

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

AUTOMATIZACIÓN DEL PROCESO DE GESTIÓN Y EVALUACIÓN DE LA CALIDAD
DE LAS DIGITALIZACIONES REALIZADAS EN EL DEPARTAMENTO DE
CONSERVACIÓN DE LA DIRECCIÓN GENERAL DEL ARCHIVO NACIONAL
UTILIZANDO DEEP LEARNING

Trabajo final de investigación aplicada sometido a la consideración de la Comisión del
Programa de Estudios de Posgrado en Tecnologías de Información y Comunicación
para la Gestión Organizacional para optar al grado y título de Maestría Profesional en
Tecnologías de la Información y Comunicación para la Gestión Organizacional

MAX ZÚÑIGA FALLAS

Ciudad Universitaria Rodrigo Facio, Costa Rica

2023

DEDICATORIA

A Nana.

AGRADECIMIENTOS

A Dios por permitirme finalizar.

A mí.

A mi esposa, por motivarme a realizarla.

A la profesora, la Mag. Sindy Porras Santamaría, por su guía.

A los profesores, la MBA. Vivian Murillo Méndez y el MBA. Carlos Alberto Vega Alvarado por su disposición como lectores.

A mis mascotas, los lomitos Pana, Lucky, Pandora, Ciruela y los michis Tita, Benito, Shere-Khan, Billo, Careys y Tocino.

A los colaboradores del Archivo Nacional que contribuyeron al desarrollo de esta investigación.

“Este trabajo final de investigación aplicada fue aceptado por la Comisión del Programa de Estudio de Posgrado en Tecnologías de Información y Comunicación para la Gestión Organizacional de la Universidad de Costa Rica, como requisito parcial para optar al grado y título en Maestría Profesional en Tecnologías de la Información y Comunicación para la Gestión Organizacional.”

M.Sc. Francisco Blanco Chavarría
**Representante de la Decana
Sistema de Estudios de Posgrado**

Mag. Sindy Porras Santamaría
Profesora Guía

MBA. Vivian Murillo Méndez
Lectora

MBA. Carlos Alberto Vega Alvarado
Lector

M.Sc. Yorleny Salas Araya
**Directora del Programa de Posgrado en Tecnologías de la Información y
Comunicación para la Gestión Organizacional**

Max Zúñiga Fallas
Sustentante

TABLA DE CONTENIDO

DEDICATORIA.....	ii
AGRADECIMIENTOS	iii
HOJA DE APROBACIÓN.....	iv
RESUMEN	viii
SUMMARY	ix
Índice de Cuadros	x
Índice de Tablas.....	xi
Índice de Figuras.....	xii
Índice de Apéndices.....	xiii
Índice de Anexos.....	xiv
Licencia de Publicación.....	xv
Capítulo I. Introducción	1
1.1 Planteamiento del problema.....	1
1.2 Justificación.....	3
1.3 Objetivos	5
1.3.1 Objetivo general	6
1.3.2 Objetivos específicos.....	6
1.4 Antecedentes	6
1.4.1 Antecedentes internacionales.....	7
1.4.2 Antecedentes nacionales.....	8
Capítulo II. Marco teórico	10
2.1 Preservación digital de documentos	10
2.1.1 Conceptualización	10
2.1.2 Alcance y objeto	11
2.1.3 Principios básicos de la preservación digital	11
2.2 Normativa Nacional de preservación digital de documentos de Costa Rica	12
2.2.1 Ley N°7202 del Sistema Nacional de Archivos.....	12
2.2.2 Ley N°8454 de Certificados, Firmas Digitales y Documentos Electrónicos	13
2.2.3 Decreto Ejecutivo N°24023-C, Reglamento a la Ley N°7202 del Sistema Nacional de Archivos	13

2.2.4	Norma técnica para la gestión de documentos electrónicos en el Sistema Nacional de Archivos	14
2.2.5	Norma Técnica Nacional – NTN-003: Digitalización de documentos textuales en soporte papel	14
2.2.6	Norma Técnica Nacional – NTN-004: Lineamientos para la digitalización de documentos con la finalidad de sustituir el soporte original	14
2.3	Gestión de la calidad	15
2.3.1	Conceptualización	15
2.3.2	Principios	16
2.3.3	Análisis de los procesos	16
2.3.4	Cultura organizacional y sistemas de gestión de la calidad.....	17
2.3.5	Normas ISO relacionadas a gestión de la calidad en el proceso de digitalización de documentos.....	17
2.3.5.1	ISO/TR 13028-201. Información y documentación. Directrices para la implementación de la digitalización de documentos.	17
2.3.5.2	ISO 13008-2013. Norma para el Proceso de migración y conversión de documentos electrónicos.	18
2.4	Automatización Robótica de Proceso	18
2.5	Deep Learning.....	18
2.5.1	Redes Neuronales.....	19
2.5.1.1	Redes neuronales prealimentadas	20
2.5.1.2	Redes neuronales convolucionales	20
Capítulo III.	Marco Metodológico.....	21
3.1	Diseño de la investigación.....	21
3.2	Sujetos de información.....	21
3.3	Técnicas e instrumentos de investigación.....	22
3.3.1	Entrevista.....	22
3.3.2	Observación	23
3.4	Análisis de los datos	23
3.5	Cronograma	26
Capítulo IV.	Análisis de los resultados	27

4.1	Presentación y análisis de los resultados	28
4.1.1	Sistematización de la información brindada por las personas funcionarias del Área de Reprografía Documental del Archivo Nacional.....	28
4.1.2	Sistematización de la información obtenida mediante la observación al control de calidad de las imágenes digitalizadas	34
4.2	Análisis FODA.....	35
Capítulo V. Propuesta de mejora		40
5.1	Software	40
5.1.1	Azure Machine Learning.....	40
5.1.2	Google Colab.....	41
5.1.3	AutoML Vision	41
5.1.4	Teachable Machine	42
5.2	Implementación.....	44
5.2.1	Entrenamiento del modelo.....	44
Capítulo VI: Conclusiones y recomendaciones		53
6.1	Conclusiones.....	53
6.2	Recomendaciones	55
Referencias		57
Apéndices		61
Anexos		66

RESUMEN

Con el propósito de adecuarse a los cambios digitales, propiamente en la preservación digital del patrimonio documental de la nación y ante la posibilidad de ofrecer nuevos servicios en línea a las personas usuarias, el Archivo Nacional de Costa Rica pasó de la microfilmación a la digitalización, modificando las labores realizadas en el Área de Reprografía, llamada anteriormente Microfilmación.

Con base en lo anterior, se estudia el proceso de digitalización, enfocándose en el control de calidad de las imágenes digitalizadas, para comprender cómo se lleva a cabo y, a partir de ahí, desarrollar el objetivo del trabajo, que conlleva elaborar una propuesta para la automatización del proceso de gestión y evaluación de la calidad de las digitalizaciones realizadas en la Dirección General del Archivo Nacional, empleando para esos fines Deep Learning.

La propuesta consiste en generar un modelo de aprendizaje supervisado, en la plataforma Teachable Machine de Google, para clasificar las imágenes digitales en correctas y defectuosas. Con dicha herramienta se obtuvo en las pruebas un 100% de efectividad, brindando así un parámetro como punto de partida para conocer la exactitud de ese proceso.

SUMMARY

In order to adapt to digital changes, specifically in the digital preservation of the nation's documentary heritage and the possibility of offering new online services to users, the National Archive of Costa Rica moved from microfilming to digitization, modifying the work carried out in the Reprographic Area, formerly called Microfilming.

In view of this, the digitization process is studied focusing on the quality control of the digitized images, to understand how it is carried out and from there develop the objective of the work, to develop a proposal for the automation of the management process and evaluation of the quality of the digitizations made in the General Directorate of the National Archives using Deep Learning.

The proposal consists of generating a supervised learning model in the Teachable Machine platform from Google, to classify digital images into correct and defective. With this tool, 100% effectiveness was obtained, offering a parameter as a starting point to know the accuracy of this process.

Índice de Cuadros

Cuadro 1. Cronograma de actividades del planteamiento metodológico.	26
Cuadro 2. Análisis FODA del proceso control de calidad de las imágenes digitalizadas.	36
Cuadro 3. Matriz FODA del proceso control de calidad de las imágenes digitalizadas.	37

Índice de Tablas

Tabla 1. Matriz análisis de datos	24
Tabla 2. Guía para el análisis de datos.....	25
Tabla 3. Datos personales de las personas.	28
Tabla 4. Categoría de análisis N° 1: Gestión de la calidad.	29
Tabla 5. Categoría de análisis N° 2: Evaluación de la calidad.	33
Tabla 6. Comparación de herramientas de software.....	43

Índice de Figuras

Figura 1. Interfaz de Teachable Machine	45
Figura 2. Ingreso de imágenes para entrenar el modelo.....	46
Figura 3. Prueba en la identificación de imágenes con errores.....	47
Figura 4. Precisión del entrenamiento.....	48
Figura 5. Margen de pérdida por época.	49
Figura 6. Precisión por clase.....	50
Figura 7. Matriz de confusión.	50
Figura 8. Exportación del modelo generado en Teachable Machine.	51

Índice de Apéndices

Apéndice A. Guía de entrevista semiestructurada para las personas funcionarias del Área de Reprografía Documental del Archivo Nacional.	61
Apéndice B. Guía de observación al control de calidad de las imágenes digitalizadas.	64

Índice de Anexos

Anexo 1. Instrumento para el control de calidad de imágenes en Microsoft Excel.....	66
Anexo 2. Imagen defectuosa, imagen con mano.	67
Anexo 3. Imagen defectuosa, imagen cortada.	68
Anexo 4. Imagen defectuosa, imagen movida.	69
Anexo 5. Certificación de revisión filológica.	70



UNIVERSIDAD DE
COSTA RICA

SEP Sistema de
Estudios de Posgrado

Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.

Yo, Max Zúñiga Fallas, con cédula de identidad 114090369, en mi condición de autor del TFG titulado Automatización del proceso de gestión y evaluación de la calidad de las digitalizaciones realizadas en el Departamento de Conservación de la Dirección General del Archivo Nacional utilizando Deep Learning

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. SI ☒ NO * ☐

*En caso de la negativa favor indicar el tiempo de restricción: ____ año (s).

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

INFORMACIÓN DEL ESTUDIANTE:

Nombre Completo: Max Zúñiga Fallas

Número de Carné: A77234 Número de cédula: 114090369

Correo Electrónico: max.zuniga@ucr.ac.cr

Fecha: _____ . Número de teléfono: _____

Nombre del Director (a) de Tesis o Tutor (a): Mag. Sindy Porras Santamaría

FIRMA ESTUDIANTE

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

Capítulo I. Introducción

La presente investigación propone el empleo de una herramienta tecnológica, basada en Deep Learning para la gestión de calidad de las imágenes digitalizadas en el Archivo Nacional de Costa Rica. Dentro del primer capítulo se explica el planteamiento del problema de forma que permite conocer la situación actual, así como la justificación, que respalda el desarrollo del proyecto; además de los objetivos (general y específicos) que reflejan las metas por lograr y, finalmente, los antecedentes, en los que se plasman trabajos similares.

1.1 Planteamiento del problema

La automatización permite definir y planificar los diferentes procesos que se llevan a cabo dentro de las organizaciones, mediante la sistematización de sus actividades cotidianas, de modo que se logran simplificar tareas con la normalización y materialización de los procesos (López, 2015). En el caso de la Dirección General del Archivo Nacional de Costa Rica (DGAN), si bien existe un manual de procedimientos para la digitalización de los documentos, no hay un modelo establecido para el control de calidad de las imágenes.

En esa misma línea de pensamiento, la estandarización de procesos tiene una gran incidencia en las organizaciones, pues son actividades que permiten llevar a cabo una operación constante y definida, facilitando la producción de bienes o servicios con normalización y costos más bajos (Cuartas, 2012). En el caso específico de este estudio, la forma en que se lleva a cabo actualmente este proceso, no ofrece un estándar de uniformidad sobre la calidad del producto, lo que resulta en inconsistencias en la versión final de las imágenes ofrecidas al público, pudiendo incluso alterarse el contenido de la información contenida en dichas digitalizaciones.

