 2022];14(3). Disponible en: http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0717-93082008000300010&lng=en&nrm=iso&tlng=en
8. Morozov S. Telemedicine-based system for quality management and peer review in radiology. 2018;5.
9. ITMS Telemedicina [Internet]. ITMS Chile. [citado 24 de enero de 2022]. Disponible en: <https://www.itms.cl>
10. Rencoret S. G. AUDITORIA MEDICA: DEMANDAS Y RESPONSABILIDAD POR NEGLIGENCIAS MEDICAS. GESTION DE CALIDAD: RIESGOS Y CONFLICTOS. Rev Chil Radiol [Internet]. 2003 [citado 10 de enero de 2022];9(3). Disponible en: http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0717-93082003000300008&lng=en&nrm=iso&tlng=en
11. sintesis.med.uchile.cl - Auditoría y acreditación [Internet]. [citado 10 de enero de 2022]. Disponible en: <http://sintesis.med.uchile.cl/index.php/profesionales/informacion-para-profesionales/medicina/condiciones-clinicas2/otorrinolaringologia/1084-7-01-3-047>

12. Vyas D. Clinical peer review in the United States: History, legal development and subsequent abuse. *World J Gastroenterol*. 2014;20(21):6357.
13. Ministro de Salud expuso sobre el Hospital Digital en seminario universitario [Internet]. Ministerio de Salud – Gobierno de Chile. [citado 10 de enero de 2022]. Disponible en: <https://qa.minsal.cl/ministro-de-salud-expuso-sobre-el-hospital-digital-en-seminario-universitario/>
14. Borgstede JP, Lewis RS, Bhargavan M, Sunshine JH. RADPEER Quality Assurance Program: A Multifacility Study of Interpretive Disagreement Rates. :7.
15. Wootton R. Experience with Quality Assurance in Two Store-and-Forward Telemedicine Networks. *Front Public Health*. 2015;3:10.
16. Larson DB, Donnelly LF, Podberesky DJ, Mellow AC, Sharpe RE, Kruskal JB. Peer Feedback, Learning, and Improvement: Answering the Call of the Institute of Medicine Report on Diagnostic Error. *Radiology*. abril de 2017;283(1):231-41.
17. Miñano Romero JA. Introducción al control de calidad en radiología digital. Madrid: Sociedad Española de Física Médica; 2013.
18. Månsson LG. METHODS FOR THE EVALUATION OF IMAGE QUALITY: A REVIEW. :12.
19. Yalcin A, Olgar T, Sancak T, Atac GK, Akyar S. Correlation Between Physical Measurements and Observer Evaluations of Image Quality in Digital Chest Radiography. :29.
20. Precht H. Radiographers' perspectives' on Visual Grading Analysis as a scientific method to evaluate image quality. 2019;5.
21. Tesselaar E. CLINICAL AUDIT OF IMAGE QUALITY IN RADIOLOGY USING VISUAL GRADING CHARACTERISTICS ANALYSIS. :7.
22. Ludewig E, Richter A, Frame M. Diagnostic imaging – evaluating image quality using visual grading characteristic (VGC) analysis. *Vet Res Commun*. 2010;7.
23. Båth M, Månsson LG. Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation. *Br J Radiol*. marzo de 2007;80(951):169-76.
24. European Commission. European guidelines on quality criteria for diagnostic radiographic images: Eur 16260 EN. European Commission; 1996.
25. Padole AM, Sagar P, Westra SJ, Lim R, Nimkin K, Kalra MK, et al. Development and validation of image quality scoring criteria (IQSC) for pediatric CT: a preliminary study. *Insights Imaging*. diciembre de 2019;10(1):95.
26. Hakansson M, Svensson S, Zachrisson S, Svalkvist A, Bath M, Mansson LG. VIEWDEX: an efficient and easy-to-use software for observer performance studies. *Radiat Prot Dosimetry*. 1 de abril de 2010;139(1-3):42-51.

27. Boita J, van Engen RE, Mackenzie A, Tingberg A, Bosmans H, Bolejko A, et al. Validation of a candidate instrument to assess image quality in digital mammography using ROC analysis. *Eur J Radiol.* junio de 2021;139:109686.
28. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol.* abril de 1989;129(4):687-702.