Además de esto, hay que ser claros en que al ser un proceso manual implica una mayor cantidad de tiempo para llevarlo a cabo, contrario a un proceso automatizado (Deloitte,

2016). Este aspecto por supuesto incide de forma directa en una mayor inversión de tiempo por parte de los funcionarios, lo que visto desde el punto de vista de la organización, se traduce en mayores costos operativos y disminución considerable en la productividad.

Por otra parte, es claro que las organizaciones, independientemente de su razón de ser, deben manejar una visión integral de todos sus procesos, de manera que puedan tener una coordinada distribución y ejecución de estos (Ministerio de Fomento, 2005). En particular el DGAN, tiene como misión la conservación, facilitación y divulgación del patrimonio documental de la nación a todos los habitantes de Costa Rica (Archivo Nacional, 2021c).

Cabe apuntar que a finales de los años 70 e inicios de los 80, el DGAN crea el Departamento de Conservación (DCONS), junto con las áreas de Restauración, Encuadernación y Microfilmación. En esta última se realizaba la microfilmación del patrimonio documental, labor se efectuó hasta el año de 2015, cuando se cambió a la digitalización y modificó sus funciones, pasando a llamarse Área de Reprografía Documental (ARD). Con este cambio buscaba no solo adaptarse a los avances tecnológicos y asegurar la permanencia del patrimonio documental a través de la preservación digital, sino que se abría la posibilidad de ofrecer nuevos servicios en línea a toda la ciudadanía (Archivo Nacional, 2021b)

En cuanto a la digitalización de documentos, el DCONS mantiene un programa de reproducción de documentos a través del escaneo y la fotografía digital, el cual se orienta a partir del Plan de Reproducción de Documentos 2021-2025. Debido a la antigüedad de los documentos, – el original más antiguo data de 1539 y la copia de 1517 – así como por medidas de conservación, estas labores se ejecutan de modo manual, incluyendo el proceso de control de calidad.

Se puede mencionar que, si bien desde el año 2015 el DGAN inició la transición de la microfilmación a la digitalización, esta última tarea se ha visto fortalecida a partir de la

estrategia de Transformación Digital hacia la Costa Rica del Bicentenario 4.0 que busca la consolidación del Gobierno Digital.

En cuanto a la calidad dentro del proceso de digitalización, es necesario citar que los procesos de control de calidad tienen como objetivo no solo mejorar la captura, el procesamiento y la gestión de los archivos de imagen, sino que además deben apoyarse en una herramienta de caracterización, la cual facilita comprenderlos mejor y optimizar el proceso como tal (Pereira-Uzal, 2018).

En ese sentido, es importante ampliar que la calidad se encuentra definida por la claridad en cómo se ve la imagen digitalizada, así como por verificar que no se presenten imperfecciones que puedan dificultar la visualización de la información contenida en la imagen. Es decir, se busca asegurar la calidad visual del contenido de los archivos generados a través de la digitalización, de modo que sean copia fiel e íntegra de los originales en físico, garantizando su autenticidad, en este caso, mediante un flujo automatizado de este proceso.

A partir de lo mencionado anteriormente, esta propuesta se orienta hacia la automatización del proceso (de la actividad) de gestión y evaluación de la calidad de las digitalizaciones realizadas, el análisis del proceso actual y aprovechamiento de los recursos, de forma que facilite comprender cómo la automatización de la gestión y evaluación de la calidad de las digitalizaciones utilizando Deep Learning, favorece la mejora continua del proceso.

1.2 Justificación

A través de la innovación tecnológica, las personas pueden interactuar con múltiples dispositivos, lo que ha llevado a incorporar distintas herramientas tecnológicas en el campo laboral, provocando que los procesos sean automatizados, incluso en ocasiones genera un ahorro de tiempo en las actividades de las organizaciones, dando como resultado una alta eficiencia (López, 2015).

En el caso del DGAN, el proceso de digitalización de documentos antiguos e históricos se realiza de forma manual, incluido el control de calidad. El procedimiento para la inspección de calidad se lleva a cabo revisando las imágenes una a una. Cuando se presentan documentos de 25 folios o menos (por una sola cara) y en documentos más grandes, se efectúa a través de un muestreo cada 5 imágenes. Esta actividad se fundamenta en acuerdos tomados entre la dirección de la institución, la jefatura del DCONS y las personas funcionarias del ARD.

Concretamente para este proceso se requiere de uno a dos funcionarios que se encarguen de ejecutar el control de calidad manual de esas imágenes, lo cual implica el uso de más recursos de los necesarios. Por ejemplo, en cuanto al tiempo, para un documento de menos de veinticinco folios, pueden necesitarse de 30 a 60 minutos, como lo anteriormente descrito, para en caso de detectar algún error, descartar dicha digitalización e indicar su corrección. Es necesario acotar que esta labor no se hace una vez que se digitaliza cada documento, sino que se ejecuta de forma masiva, es decir, se digitaliza una cantidad de documentos y posteriormente se lleva a cabo el control de calidad a ese segmento.

Hay que agregar que uno de los beneficios obtenidos al implementar un sistema automatizado del proceso de calidad, es que contribuye, no solo a la mejora cuantitativa de determinadas características como mayor nitidez o menor ruido, sino que además faculta una mejora global del proceso, que se ve reflejado en una sólida automatización de tareas o sistemas para el control de errores, entre otras actividades (Pereira-Uzal, 2018).

De acuerdo con el autor, el control de calidad automatizado brinda un valor agregado a este tipo de procesos, disminuyendo considerablemente los errores que podrían venir en detrimento de la visualización y conservación de la información. Adicionalmente, se brinda homogeneidad al producto final, todo esto frente a un proceso manual en el que inciden aspectos como el azar, de la mano de la subjetividad de quien revisa, condiciones

que actualmente no asegura un modelo de calidad estandarizada a la totalidad de las imágenes.

Es necesario mencionar que el ARD se encontraba conformada por cuatro personas funcionarias, de los cuales dos se acogieron a su jubilación en diciembre del 2021 y sus plazas no se han podido recuperar, debido a las medidas tomadas por el gobierno para reducir el gasto público entre el 2021 y 2025. Esta situación genera una carencia de recurso humano e involucra invertir una mayor cantidad de tiempo para cumplir con las metas anuales de digitalización establecidas en el Plan Operativo Institucional.

A raíz de lo anterior, se pretende con esta investigación desarrollar una propuesta para la automatización del proceso de gestión y evaluación de la calidad de las digitalizaciones hechas en el DGAN, mediante la utilización de Deep Learning, que contribuya en la modernización y agilización de la gestión global del proceso, así como asegurar la integridad de las imágenes disponibles.

Con la finalidad de identificar las mejores prácticas, ofrecerle a la ciudadanía el acceso a los documentos sin la necesidad de trasladarse hasta la institución, garantizar la integridad y autenticidad de los documentos y apegarse a la normativa vigente en esta materia, resulta primordial automatizar el proceso de control de calidad de las imágenes digitalizadas de los documentos custodiados por el DGAN.

1.3 Objetivos

A continuación, se presentan el objetivo general y los específicos de la presente investigación:

1.3.1 Objetivo general

- 1.3.1.1 Elaborar una propuesta para la automatización del proceso de gestión y evaluación de la calidad de las digitalizaciones realizadas en el Departamento de Conservación de la Dirección General del Archivo Nacional utilizando Deep Learning.

1.3.2 Objetivos específicos

- 1.3.2.1 Identificar las fortalezas, oportunidades, debilidades y amenazas del proceso de gestión y evaluación de la calidad de las digitalizaciones que se lleva a cabo en la Dirección General del Archivo Nacional.
- 1.3.2.2 Diseñar un plan para la automatización de la gestión y evaluación de la calidad de las digitalizaciones realizadas en la Dirección General del Archivo Nacional.

1.4 Antecedentes

Dentro del proceso de investigación, la revisión documental es fundamental, ya que permite conocer trabajos similares, facilitando la claridad y un panorama más amplio, así como la obtención de información, métodos y técnicas que se pueden relacionar al trabajo que se desarrolla.

Para el presente estudio se consultaron sitios web, repositorios y bases de datos de las universidades públicas, tales como el Sistema de Bibliotecas, Documentación e Información (SIBDI) y el repositorio Kerwá, ambos de la Universidad de Costa Rica. También del Centro de Información, Documentación y Recursos Bibliográficos de la Universidad Estatal a Distancia (CIDREB). De igual modo, se consultaron bases de datos y repositorios a nivel internacional, como el caso de la Universidad Católica del Perú, entre otros.

1.4.1 Antecedentes internacionales

El investigador Romero (2009), en su trabajo de grado realiza una propuesta para la automatización de procesos de verificación y despacho en una panificadora. El estudio utilizó diferentes métodos, como el diagrama de pescado espina o el estudio de métodos para identificar los principales problemas de los procesos, en los que se destacan tiempos muertos por la doble verificación de forma manual hecha por los operarios. A partir de esto, el autor plantea que la automatización del proceso de distribución y verificación, provoca una mejora en el proceso general de la empresa debido a la disminución del tiempo y errores en el proceso.

Por su lado, Barbachan et al. (2015), indagaron si las Empresas del Sector de Servicios Públicos del Gobierno Central en el Departamento de Lima con un Sistema de Gestión de Calidad (SGC), poseían un mayor nivel de calidad comparado con las que no lo tienen. Para esto llevaron a cabo una investigación no experimental cuantitativa, utilizando un muestreo probabilístico y la aplicación de encuestas basadas en el instrumento de la Administración de la Calidad Total (TMQ) de Benzaquen (2013). Los resultados indicaron que las empresas que no cuentan con un SGC no tienen la forma de medir sus resultados o lo hacen de forma subjetiva.

En el caso de Lombardi (2016), se pretendía identificar qué mecanismos de un sistema de calidad pueden ser la base para establecer un sistema de innovación, destacando la mejora continua, ya que funciona como herramienta para minimizar el temor al fracaso y disminuir los errores. Entre los resultados más importantes, se señalan que la empresa no cuenta con una ruta hacia la innovación; sin embargo, se podría desarrollar a partir de la gestión de personas. Asimismo, se destaca la necesidad de hacerlo bien a la primera, pues al ser una empresa mediana, los recursos son limitados.

Los autores Robledano-Arillo et al. (2016), a partir del empleo del algoritmo de aprendizaje automático C4.5, utilizado en el control automatizado de calidad de imágenes, de obras de tipo gráfico, digitalizadas, encontraron que se pueden obtener

resultados eficaces superiores al 85%. Tomando en cuenta que el porcentaje de error en la valoración de experto es un 10,87% y un 20%, establecieron que el grado de acierto, de este algoritmo, a través de la automatización del proceso de control de calidad, es relativamente alto.

Cabe citar a Güendián (2019), quien empleó el Deep Learning en el proceso de la digitalización de documentos históricos, pero su trabajo se enfocó en la automatización del proceso de la transcripción de estos documentos, con el uso de la herramienta Ocropus. Ese proceso se basó en alimentar o entrenar la herramienta a partir de glosas para que esta realizará la transcripción de los documentos una vez digitalizados. Entre los resultados más importantes, menciona el autor que la red pre-entrenada con 160 líneas y posteriormente ajustada, obtiene buenos resultados con muchos menos ciclos de trabajo, para un porcentaje de error del 1.34%.

1.4.2 Antecedentes nacionales

Entre estos antecedentes se encuentra Rodríguez (2019), con una propuesta de un plan de utilización de metodologías ágiles como SCRUM, Programación Extrema (XP) o KANBAN, para apoyar los procesos de calidad en una empresa privada de desarrollo de software, que se apega a los métodos tradicionales para realizar sus procesos. Entre los resultados destaca que, si bien con la metodología tradicional se obtiene la calidad esperada, con las metodologías ágiles se puede obtener el mismo efecto, pero se pueden mejorar aspectos como el tiempo o el recurso humano, haciendo el proceso más eficiente.

Por su lado, Duarte (2020) desarrolla una aplicación web responsiva para la automatización del proceso de gestión de expedientes de información en el Área Rectora de Puntarenas del Ministerio de Salud. Los resultados indican que la automatización de la gestión de expedientes permite realizar nuevas características que no se consideraban antes y recolectar los errores que aparezcan para depurar los parámetros de calidad y favorecer la mejora continua.

Vásquez (2020) plantea la utilización de una herramienta de aprendizaje automatizado (Machine Learning) para la predicción del riesgo de dengue en Costa Rica. Entre los resultados señala que, a pesar de las variaciones entre las regiones del país, el empleo de la herramienta ofrece una oportunidad de mejora en el manejo de programas y la toma de decisiones de forma temprana en la lucha contra el dengue.

Capítulo II. Marco teórico

En este capítulo se detallan los postulados y aspectos teóricos que dan sustento a esta investigación, abarcando temas como la preservación digital, la gestión de calidad, Deep Learning y el marco normativo sobre la digitalización de documentos que rige en Costa Rica.

2.1 Preservación digital de documentos

La Conservación de documentos como disciplina, se ha apoyado en las TIC, buscando los medios, no solo para ponerlos a disposición de las personas, sino que facilita desarrollar nuevos medios para conservar los documentos originales y evitar su deterioro ante factores como la inapropiada manipulación.

2.1.1 Conceptualización

La digitalización es entendida como la conversión de un objeto o proceso análogo a un formato digital (Delgado, 2016). En el caso de un libro, documento o imagen, es pasarlos de su soporte en papel a uno digital para facilitar su transporte. No obstante, la digitalización para preservar procura sacar de circulación el documento físico original para conservarlo, sin pretender sustituirlo (Archivo Nacional, 2021d), para lo cual es necesario elaborar y seguir un plan que abarque diferentes aspectos técnicos para garantizar la adecuada manipulación del original y la autenticidad de la reproducción.

La preservación digital se define como el conjunto de políticas, estrategias y medidas específicas que tienen como fin mantener la integridad de los documentos, así como garantizar su inalterabilidad, autenticidad, además de asegurar la permanencia y el acceso a estos (Archivo General de la Nación de Colombia, 2018).

Se debe mencionar que en el proceso de digitalización se elaboran copias maestras y derivadas. Las primeras corresponden a documentos que se ejecutan con las mayores especificaciones y se mantienen como un documento independiente e inviolable, siendo a partir de estas que se obtienen las copias derivadas (UNE, 2013).

2.1.2 Alcance y objeto

Según la UNESCO (2009), el alcance de la preservación digital no solo incluye elementos culturales, de educación o científicos, abarca además recursos administrativos, técnicos, legales o médicos, que se hayan creado inicialmente en un formato digital o bien, que se hayan convertido a este formato a partir de un material analógico ya existente. Con respecto a Costa Rica, los documentos custodiados por la DGAN cuentan con valor científico-cultural. Y se debe aclarar que este carácter se lo da la Comisión Nacional de Selección y Eliminación de Documentos (CNSD), órgano asesor del Archivo Nacional, cuya función es establecer las normas relacionadas a la selección y eliminación de documentos, de acuerdo con su valor científico – cultural (Archivo Nacional, 2021a).

2.1.3 Principios básicos de la preservación digital

La preservación digital no se limita simplemente a un proceso de escaneo o de fotografiar documentos para tener una copia digital y eliminar el original, pues cumple una función de continuidad entre el pasado, el presente y el futuro (Umaña y Marín, 2021). Esto basándose en una serie de principios que no pretenden sustituir un soporte por otro, al contrario, busca asegurar la conservación del original analógico y generar reproducciones fieles que se puedan facilitar a los usuarios. Es decir, permite la adopción del sistema digital (Delgado, 2016).

Los principios que se mencionan a continuación, pretenden garantizar la autenticidad e integridad del documento digitalizado y su preservación en el tiempo.

- Integridad: Busca asegurar que el contenido del documento analógico original no se ha modificado; asimismo, que los formatos empleados realicen una reproducción de calidad que deje visualizar apropiadamente el documento.
- Equivalencia: Consiste en la migración de los formatos, sin que esto altere el documento, de manera que se pueda hacer frente a la obsolescencia tecnológica.
- Economía: Asegurar que los procesos de digitalización se puedan mantener en el tiempo y sean viables, para garantizar la sostenibilidad técnica y económica.
- Actualidad: Consiste en evolucionar con los avances tecnológicos, así como con los formatos que surgen, para evitar los desfases.
- Normalización: Elaborar y emplear lineamientos y estándares para mantener un control normalizado de los documentos digitalizados (Archivo General de la Nación de Colombia, 2018).

2.2 Normativa Nacional de preservación digital de documentos de Costa Rica

Dentro de la normativa aplicable en Costa Rica existen, desde leyes hasta normas técnicas, con la cuales se rigen las instituciones en materia de digitalización de documentos y su manejo apropiado.

2.2.1 Ley N° 7202 del Sistema Nacional de Archivos

En el artículo 23 inciso “b” de la Ley N° 7202, se establece como una función esencial del Archivo Nacional:

Reunir, conservar, clasificar, ordenar, describir, seleccionar, administrar y facilitar los documentos textuales, gráficos, audiovisuales, y legibles por máquina, pertenecientes a la Nación, que constituyan el patrimonio documental nacional, así como la documentación privada y particular que le fuere entregada para su custodia (24 de octubre de 1990).

2.2.2 Ley N° 8454 de Certificados, Firmas Digitales y Documentos Electrónicos

En el artículo 3 de esta Ley se determina que los documentos electrónicos tienen jurídicamente valor equivalente a los documentos que se otorguen, residan o transmitan por medios físicos. Se debe aclarar que los “documentos digitalizados pueden considerarse documentos electrónicos simples..., si contienen una validación jurídica” (Franco y Pérez, 2014, p. 9).

Posteriormente, en el artículo 6, se indica que cuando legalmente se requiera que un documento sea conservado para futura referencia, se podrá optar por hacerlo en soporte electrónico, siempre y cuando se apliquen las medidas de seguridad necesarias para garantizar su inalterabilidad, se posibilite su acceso o consulta posterior y se preserve, además, la información relativa a su origen y otras características básicas.

Adicionalmente, establece que la transición o migración a soporte electrónico, cuando se trate de registros, archivos o respaldos que por Ley deban ser conservados, deberá contar, previamente, con la autorización de la autoridad competente.

2.2.3 Decreto Ejecutivo N° 24023-C, Reglamento a la Ley N° 7202 del Sistema Nacional de Archivos

En los artículos 99 y 100 del Reglamento a la Ley N° 7202, se indica que el Archivo Nacional facilitará los servicios reprográficos necesarios para que los usuarios puedan tener acceso al acervo documental, sufragar los costos de estos y la obligación de citar la fuente cuando se usare para fines de investigación y publicación.

2.2.4 Norma técnica para la gestión de documentos electrónicos en el Sistema Nacional de Archivos

Su finalidad es sistematizar la gestión de documentos electrónicos a partir de la normalización de la producción de documentos en las diferentes unidades de una institución, tomando en cuenta el marco jurídico vigente (Junta Administrativa del Archivo Nacional, 2018).

2.2.5 Norma Técnica Nacional – NTN-003: Digitalización de documentos textuales en soporte papel

Esta norma busca regular la digitalización de documentos textuales que fueron producidos en soporte de papel a través de lineamientos y requerimientos que se deberán aplicar, tomando en cuenta las siguientes variables: 1. Forma de producción. 2. Valor y plazos de los documentos 3. Objetivo de la digitalización (Junta Administrativa del Archivo Nacional, 2021).

2.2.6 Norma Técnica Nacional – NTN-004: Lineamientos para la digitalización de documentos con la finalidad de sustituir el soporte original

El objetivo de esta norma técnica es promover el correcto desarrollo, implementación, mantenimiento y mejora continua de los procesos de digitalización de documentos, para garantizar la integridad, confiabilidad, autenticidad, valor legal y acceso a estos, de manera que se pueda asegurar la preservación de las características fundamentales y originales de los documentos al efectuar un cambio en su soporte, de manera que conserven su validez administrativa y legal en el registro de los diferentes actos de la administración pública (Junta Administrativa del Archivo Nacional, 2021).

2.3 Gestión de la calidad

La heterogeneidad de términos y las distintas experiencias en organizaciones, hacen que sea una tarea difícil poder determinar con exactitud qué es la gestión de la calidad, generando confusión en jerarcas y mandos medios acerca de cómo establecer parámetros exactos para gestionar la calidad de los productos o servicios que ofrecen.

2.3.1 Conceptualización

Se puede definir como un conjunto de variables necesarias para la puesta en práctica de técnicas y principios necesarios para la mejora de la calidad (Camisón et al., 2007). Es decir, en este sentido, implica el diseño de un proceso que enlace distintos componentes, así como los aspectos relacionados al concepto de calidad de la organización, ya sea que ofrezcan productos o servicios.

Por otra parte, Vesga (2013), citando ICONTEC (2009), explica que un sistema de gestión de la calidad es el área de organización enfocada en el logro de los resultados, a partir de los objetivos de calidad establecidos, que buscan cumplir con los requerimientos esperados para satisfacer las necesidades de las partes interesadas.

En cuanto a la digitalización de documentos, se deben considerar una serie de aspectos tales como los define el Ministerio de Educación, Cultura y Deporte de España (2015), al detallarlos de esta forma:

- Corrección de las imágenes obtenidas en cuanto a calidad de la imagen o en cuanto a criterios técnicos:
 - Resolución adecuada al tipo documental.
 - Color adecuado al tipo documental.
 - Formato adecuado al tipo documental.
 - Alineación correcta de la imagen.
- Corrección de las imágenes obtenidas en cuanto a fidelidad con el original:
 - Digitalización de todas las páginas del documento.

- Digitalización sin incluir información que no aparece en el original.
- Visualización y legibilidad de la imagen.

2.3.2 Principios

Según la norma NTP-ISO 9004:2001, existen ocho principios sobre los cuales se desarrollan los procesos de gestión de la calidad en cualquier institución. En cuenta la *organización enfocada al cliente* donde se trata de conocer las necesidades del público meta, satisfacer e incluso exceder estas expectativas. El *liderazgo* que trata de que la cabeza de las organizaciones propone la línea a seguir, involucrando al personal en el logro de los objetivos. La *participación del personal* y compromiso en todos los niveles operacionales. El *enfoque a los procesos* para la obtención de resultados. Un *sistema enfocado hacia la gestión* para la interrelación de todos los procesos. El *mejoramiento continuo* como un objetivo permanente. La *toma de decisiones* basada en hechos fiables como los datos y la información. Finalizando, con la *relación mutuamente benéfica con proveedores* para aumentar la capacidad de ambos para crear valor entre sí (García, Quispe & Ráez, 2003).

2.3.3 Análisis de los procesos

Previamente al inicio de cualquier proceso de mejora es fundamental un análisis de los procesos existentes, es decir, los métodos empleados actualmente para realizar las tareas (Rincón, 2009). Lo anterior, por cuanto estas tareas son las que inciden directamente sobre los resultados al final de la cadena de trabajo y al examinarlas se pueden determinar las vías y oportunidades de mejora.

2.3.4 Cultura organizacional y sistemas de gestión de la calidad

La cultura organizacional juega un papel clave en la implementación de cualquier sistema de gestión, sin importar el enfoque o propósito que tenga, debido a que dependiendo de esta se facilita o dificulta el proceso de cambio y mejora.

Para Vesga (2013) citando a Fernández (2002), la cultura organizacional se concibe como un sistema de valores que se ha creado dentro de un entorno específico el cual determina la ideología y el manejo general dentro de la organización. Este mismo autor desglosa los componentes de la cultura organizacional en formales o expresivos (artefactos materiales: tecnología, distribución del espacio, forma de vestir; formas lingüísticas: lenguaje, mitos, leyendas e historias; personajes: fundadores, héroes y líderes; ritos, rituales: sociales, comunicación, trabajo, administrativos y reconocimiento) y nucleares (de contenido y significado: valores, creencias y presunciones básicas).

2.3.5 Normas ISO relacionadas a gestión de la calidad en el proceso de digitalización de documentos

Las normas ISO ofrecen un estándar que ayuda a las organizaciones a normalizar y simplificar sus procesos, de modo que las tareas sean más efectivas y con un costo menor, además de establecer los responsables y partes involucradas en cada proceso.

2.3.5.1 ISO/TR 13028-201. Información y documentación. Directrices para la implementación de la digitalización de documentos.