29. Hubbard LD, Brothers RJ, King WN, Clegg LX, Klein R, Cooper LS, et al. Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the atherosclerosis risk in communities study¹¹The authors have no proprietary interest in the equipment and techniques described in this article. *Ophthalmology.* diciembre de 1999;106(12):2269-80.
30. Lamirel C, Bruce BB, Wright DW, Delaney KP, Newman NJ, Biousse V. Quality of Nonmydriatic Digital Fundus Photography Obtained by Nurse Practitioners in the Emergency Department: The FOTO-ED Study. *Ophthalmology.* marzo de 2012;119(3):617-24.
31. Patton N, Aslam TM, MacGillivray T, Deary IJ, Dhillon B, Eikelboom RH, et al. Retinal image analysis: Concepts, applications and potential. *Prog Retin Eye Res.* enero de 2006;25(1):99-127.
32. Wang S, Jin K, Lu H, Cheng C, Ye J, Qian D. Human Visual System-Based Fundus Image Quality Assessment of Portable Fundus Camera Photographs. *IEEE Trans Med Imaging.* abril de 2016;35(4):1046-55.
33. Pires Dias JM, Oliveira CM, da Silva Cruz LA. Retinal image quality assessment using generic image quality indicators. *Inf Fusion.* septiembre de 2014;19:73-90.
34. Whorms DS, Giess CS, Golshan M, Freedman RA, Bunnell CA, Alper EC, et al. Clinical Impact of Second Opinion Radiology Consultation for Patients With Breast Cancer. *J Am Coll Radiol.* junio de 2019;16(6):814-23.
35. Agrawal A, Koundinya DB, Raju JS, Agrawal A, Kalyanpur A. Utility of contemporaneous dual read in the setting of emergency teleradiology reporting. *Emerg Radiol.* abril de 2017;24(2):157-64.
36. Moriarity AK, Hawkins CM, Geis JR, Dreyer KJ, Kamer AP, Khandheria P, et al. Meaningful Peer Review in Radiology: A Review of Current Practices and Potential Future Directions. *J Am Coll Radiol.* diciembre de 2016;13(12):1519-24.
37. To Err Is Human: Building a Safer Health System [Internet]. Washington, D.C.: National Academies Press; 2000 [citado 11 de enero de 2022]. Disponible en: <http://www.nap.edu/catalog/9728>
38. Jackson VP, Cushing T, Abujudeh HH, Borgstede JP, Chin KW, Grimes CK, et al. RADPEER™ Scoring White Paper. *J Am Coll Radiol.* enero de 2009;6(1):21-5.
39. Committee on Diagnostic Error in Health Care, Board on Health Care Services, Institute of Medicine, The National Academies of Sciences, Engineering, and Medicine. Improving

- Diagnosis in Health Care [Internet]. Balogh EP, Miller BT, Ball JR, editores. Washington, D.C.: National Academies Press; 2015 [citado 14 de mayo de 2021]. Disponible en: <http://www.nap.edu/catalog/21794>
40. Goldberg-Stein S, Frigini LA, Long S, Metwalli Z, Nguyen XV, Parker M, et al. ACR RADPEER Committee White Paper with 2016 Updates: Revised Scoring System, New Classifications, Self-Review, and Subspecialized Reports. *J Am Coll Radiol*. agosto de 2017;14(8):1080-6.
 41. Renfrew DL, Franken EA, Berbaum KS, Weigelt FH, Abu-Yousef MM. Error in radiology: classification and lessons in 182 cases presented at a problem case conference. *Radiology*. abril de 1992;183(1):145-50.
 42. Gálvez M M, Montoya M C. Error en el informe radiológico: La paradoja del elefante en la habitación y otros tropiezos. *Rev Chil Radiol*. 2017;23(2):80-9.
 43. Shreve J. Clinical interoperability and FHIR [Internet]. Dynamic Health IT - Quality Measure and Interoperability Solutions. 2015 [citado 21 de marzo de 2022]. Disponible en: http://dynamichealthit.blogspot.com/2015_12_01_archive.html
 44. Benson T, Grieve G. Principles of Health Interoperability: SNOMED CT, HL7 and FHIR [Internet]. Cham: Springer International Publishing; 2016 [citado 11 de enero de 2022]. (Health Information Technology Standards). Disponible en: <http://link.springer.com/10.1007/978-3-319-30370-3>
 45. Ríos Melo, Elvis. Interoperabilidad de los Sistemas Informáticos en el Área Salud. Memoria para optar al grado de Ingeniero Civil Informático. Universidad de Concepción.