El propósito de esta norma es especificar las directrices, tanto para la creación, como para el mantenimiento de los documentos digitalizados a partir de los originales, en soportes tradicionales como el papel. De igual forma, establece los requerimientos del proceso de digitalización con el propósito de asegurar la veracidad, fiabilidad y validez jurídica de los documentos digitalizados. También indica las estrategias para la creación

de documentos digitalizados, su conservación a largo plazo y la gestión de los documentos originales tras su digitalización (UNE, 2011).

2.3.5.2 ISO 13008-2013. Norma para el Proceso de migración y conversión de documentos electrónicos.

Esta norma establece los procesos a seguir para la confección o diseño de un plan de conversión o migración de soportes, asegurando la mayor cantidad de garantías posibles. Se hace necesario entender que la conversión es el proceso de cambio de formato o soporte, en ocasiones útil para el mantenimiento o usabilidad de los documentos. Por otra parte, migración es el cambio de sistema, aplicación o medio de almacenamiento, proceso que se debe llevar a cabo porque la vida útil de los documentos es mayor que la de los sistemas o formatos que los soportan (UNE, 2013).

2.4 Automatización Robótica de Proceso

La Automatización Robótica de Proceso (RPA por sus siglas en inglés), se puede definir como un método para automatizar procesos a partir de reglas específicas. No se trata de un robot físico, sino de un software que aprende de un usuario y lo asiste con tareas sencillas. Estas reglas son lógicas pre-construidas para entregar resultados (Deloitte, 2017).

Cabe señalar que el RPA y la Inteligencia Artificial (IA) son dos temas distintos. El primero emplea tecnologías de automatización para imitar tareas humanas y la segunda es la respuesta que tiene un sistema informático para trabajar ante la ambigüedad, similar a la inteligencia humana (KIO Networks, 2021).

2.5 Deep Learning

La finalidad de la inteligencia artificial es ofrecer soluciones mediante algoritmos y técnicas con los cuales se pueden ejecutar tareas de forma automática que las personas

ejecutan manualmente, en este caso, el control de calidad de imágenes digitalizadas. Es importante mencionar que los fabricantes emplean cada vez más recursos de Deep Learning para automatizar diferentes procesos en sus organizaciones, principalmente en inspecciones de fabricación, con lo cual pretenden reducir el margen de error y acelerar los tiempos de revisión (Cognex Corporation, 2022).

El Deep Learning o aprendizaje profundo es parte del Machine Learning o aprendizaje automático, que forma parte, a su vez, de la inteligencia artificial (IA) (López, 2017). El término profundo o Deep, hace referencia a la cantidad de capas empleadas en el modelo, las cuales aprenden automáticamente a medida que el modelo es entrenado con los datos (López, 2017).

Para entender el funcionamiento de las capas, se requiere tener claro que cualquier herramienta de Machine Learning ejecuta una asignación de entradas a salidas. Por ejemplo, a una imagen, la etiqueta "gato", por medio de la observación de un gran número de ejemplos de este tipo. El Deep Learning realiza este mismo proceso de mapeo a través de una red neuronal artificial compuesta de un número grande de *capas* colocadas en forma de jerarquía (López, 2017).

En ese orden la red, en su capa inicial aprende y luego envía esta información a la siguiente capa. La siguiente capa toma esta información tal cual esta y la combina, generando algo más complejo, para pasarlo a la tercera capa. Este proceso continúa en cada capa, donde se irá formando algo más complejo a partir de la entrada de la capa anterior y así la red irá *aprendiendo* por medio de la exposición a los datos de ejemplo (López, 2017).

2.5.1 Redes Neuronales

Las redes neuronales son un modelo computacional basado en un gran conjunto de unidades neuronales simples (neuronas artificiales). Esto se da en forma aproximadamente análoga al comportamiento observado en los axones de las neuronas

en los cerebros biológicos (López, 2017). En otras palabras, las redes neuronales intentan simular el funcionamiento de las neuronas de un cerebro humano.

2.5.1.1 Redes neuronales prealimentadas

Fueron las primeras que se desarrollaron y son las más sencillas. En estas redes la información se mueve en una sola dirección: hacia adelante. Los mayores ejemplos de esta arquitectura son el perceptrón y el perceptrón multicapa. Son empleadas en problemas de clasificación simples (López, 2017).

2.5.1.2 Redes neuronales convolucionales

Son muy similares a las redes neuronales ordinarias como el perceptron multicapa. Estas neuronas están compuestas de pesos y sesgos que pueden aprender. Cada neurona recibe algunas entradas, efectúa un producto escalar y luego aplica una función de activación. Lo que diferencia a las redes neuronales convolucionales es que suponen explícitamente que las entradas son imágenes, lo que facilita codificar ciertas propiedades en la arquitectura; permitiendo ganar en eficiencia y reducir la cantidad de parámetros en la red (López, 2017).

Capítulo III. Marco Metodológico

En este capítulo se aborda la metodología empleada y detallando la forma en que se llevó a cabo la investigación. Para su ejecución se utilizaron instrumentos cualitativos, tal como la entrevista que ayuda a la identificación y obtención de datos más precisos, de forma que enriquece el análisis de los datos y ofrece un amplio panorama al momento de abordar la propuesta del trabajo.

3.1 Diseño de la investigación

Este trabajo se desarrolla bajo el enfoque cualitativo, pues permite profundizar en el entorno para describir la problemática (Hernández et al., 2014). El tipo de investigación es descriptivo, ya que se busca describir los eventos y posteriormente organizar, tabular y presentar los datos (Glass y Hopkins, 1984). Con esto se procura conocer el contexto estudiado y brindarle un acompañamiento adecuado para la satisfacción de sus necesidades. En este caso se pretende conocer la situación desde la perspectiva de las propias personas funcionarias del ARD.

3.2 Sujetos de información

Los sujetos de información corresponden a los qué o quiénes de la investigación (Hernández et al., 2014). Para efectos de este trabajo son las dos personas funcionarias que laboran, al momento de elaborar este proyecto en el ARD, quienes llevan a cabo el proceso de digitalización del patrimonio documental y control de calidad de las imágenes. Tienen una categoría de Técnico de Servicio Civil 2 G. de E. (Grupo de especialidad), especialidad microfilmación. Sus funciones son las siguientes:

- Digitalizar documentos recibidos de los diferentes departamentos del DGAN.
- Digitalizar de documentos del patrimonio documental del fondo Gobernación.
- Digitalizar documentos históricos, a solicitud de las personas usuarias.

- Digitalizar los documentos de transferencias y otra previa coordinación con los departamentos correspondientes.
- Verificar que las imágenes digitalizadas se puedan visualizar correctamente bajo los parámetros de calidad establecidos.

3.3 Técnicas e instrumentos de investigación

Son los instrumentos de recolección de datos, como la entrevista semiestructurada y la observación sobre la actividad del control de calidad de las imágenes. Además de los documentos analizados, como libros, trabajos finales de graduación, investigaciones de campo, bases de datos en línea y publicaciones periódicas, durante la revisión de literatura.

3.3.1 Entrevista

La entrevista, para Denzin y Lincoln (2005) es “una conversación, es el arte de realizar preguntas y escuchar respuestas” (p. 643). Para esta investigación se ha optado por la entrevista semiestructurada, pues a pesar de llevar una guía, le propone al investigador agregar preguntas adicionales para completar ideas o conceptos (Hernández et al., 2014). Esta corresponde al primer objetivo del trabajo y se encuentra dirigida a las dos personas funcionarias del ARD, la cual se aplicó de forma presencial.

La entrevista citada consta de una primera parte referida a datos personales; la segunda a la variable denominada gestión de la calidad, que consta de 12 preguntas abiertas y la tercera parte, a la variable llamada evaluación de la calidad, con 3 preguntas abiertas. Estas consultas responden al primer objetivo específico, que es parte de los instrumentos a utilizar para la posterior elaboración del FODA (Apéndice A).

3.3.2 Observación

La observación le proporciona al investigador el conocimiento de forma directa del fenómeno a estudiar. Es decir, es una aproximación sensible al fenómeno, delimitada por el problema, el plan de recolección y el análisis e interpretación de los datos (Barrantes, 2015). Para el desarrollo de este trabajo se utilizó la observación participante limitada, donde la relación entre observador y observado es indiferente e impersonal, siendo el primero un simple espectador (Riba Campos, 2013). Esto con el fin de conocer y describir los pasos a seguir por las personas funcionarias del ARD en el control de calidad de las imágenes digitalizadas.

Esta observación se llevó a cabo con una de las personas funcionarias del ARD, mientras hacía dicha actividad, de manera que fuera posible analizar cómo se ejecutaba según la cantidad de folios, el tiempo empleado, si son copias maestras o derivadas, los pasos a seguir en caso de encontrar errores en las imágenes, quién solventa los errores, si se generan informes o estadísticas de errores encontrados y solucionados, así como los filtros existentes

El instrumento consiste en una tabla con 15 ítems (Apéndice B), que permitirán obtener información sobre el control de calidad. Se encuentra dividida en dos partes, la primera responde a la variable, gestión de la calidad, que consta de seis ítems. La segunda parte a la variable evaluación de la calidad, con nueve, los cuales responden al primer objetivo específico, del cual surgen los instrumentos a utilizar para la posterior elaboración del FODA.

3.4 Análisis de los datos

El análisis de la información obtenida de forma cualitativa se presenta por medio de tablas, en las que se expondrán los datos derivados de las entrevistas y la observación. Además, se empleará la matriz FODA. Las tablas llevan la estructura que se denota en la siguiente Tabla 1.

Tabla 1. Matriz análisis de datos

Pregunta	Entrevistado 1	Entrevistado 2
-----------------	-----------------------	-----------------------

Nota: Elaboración propia, 2022.

- La información se estructuró a través de categorías a partir de las unidades que son relevantes para el interés en la investigación. Asimismo, para la presentación de una forma organizada de la información, se empleó la herramienta Microsoft Word 2016.
- Se elaboró una matriz de análisis FODA (Araya, 2019), que se basa en los factores internos – fortalezas y debilidades – y factores externos –oportunidades y amenazas – (Romero y Cordero, 2020).
- En la Tabla 2 se indica la guía para realizar el análisis de los datos, empezando por el objetivo específico al cual responde y los instrumentos y actividad correspondiente.
- En el Cuadro 1 se explica el cronograma y cómo se fueron desarrollando cada una las fases de esta investigación.

Tabla 2. Guía para el análisis de datos.

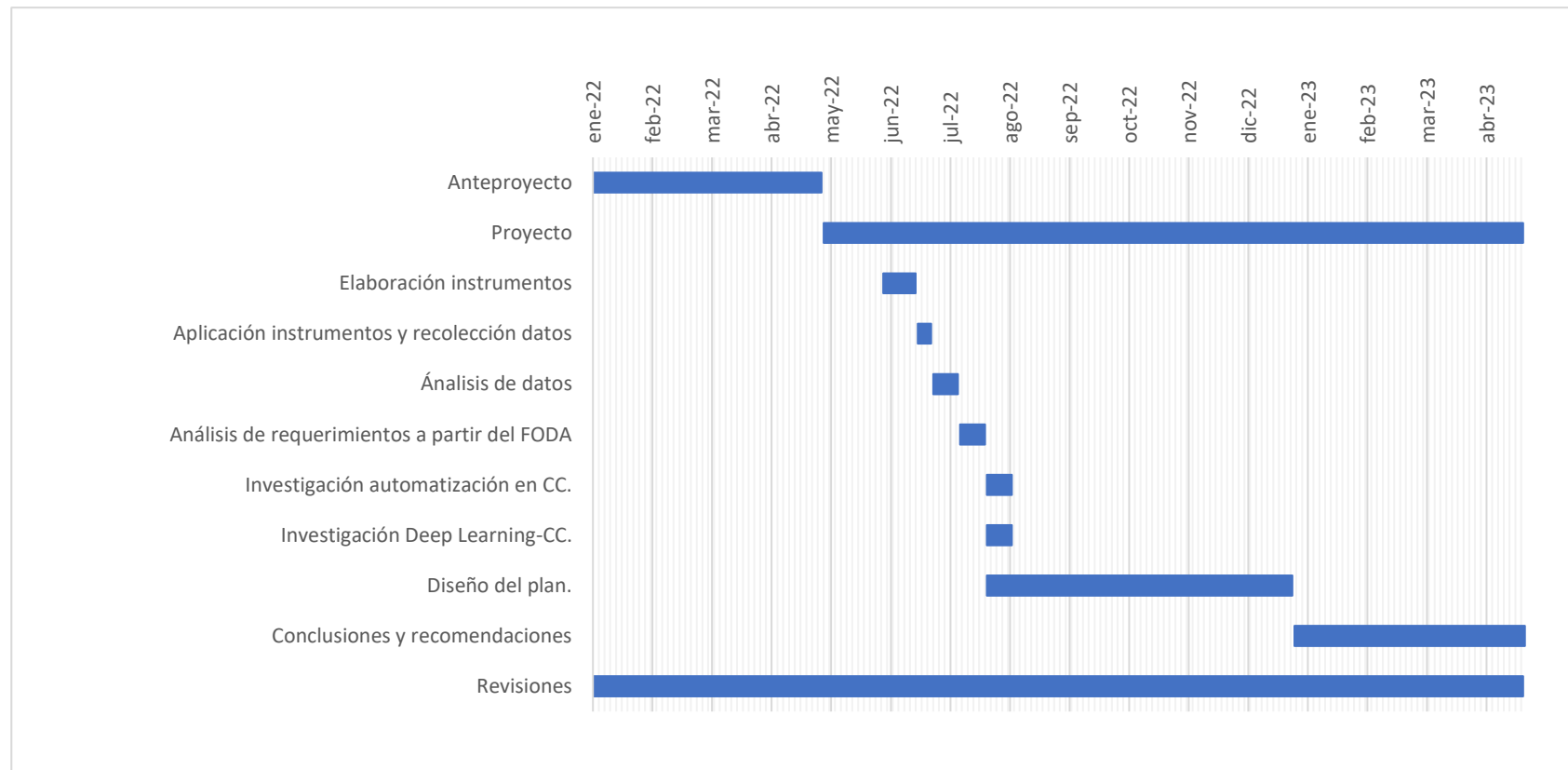
Objetivo	Variables	Técnica	Instrumento	Actividades
Identificar las fortalezas, oportunidades, debilidades y amenazas del proceso de gestión y evaluación de la calidad de las digitalizaciones.	1. Gestión de la calidad. 2. Evaluación de la calidad.	1. Entrevista 2. Observación	1. Entrevista estructurada. 2. Observación participante limitada.	1. Elaboración de instrumentos. 2. Aplicación de instrumentos. 3. Análisis de datos.
Diseñar un plan para la automatización de la gestión y evaluación de la calidad de las digitalizaciones realizadas en la Dirección General del Archivo Nacional.	1. Requerimientos del DCONS. 2. Necesidades de hardware y software. 3. Requerimientos de programación de Deep Learning.	1. Revisión de documentos	1. Manual de procedimientos Digitalización de documentos de conservación permanente.	1. Investigación sobre automatización en el control de calidad. 2. Investigación sobre Deep Learning en el control de calidad. 3. Diseño del plan.

Nota: Elaboración propia, 2022.

3.5 Cronograma

En el Cuadro 1 se puede observar la distribución de la investigación a lo largo del tiempo.

Cuadro 1. Cronograma de actividades del planteamiento metodológico.



Nota: Elaboración propia, 2022.

Capítulo IV. Análisis de los resultados

El presente capítulo se compone de dos apartados y el primero corresponde a la presentación y análisis de los resultados, en el que se plasman los datos obtenidos mediante la aplicación de los instrumentos de recolección denominados “Guía de entrevista semiestructurada para las personas funcionarias del Área de Reprografía Documental del Archivo Nacional” (Apéndice A) y “Guía de observación al control de calidad de las imágenes digitalizadas” (Apéndice B), con su respectivo análisis.

Es importante indicar que la recolección de datos llevada a cabo a través de las entrevistas, se realizó de manera anónima y confidencial, sin mantener el registro de nombres, correos electrónicos o algún otro dato en las respuestas, de modo que se pudiera revelar la identidad de los participantes.

Por otro lado, en el segundo apartado se desarrolla el análisis FODA a partir de los resultados, para conocer los aspectos más importantes y tener un panorama más amplio sobre la actividad en cuestión.

La recolección y presentación de los resultados se orientó desde el objetivo del diagnóstico planteado en la investigación:

1. Identificar las fortalezas, oportunidades, debilidades y amenazas del proceso de gestión y evaluación de la calidad de las digitalizaciones.

Con base en el objetivo mencionado se extrajeron las categorías (variables) para diagnosticar la población de estudio de acuerdo con el enfoque de investigación cualitativo:

1. Categoría N° 1. Gestión de la calidad.
2. Categoría N° 2. Evaluación de la calidad.

4.1 Presentación y análisis de los resultados

A continuación, se muestra el análisis de la aplicación de los instrumentos, que se inicia con la entrevista y finaliza con los datos obtenidos a través de la observación.

4.1.1 Sistematización de la información brindada por las personas funcionarias del Área de Reprografía Documental del Archivo Nacional

La entrevista (Apéndice A) tiene como sujetos de información a las dos personas funcionarias del ARD. La primera parte del instrumento busca conocer información personal (grado académico, años de laborar en la institución y edad) como se observa en la Tabla 3, por lo que los datos a continuación son dados desde el punto de vista de las personas sujetos de información, a las cuales se les asignó un código para identificarlas, E.1 y E.2.

Tabla 3. Datos personales de las personas.

Preguntas	E. 1	E. 2
Grado académico	Bachiller en Tecnología de la Imagen (UTN)	Diplomado en Fotografía (UTN)
Años de laborar en la institución	12 años de experiencia	7 años de experiencia
Edad	47 años	41 años

Nota: Elaboración propia a partir de los datos obtenidos de la entrevista, 2022.

Con respecto a la primera categoría de análisis, “Gestión de la calidad”, en el Tabla 4 se pueden observar las respuestas a las preguntas que la conforman por parte de los entrevistados.

Tabla 4. Categoría de análisis N° 1: Gestión de la calidad.

Preguntas	E. 1	E. 2
Pregunta 4 ¿Conoce los procedimientos establecidos para la digitalización de documentos en el Archivo Nacional? Descríbalos	Sí. Se inicia con el cotejo de firmas, se revisa que el programa esté con los parámetros establecidos y se procede a digitalizar.	Sí. Es para saber cómo se deben digitalizar los documentos, según su tipo.
Pregunta 5 ¿Cuál es el protocolo de digitalización de documentos? Descríbalos.	Crear una carpeta con la firma del documento, redireccionar el guardado de las imágenes en la carpeta creada, verificar los parámetros establecidos, seleccionar el formato indicado, se hace un control de calidad, en un documento en Excel se anotan las fallas y se hace un acta de verificación cuando se corrigen.	Hacer una carpeta con la firma del documento, establecer el guardado en la carpeta creada, revisar los parámetros para digitalizar, se hace un control de calidad, en un Excel se anotan los errores y se hace un acta de verificación cuando se corrigen.
Pregunta 6 ¿Cuáles son las actividades para llevar a cabo la digitalización de documentos históricos? Enumérelas y descríbalas y ¿Cuánto se tarda digitalizando un documento?	Es donde se indican los parámetros para digitalizar un documento realizado en procesadores de texto, máquinas de escribir o un documento manuscrito. Depende del grosor, pero se han hecho 400 imágenes por día.	Generalmente es revisar los números de cajas y documentos y empezar a digitalizarlos. Relativo al tamaño del documento.
Pregunta 7 ¿Cuáles son los recursos tecnológicos con los que cuentan para el proceso de digitalización de documentos históricos?	Aparte del equipo de cómputo, un escáner plantario y otro plano. Los programas y licencias los manejan el DTI. La preocupación es la obsolescencia del equipo o de los programas, pues en ocasiones el equipo se pone lento.	La computadora de escritorio, un escáner plano y los programas, que los maneja DTI.
Pregunta 8 ¿En qué formatos se digitalizan los documentos históricos?	Generalmente en TIFF, pero actualmente es en JPEG.	JPEG actualmente, antes eran en TIFF.
Pregunta 9 ¿Existe algún manual o procedimiento sobre el control de calidad de las imágenes digitalizadas?	No. Es una actividad más dentro de la digitalización.	Actualmente no, es algo más verbal. Se hizo una investigación por aparte y se determinó que se podía hacer cada 10 imágenes.
Pregunta 10 ¿Cuántas personas realizan el control de calidad de las imágenes digitalizadas? Y ¿Cuáles son sus puestos?	Dos por el momento, los que trabajamos en esta área. Se realiza una vez a la semana y se hace cruzado, o sea, yo reviso el trabajo de la otra persona y viceversa.	Dos. Aunque se está integrando el fotógrafo también.

Preguntas	E. 1	E. 2
Pregunta 11 ¿Cuáles son las actividades que se desarrollan durante el control de calidad de las imágenes?	Depende del grosor del documento se hace folio por folio, o por muestreo. La foliación o numeración no se puede controlar, porque algunos documentos la traen, otros no, o tienen diferentes numeraciones.	Revisar las imágenes digitalizadas, y dependiendo del grueso del documento se hace uno por uno, o por muestreo. Actualmente se está llevando un control en Excel, para que la persona que digitalizó pueda guiarse para corregir los errores y luego se realiza un acta de verificación.
Pregunta 12 Aparte de usted, ¿existe otro filtro de control de calidad de las imágenes? ¿Cuál es?	Se supone que en el Departamento Archivo Histórico realizan otro control, pero es sobre los metadatos y descripción del documento, para verificar que correspondan las imágenes y la información.	No.
Pregunta 13 ¿Llevan estadísticas del tiempo empleado en el proceso completo de la digitalización de documentos?	No. Solo informes de labores donde se pone a modo general que se realizó la corrección de errores de la digitalización.	No.
Pregunta 14 Del tiempo empleado en la digitalización, ¿Cuánto corresponde al control de calidad?	Un día a la semana.	Un día completo por semana.
Pregunta 15 ¿Quién corrige los errores encontrados?	La persona que digitalizó	Quien digitalizó

Nota: Elaboración propia a partir de las respuestas textuales obtenidos de la entrevista, 2022.

Si bien en los cuadros se presentan respuestas más puntuales, al ser una entrevista, se pudo recolectar una cantidad mayor de información, ya que muchas preguntas desencadenaron otras más, enriqueciendo la recolección de datos.

De las preguntas sobre la primera categoría de análisis, relacionadas al procedimiento de digitalización, se destaca que ambos entrevistados lo conocen y aplican, que concuerda con los principios de básicos de la preservación digital, la normalización, la cual, según el Archivo General de la Nación de Colombia (2018), consiste en la elaboración y empleo de lineamientos y estándares para mantener un control normalizado de los documentos digitalizados.

Por otra parte, se debe resaltar que solo uno de los funcionarios tiene claros los conceptos de qué es una copia maestra y una derivada. Por ese motivo es necesario aclarar que una copia maestra corresponde a los documentos que se digitalizan con las mayores especificaciones y se mantienen como un documento independiente e inviolable. Por su parte las copias derivadas, son las que se obtienen a partir de las maestras (UNE, 2013).

En cuanto a los recursos tecnológicos, se debe mencionar que los dos coincidieron en que los equipos de cómputo son óptimos para las tareas por el momento. Sin embargo, en ocasiones se ponen lentos, sobre todo al abrir carpetas con una gran cantidad de imágenes y “duran para cárgalas”, aspecto que a ambos preocupa a futuro, debido a la obsolescencia, según señalaron. Al respecto el Archivo General de la Nación de Colombia (2018), propone que se requiere tener en cuenta fundamentos tales como la equivalencia y la actualidad. El primero consiste en la migración de los formatos, sin que esto altere el documento y el segundo en evolucionar con los avances tecnológicos y los formatos que surgen de manera que se pueda hacer frente a la obsolescencia tecnológica.

También mencionaron que el Departamento Tecnologías de Información (DTI) solo se encarga de darles soporte con relación a los aspectos técnicos. Y más bien el tema de

la preservación digital y control de calidad, se maneja en el DCONS. En cuanto a este punto se debe indicar que la participación del personal y compromiso en todos los niveles operacionales es uno de los principios de la gestión de la calidad incluidos en la norma NTP-ISO 9004:2001 (García, Quispe & Ráez, 2003).