 46. ¿Qué es una API? [Internet]. [citado 21 de marzo de 2022]. Disponible en: <https://www.redhat.com/es/topics/api/what-are-application-programming-interfaces>
 47. What is a REST API? Definition and Principles - Seobility Wiki [Internet]. [citado 30 de marzo de 2022]. Disponible en: https://www.seobility.net/en/wiki/REST_API
 48. Healthcare N. Mirth® Connect [Internet]. NextGen Healthcare. [citado 21 de marzo de 2022]. Disponible en: <https://www.nextgen.com/products-and-services/integration-engine>
 49. Orthanc - DICOM Server [Internet]. [citado 21 de marzo de 2022]. Disponible en: <https://www.orthanc-server.com/>
 50. Understanding DICOM with Orthanc — Orthanc Book documentation [Internet]. [citado 21 de marzo de 2022]. Disponible en: <https://book.orthanc-server.com/dicom-guide.html>
 51. Eisenberg RL, Cunningham ML, Siewert B, Kruskal JB. Survey of Faculty Perceptions Regarding a Peer Review System. *J Am Coll Radiol*. abril de 2014;11(4):397-401.
 52. Calculadora de tamaño muestral para indicadores de calidad - Calculadora de tamaño muestral para indicadores de calidad [Internet]. Observatorio de Calidad en Salud. Superintendencia de Salud, Gobierno de Chile. [citado 13 de marzo de 2022]. Disponible en: <http://www.supersalud.gob.cl/observatorio/671/w3-article-7652.html>

53. Härtel-Gründler S. IA-TRad Chile: Inteligencia Artificial para Apoyo Diagnóstico y Procesos de Atención en Radiología y Tele-Radiología Clínica. 2020 [citado 3 de septiembre de 2021]; Disponible en: <http://repositorio.conicyt.cl/handle/10533/249886>
54. Zhang H, Fritts JE, Goldman SA. Image segmentation evaluation: A survey of unsupervised methods. *Comput Vis Image Underst.* mayo de 2008;110(2):260-80.
55. Berta L, Rizzetto F, De Mattia C, Lizio D, Felisi M, Colombo PE, et al. Automatic lung segmentation in COVID-19 patients: Impact on quantitative computed tomography analysis. *Phys Med.* julio de 2021;87:115-22.
56. Wasilewski P, Mruk B, Mazur S, Pótorak-Szymczak G, Sklinda K, Walecki J. COVID-19 severity scoring systems in radiological imaging – a review. *Pol J Radiol.* 2020;85(1):361-8.
57. Li K, Fang Y, Li W, Pan C, Qin P, Zhong Y, et al. CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). *Eur Radiol.* agosto de 2020;30(8):4407-16.
58. Haralick RM, Shapiro LG. Image segmentation techniques. *Comput Vis Graph Image Process.* enero de 1985;29(1):100-32.
59. Gatica JL, Bertoló S, Morales E, Espinoza M, Contreras C. Tele dermatología en Chile, un aporte a la atención primaria de salud. *Piel.* marzo de 2015;30(3):148-54.
60. Minsal, Orientación Técnica Programa de Salud Cardiovascular, 2017.
61. Farias Araya, M. Satisfacción De Usuarios Clínicos Y Administrativos Con Aspectos De Calidad De Servicios De Telerradiología. Tesis para optar a grado Magíster en Informática Médica. Universidad de Chile. Año 2014.
62. Cubillos Moreno, P. Fortalecimiento de competencias diagnósticas en citopatología ginecológica mediante Microscopía Virtual y Módulos de Aprendizaje Electrónico-Adaptativo. Tesis para optar a grado Magíster en Informática Médica. Universidad de Chile. Año 2021.
63. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med.* junio de 2016;15(2):155-63.
64. Radiology Workflow Innovation & Technology [Internet]. vRad. [citado 24 de junio de 2022]. Disponible en: <https://www.vrad.com/service/peer-review/>