Una pregunta que surgió a partir del tema del recurso tecnológico, fue acerca de si tenían conocimiento sobre Inteligencia Artificial (IA), a lo que respondieron que era poco lo que sabían, principalmente sobre los robots e información encontrada en redes sociales. El poco conocimiento que los funcionarios tienen sobre IA, puede ser parte del panorama nacional en el que hay pocas empresas que cuentan con proyectos de investigación en este campo. Al respecto comenta Pensis (2018), “En Costa Rica todavía los avances son limitados, especialmente porque la mayoría de empresas privadas no tiene el músculo económico para financiar investigaciones en el campo” (párr. 12). Se debe destacar que uno de los entrevistados comentó que en algún momento se planteó la idea de trabajar con sistema para el control de calidad, pero quedó solo como una idea.

En cuanto a las actividades del control de calidad, es relevante el hecho que ambos coinciden en que el control de calidad es una actividad más y no existe un documento o procedimiento formal como tal, en el que se indique su ejecución, sino que funciona con acuerdos verbales sobre las actividades a realizar. En relación con este aspecto, Camisón et al. (2007) y la norma ISO 13008-2013 (UNE, 2013), mencionan que es necesario diseñar un proceso que facilite vincular distintos elementos y que a su vez se puedan relacionar al concepto de calidad, es decir, establecer formalmente el proceso para el control de calidad.

Asimismo, al consultarles a los entrevistados sobre los filtros existentes, ambos respondieron que solo son ellos dos quienes llevan a cabo dicha actividad, con un control cruzado. No obstante, uno de los entrevistados indicó que se estaba involucrando al fotógrafo en dicha labor para solventar la falta de personal y establecer otro filtro.

Para la segunda categoría de análisis, “Evaluación de la calidad”, en la Tabla 5 se aprecian las respuestas brindadas por los entrevistados.

Tabla 5. Categoría de análisis N° 2: Evaluación de la calidad.

Preguntas	E. 1	E. 2
Pregunta 16 ¿Conoce alguna norma sobre la digitalización de documentos?	Sí. Las normas institucionales, además de otras que estoy viendo en un curso virtual con la Fonoteca de México.	Solo las normas de digitalización. De control de calidad no.
Pregunta 17 ¿Sabe qué es una copia maestra o derivada?	Sí. Se estuvieron realizado en TIFF. Actualmente se hacen JPEG, que es un formato para copia derivada, pero menos pesado para almacenarlo.	No.
Pregunta 18 ¿Cuál es el nivel de precisión o margen de error en el control de calidad que realizan?	Depende del grosor del documento, si son delgados es mínimo porque se hace folio a folio. Si es grueso el documento no se puede medir, pues al hacerlo por muestreo pueden escaparse errores en las imágenes.	No existe uno como tal. Cuando se hace por muestreo se pueden escapar errores.

Nota: Elaboración propia a partir de las respuestas textuales obtenidos de la entrevista, 2022.

Con respecto a la segunda categoría de análisis, las preguntas estaban orientadas a saber si conocían la normativa relacionada a la preservación digital y control de calidad. Ambos funcionarios indicaron que conocen las normas vigentes sobre preservación digital, pero no alguna específica del control de calidad de imágenes digitalizadas.

También se les consultó si llevaban estadísticas o tenían algún margen de precisión en cuanto a los errores detectados en el control de calidad y ambos indicaron que no se llevan estadísticas y no pueden establecer un margen, pues cuando el documento es de gran tamaño y se realiza por muestreo se pueden pasar por alto errores. Solamente en documentos pequeños (en cantidad de folios) se puede asegurar, ya que deben cotejar todas las imágenes e incluso pueden confrontarlas contra el original en físico.

4.1.2 Sistematización de la información obtenida mediante la observación al control de calidad de las imágenes digitalizadas

La información recopilada mediante la guía de observación facilitó comprender con mayor detalle la forma en cómo se lleva a cabo el control de calidad de las imágenes digitalizadas, complementado la información brindada por los entrevistados, para tener una idea más clara de dicha actividad. Los ítems evaluados durante la observación corresponden a las categorías (variables) del objetivo de diagnóstico: Categoría N° 1. Gestión de la calidad y Categoría N° 2. Evaluación de la calidad.

Con respecto a la categoría N° 1. Gestión de la calidad, se pudo observar que la persona funcionaria lleva a cabo el protocolo de digitalización, verificando si es una copia maestra (máster) o una copia derivada, porque en ocasiones la orden de trabajo corresponde a la solicitud de un usuario de una reproducción digital de un documento que no se encuentra establecido en el plan de trabajo de digitalización. Pero generalmente estas solicitudes son acerca de algún grupo de folios de un documento.

Cuando es una copia maestra, los parámetros establecidos son un formato TIFF 300 dpi 24 bits de profundidad, aunque actualmente están trabajando con formato JPEG 600 dpi 24 bits de profundidad sin compresión y se consideran copias derivadas, esto debido al peso de cada imagen, porque se está agotando el espacio.

El tiempo requerido para la digitalización fue de aproximadamente una hora para un documento de 12 folios (frente y vuelto cada uno) en un buen estado de conservación. En otras palabras, se digitalizaron 24 imágenes. Sin embargo, esto no es una métrica general, pues ellos trabajan por metas, o sea, se les asigna una cantidad de cajas para digitalizar en el año. Cada caja tiene un número variado de documentos, porque se acomodan de acuerdo con su tamaño.

Otro aspecto por considerar, en conjunto con la duración es el estado de conservación, pues si el documento se presenta muy deteriorado o débil, la manipulación debe ser más cuidadosa, lo que implica mayor tiempo de trabajo.

Para la Categoría N° 2. Evaluación de la calidad, se realizó el control de calidad, pero al ser un documento pequeño, el mismo funcionario realizó un breve control de calidad en el cual duró aproximadamente 18 minutos, ejemplificando de esa manera cómo se ejecuta esta actividad. Posteriormente se llevó a cabo un control de calidad a 92 imágenes, en lo que se invirtió una hora y veintidós minutos. Para esto se efectuó un muestreo de imágenes, es decir de 5 en 5, buscando errores como dobleces, imágenes desenfocadas u otras fallas que no permitieran apreciar el texto.

En cuanto a los errores, actualmente trabajan con un libro en Microsoft Excel (Anexo 1), en el que llevan un registro general, sin especificar cuáles son los errores encontrados, para que la persona funcionaria que hizo la digitalización se guíe y pueda subsanar las fallas encontradas. En este caso no se detectaron errores.

Sobre los informes, simplemente se levanta un acta de verificación cuando los errores han sido corregidos para dar fe que el documento es legible. El control de calidad no cuenta con un informe como tal y más bien se lleva como parte de las actividades semanales de cada funcionario.

4.2 Análisis FODA

La finalidad, tanto de las entrevistas como de la observación es efectuar un análisis FODA acerca de la gestión y evaluación de la calidad de las digitalizaciones. En el cuadro 2 se muestra la información sobre aspectos resaltados por los entrevistados y registrados en esta observación.

En ese sentido se puede mencionar que la implementación del análisis FODA determina la situación actual de la organización y, a su vez, contribuye a su diagnóstico. Este FODA se desarrolla a partir de las fortalezas, como los factores positivos cruciales con los que se cuenta. Las oportunidades, que son aspectos positivos externos que se puede aprovechar. Las debilidades, tales como factores negativos cruciales que se deben disminuir o eliminar. Y por último, las amenazas, que se refiere a los aspectos negativos externos que pueden dificultar el alcance de metas (Castro, 2013).

En el Cuadro 2 se observa el análisis FODA llevado a cabo durante esta investigación.

Cuadro 2. Análisis FODA del proceso control de calidad de las imágenes digitalizadas.

Fortalezas	Debilidades
1. Dominio de las tareas por parte de las personas funcionarias. 2. Disposición para aprender a trabajar con herramientas que agilicen el control de calidad. 3. Conocimiento sobre la normativa vigente. 4. Trabajo en equipo.	1. Poco conocimiento de temas sobre Inteligencia Artificial (incipiente). 2. Falta de un procedimiento formal sobre la actividad de control de calidad. 3. Poco personal (actualmente 2 personas funcionarias). 4. Falta de apoyo del área de TI en la elaboración de un sistema para el control de calidad.
Oportunidades	Amenazas
1. Automatizar el control de calidad para una mayor precisión. 2. Herramienta innovadora, gratuita y fácil de usar.	1. Medidas restrictivas por parte del gobierno (congelamiento de plazas). 2. Falencias al poner a disposición las imágenes por posibles errores no detectados. 3. Obsolescencia de recursos tecnológicos.

Nota: Elaboración propia a partir de los datos obtenidos en los instrumentos, 2022.

Mediante los datos recolectados por medio de los instrumentos y el análisis realizado a través del FODA, se obtuvo conocimiento de la situación actual acerca del control de calidad de las digitalizaciones y es con base en este, que se desarrolla la propuesta que busca agilizar esta actividad a través de su automatización con Deep Learning.

A continuación, se presenta el cruce de los factores internos (fortalezas y debilidades) con los factores externos (oportunidades y amenazas), de manera que se puedan identificar las buenas prácticas existentes y encontrar los puntos de mejora. Para mayor comprensión de la matriz se utilizó un código para identificar las estrategias, el cual está conformado por la letra inicial, según sea el factor y un número. Por ejemplo, Fortaleza 1 (F1), Debilidad 1 (D1), Oportunidad 1 (O1), Amenaza 1 (A1) y así sucesivamente.

Cuadro 3. Matriz FODA del proceso control de calidad de las imágenes digitalizadas.

Estrategias FO	Estrategias DO
F2 – O1 y O2: La disposición para aprender a trabajar por parte de las personas funcionarias con una herramienta innovadora, gratuita y fácil de usar que permita automatizar el control de calidad de las imágenes digitalizadas para obtener una mayor precisión, establecer procedimiento y generar mejoras a futuro en el producto.	D2 – O1: La oportunidad de automatizar esta actividad, generaría un procedimiento formal, creando un documento que puede servir de insumo para otras instituciones que pretendan llevar a cabo un proceso similar. D4 – O2: Falta de apoyo del área de TI en la elaboración de un sistema para el control de calidad y una herramienta innovadora, gratuita y fácil de usar.
Estrategias FA	Estrategias DA
F1 y F4 – A1: El dominio de las tareas y el trabajo en equipo por parte de las personas funcionarias del ARD, son factores mitigantes ante las medidas restrictivas por parte del gobierno.	D2 – A2: Ante la falta de un procedimiento formal sobre la actividad, puede llegar a presentarse posibles errores no detectados en las imágenes. D3 – A1: La presencia de solo 2 personas funcionarias y las medidas restrictivas del gobierno pueden repercutir en la prestación de servicios, principalmente en la digitalización de documentos históricos, a solicitud de las personas usuarias.

Nota: Elaboración propia a partir de los datos obtenidos en los instrumentos, 2022.

A partir de la matriz se pueden desarrollar las líneas estratégicas, las cuales se dividen en éxito, adaptación, reacción y defensa o supervivencia. Estas surgen al comparar las fortalezas, oportunidades, amenazas y debilidades analizadas.

Para la línea de éxito, se ha definido: *F2 – O1 y O2: Aprovechar el dominio de las tareas y la disposición para aprender y trabajar por parte de las personas funcionarias con una herramienta innovadora, gratuita y fácil de usar que facilite automatizar el control de*

calidad de las imágenes digitalizadas para obtener una mayor precisión, establecer procedimiento y generar mejoras a futuro en el producto. La unión de estos puntos facilitaría la implementación del sistema, así como obtener una retroalimentación en procura de la mejora continua.

Con respecto a la línea de adaptación se determinó: *D2 – O1: La oportunidad de automatizar esta actividad, generaría a su vez un procedimiento formal, creando un documento que puede servir de insumo para otras instituciones que pretendan llevar a cabo un proceso similar.* Es que se diseñaría un procedimiento formal para esta actividad, máxime que la DGAN es el ente rector del Sistema Nacional de Archivos (SNA).

Además, se puede acotar que: *D4 – O2: Ante la falta de apoyo del área de TI en la elaboración de un sistema para el control de calidad y la existencia de una herramienta innovadora, gratuita y fácil de usar.* Esto les permitirá a los funcionarios del ARD cierto grado de autonomía, es decir, no depender completamente del área de cómputo para la gestión y control de calidad de las imágenes digitalizadas.

En cuanto a la línea de reacción se puede mencionar el: *F1 y F4 – A1: El dominio de las tareas y el trabajo en equipo por parte de las personas funcionarias del ARD, son factores mitigantes ante las medidas restrictivas por parte del gobierno.* En este punto se debe resaltar el compromiso de las personas funcionarias, ya que, por las restricciones actuales en el sector público, de cierto modo, tienen una recarga en sus labores diarias.

En la línea de defensa se tiene que: *D2 – A2: Ante la falta de un procedimiento formal sobre la actividad, puede llegarse a presentar posibles errores no detectados en las imágenes.* Esto se plantea ya que actualmente no existen estadísticas, grado de precisión o margen de error, lo que se podría alcanzar con la implementación de un sistema automatizado.

Por último, el *D3 – A1: La presencia de solo 2 personas funcionarias y las medidas restrictivas del gobierno pueden repercutir en la prestación de servicios, principalmente en la digitalización de documentos históricos, a solicitud de las personas usuarias*. Lo descrito sucede al encontrarse solamente dos personas funcionarias realizando las labores de cuatro, por lo que en algún momento tendrán que priorizar tareas, dejando alguna en segundo plano.

Capítulo V. Propuesta de mejora

La finalidad del presente trabajo es proponer la ejecución del control de calidad de las imágenes digitalizadas a través del Deep Learning, mediante el entrenamiento de una red neuronal convolucional con el uso de una herramienta web que pueda distinguir de forma eficiente las imágenes que presentan errores en su visualización. La automatización de este proceso reduciría una tarea que, según la cantidad de imágenes, puede durar varias horas o no ser tan precisa al practicarle un muestreo.

Se expone el modelo empleado para el desarrollo del trabajo, acompañado de breves explicaciones acerca de su funcionamiento. Además, se muestran los recursos tecnológicos utilizados, como son las librerías y el entorno empleado, la necesidad por satisfacer y el procedimiento para llevar a cabo la implementación de una red neuronal convolucional, iniciando con la creación y tratamiento del conjunto de datos (dataset) para trabajar, el funcionamiento, entrenamiento y prueba del modelo. A continuación, se presentan los recursos tecnológicos.

5.1 Software

Seguidamente se detallan algunos de los modelos más citados durante el desarrollo de esta investigación.

5.1.1 Azure Machine Learning

Azure Machine Learning (Azure ML) es un servicio en la nube establecido a partir de la plataforma Microsoft Azure, es decir, trabaja como plataforma de servicio (PaaS por sus siglas en inglés), por lo que no requiere infraestructura de ningún tipo, lo cual es una ventaja. Se debe mencionar que es de paga.

El Azure ML facilita el acelerar y administrar el ciclo de vida de los proyectos de aprendizaje automático... Además, se puede usar en flujos de trabajo diarios:

entrenamiento e implementación de modelos y administración de MLOps, asimismo permite la creación modelos precisos rápidamente con el aprendizaje automático automatizado para modelos tabulares, de texto e imagen, usando ingeniería de características y barrido de hiperparámetros (Microsoft, 2022).

5.1.2 Google Colab

Google Colaboratory es un servicio de Google gratuito que permite la ejecución del código en la web, por lo que no se necesita de infraestructura de hardware, además de facilitar el trabajo colaborativo al momento de entrenar los modelos o de procesar la información. Este es un servicio en la nube basado en Jupyter Notebooks para difundir la educación y la investigación del aprendizaje automático. Proporciona un tiempo de ejecución totalmente configurado para el aprendizaje profundo y el acceso gratuito (Carneiro et al., 2018).

5.1.3 AutoML Vision

Forma parte de Google Cloud, el cual es una suite de infraestructuras y servicios que la Google utiliza a nivel interno y ahora está disponible para cualquier empresa, de tal forma que sea aplicable a multitud de procesos empresariales (Ordorica, 2020). Por su parte, AutoML Vision proporciona entrenar modelos de aprendizaje automático para clasificar las imágenes a partir de imágenes etiquetadas y evalúa su rendimiento, lo que permite registrar modelos entrenados para la entrega a través de la API de AutoML. Es un servicio de paga. (Google Cloud, 2022).

5.1.4 Teachable Machine

Es una herramienta de IA desarrollada por Google basada en la web y que facilita la creación modelos de aprendizaje automáticos de manera rápida, sencilla y accesible para todos. Es de fácil uso, pues solamente se debe preparar un ordenador para que reconozca las imágenes, sonidos y posturas sin escribir código de aprendizaje automático y, posteriormente, se exporta el modelo para emplearse en diferentes proyectos y en sitios o aplicaciones, entre otros. Entre sus ventajas se destacan que no es necesario que la persona cuente con conocimientos previos sobre aprendizaje automático. Esta herramienta facilita recopilar y agrupar lo que se desee que el ordenador aprenda; preparar el modelo y comprobar si hace la clasificación de forma correcta y exportar el modelado de un proyecto de forma gratuita y online (Teachable Machine, 2017).

En la Tabla 6 se realiza una comparación con los principales aspectos que se deben considerar al momento de elegir el software.

Tabla 6. Comparación de herramientas de software.

Software Criterio	Azure Machine Learning	Google Colab	AutoML Vision	Teachable Machine
Requerimientos	Usuario Office	Usuario Google	Usuario Google	No necesita.
Usabilidad	Depende de la experticia del usuario.	Sencillo de utilizar e intuitivo.	Sencillo de utilizar e intuitivo.	Es muy sencillo.
Conocimientos	Es necesario que el usuario cuente con experiencia en el entrenamiento de modelos de aprendizaje.	Necesita conocimientos mínimos.	Poca experiencia en la materia de entrenar modelos	No requiere conocimientos.
Interoperabilidad	Puede trabajar con marcos como PyTorch TensorFlow scikit-learn XGBoost LightGBM	Es compatible con Python y su ecosistema de herramientas de terceros	Es compatible con Tensorflow Lite, Core ML	Los modelos creados con esta herramienta son propios de TensorFlow.js, de modo que se pueden exportar y usar en cualquier lugar con JavaScript,
Licencia	De paga.	De paga.	De paga.	Libre.

Nota: Elaboración propia, 2022

Una vez estudiado cada software y confrontada la información de cada uno con los datos recolectados a través de las entrevistas a las personas funcionarias del ARD, se optó por escoger Teachable Machine, ya que para utilizar esta herramienta no se requiere conocimiento en programación o aprendizaje automático. Este aspecto puede favorecer la incorporación de esta tecnología dentro de las funciones del ARD, porque los entrevistados mencionaron en su momento que no tienen acercamiento con estos temas y que, de lo poco que han escuchado, se refiere a los robots propiamente.

Es importante mencionar que esta herramienta facilita programar un modelo propio de aprendizaje automático con solo hacer clic en un botón, sin los requerimientos de codificación. Asimismo, se debe agregar que trabaja con bibliotecas de código abierto, lo cual ayuda a realizar las modificaciones para su mejor progreso. El modelo desarrollado se puede exportar a diferentes recursos como sitios web, aplicaciones, máquinas físicas y otras plataformas.

Otro aspecto por considerar es que es de uso libre, por lo que no requiere de paga, con lo que se brinda un acercamiento más amigable para la institución, en especial debido a la coyuntura económica por la que atraviesa el país y las restricciones presupuestarias en el sector público.

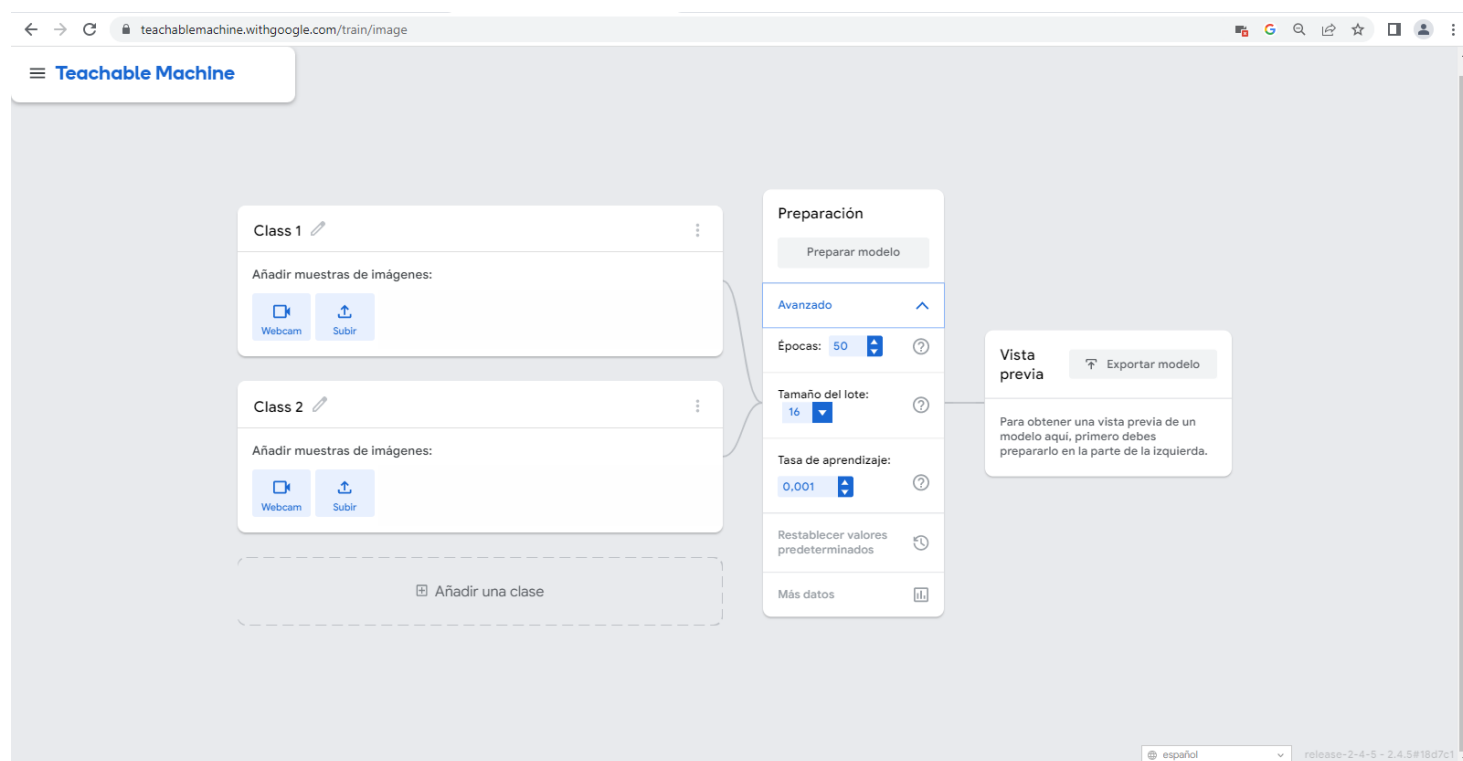
5.2 Implementación

A continuación, se explica el proceso para llevar a cabo la implementación de la herramienta elegida.

5.2.1 Entrenamiento del modelo

Para el entrenamiento del modelo, se utilizó Teachable Machine, el cual presenta dos paneles: con el primero se agregan los elementos que sirven como base para entrenar el modelo y el segundo una “avanzada” en la cual se pueden ajustar los parámetros. Este panel se encuentra en el módulo de capacitación y puede expandirse para visualizar las opciones que ajustan los parámetros de épocas, tamaño de lote y tasa de aprendizaje, tal como se observa en la Figura 1.

Figura 1. Interfaz de Teachable Machine

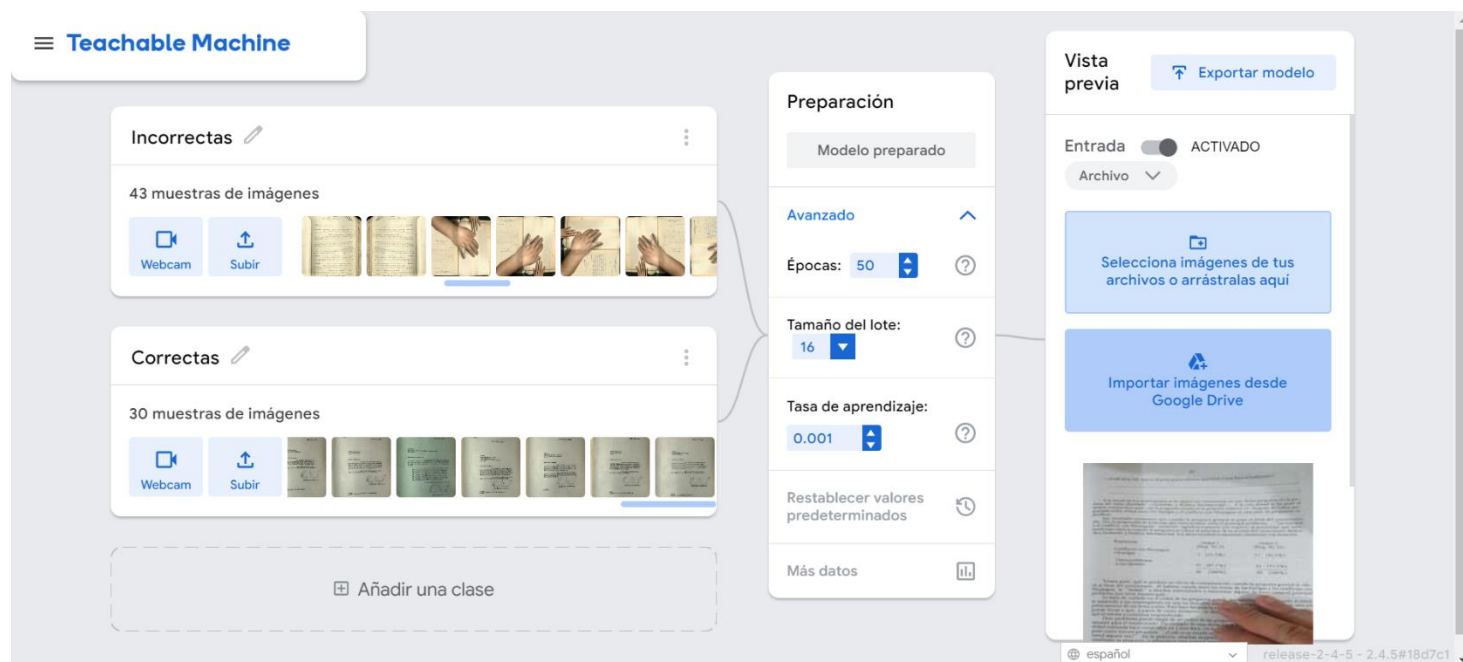


Nota: Elaboración propia, 2022.

Cabe destacar que la herramienta web para clasificación de imágenes tiene parámetros establecidos de forma predeterminada; sin embargo, estos pueden ser modificados según se necesite, como se observa en la Figura 1.

El banco de imágenes (dataset) de Teachable Machine se conforma de 50 épocas y cada una representa la revisión total de todos los datos recopilados. Para el proyecto se empleó un total de 43 imágenes para procesar, con errores detectados en las imágenes digitalizadas (Anexo 2, 3, 4).

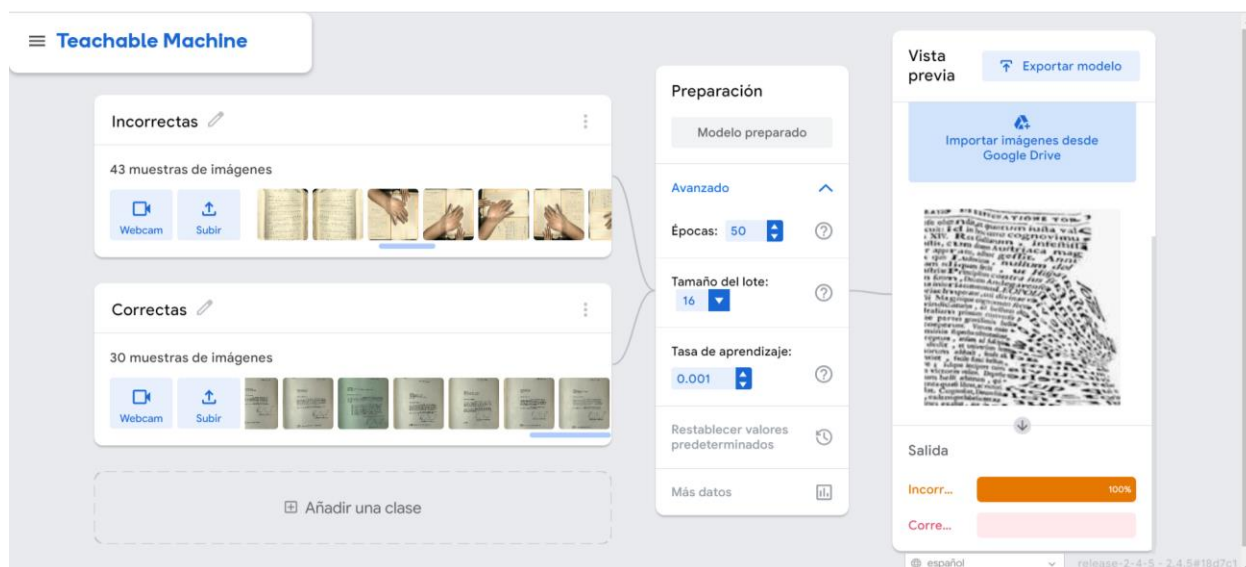
Figura 2. Ingreso de imágenes para entrenar el modelo.



Nota: Elaboración propia, 2022.

Una vez ingresadas las imágenes (Figura 2), se procede a entrenar el modelo con los datos establecidos y luego se efectúan las pruebas para verificar su funcionamiento, como se puede observar en Figura 3.

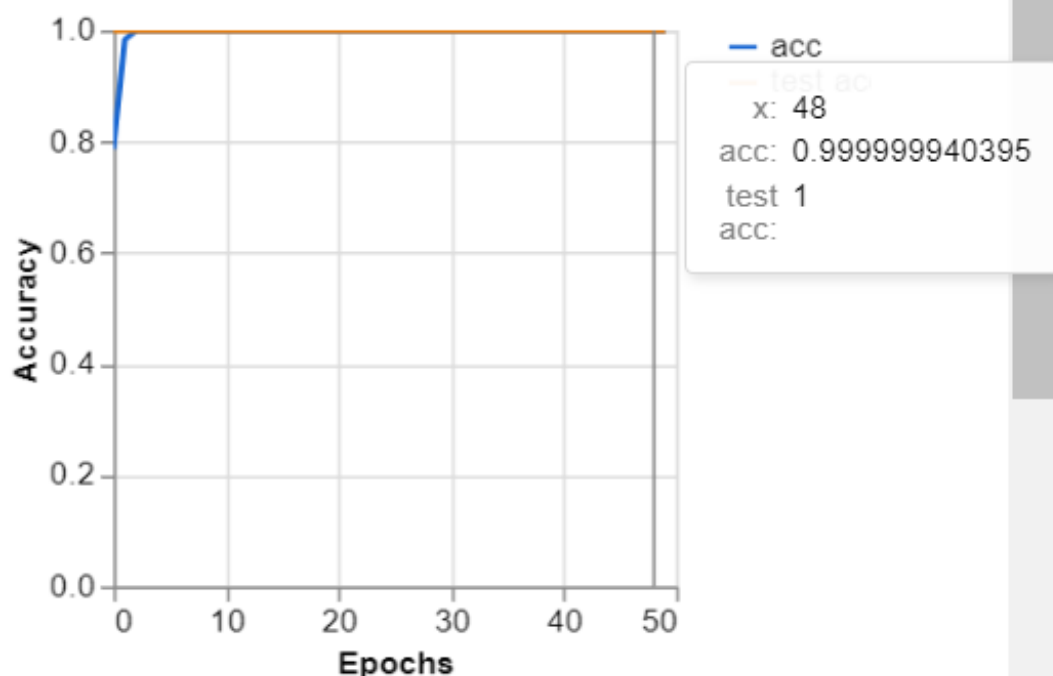
Figura 3. Prueba en la identificación de imágenes con errores.



Nota: Elaboración propia, 2022.

En la Figura 3 se observa el proceso de prueba al ingresar una imagen movida, para lo cual el modelo arroja un 100% relacionado a imágenes incorrectas, en este caso con errores. Es importante para aclarar que, si bien existen diferentes tipos errores como se puede observar en los Anexos 2, 3 y 4, para este trabajo se pretende solo trabajar en dos categorías, que son imágenes correctas e incorrectas.

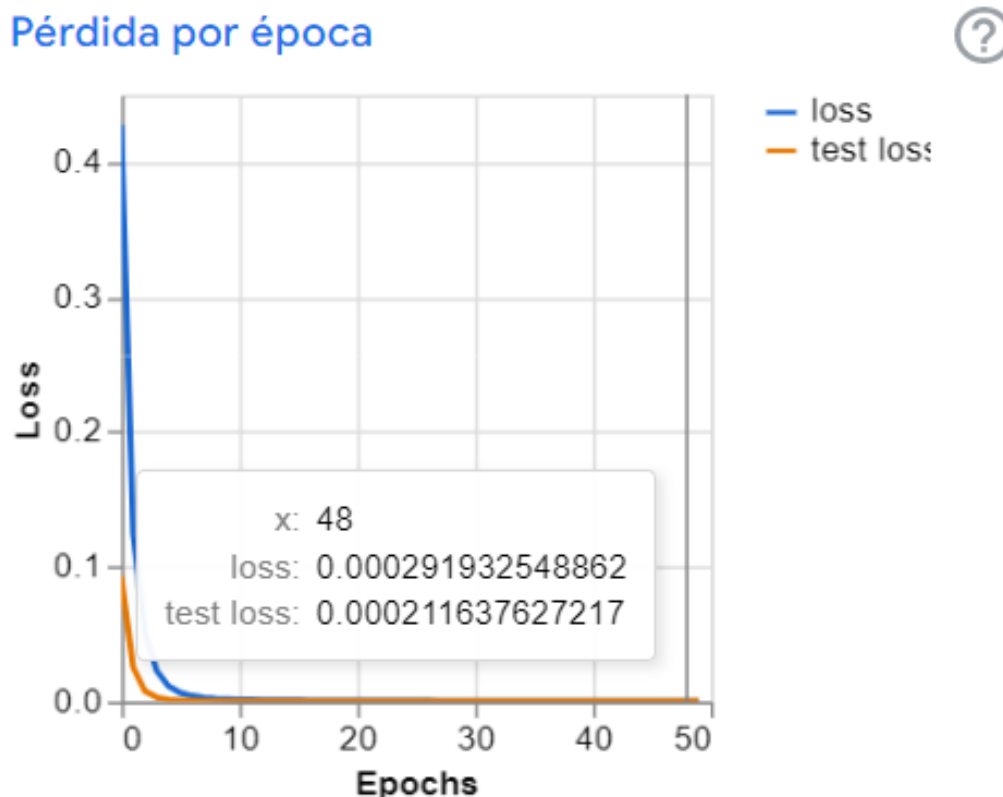
En cuanto a la evaluación, Teachable Machine ofrece un panel secundario que monitorea el funcionamiento (Figura 4), así como valora el aprendizaje del modelo y su exactitud. Se debe mencionar que esta herramienta utiliza el 85% de las muestras para enseñarle al modelo como clasificar y 15% restante se emplea para comprobar el rendimiento del modelo. Bajo este contexto, la Figura 4 presenta la precisión por época, siendo esta el porcentaje de clasificación que ha acertado el modelo en el proceso de entrenamiento, dado que la precisión es igual a 1, lo que se considera exitoso.

Figura 4. Precisión del entrenamiento**Precisión por época**

Nota: Elaboración propia, 2022.

Además de la precisión como verificador, esta herramienta presenta las pérdidas por época, con las que se calcula el nivel de aprendizaje del modelo, al momento de predecir la clasificación de las imágenes. Como se denota en la Figura 5, el margen de pérdida es 0.00029, que es un valor cercano a cero, que no es representativo, lo que indica que las predicciones del modelo son y serán correctas.

Figura 5. Margen de pérdida por época.



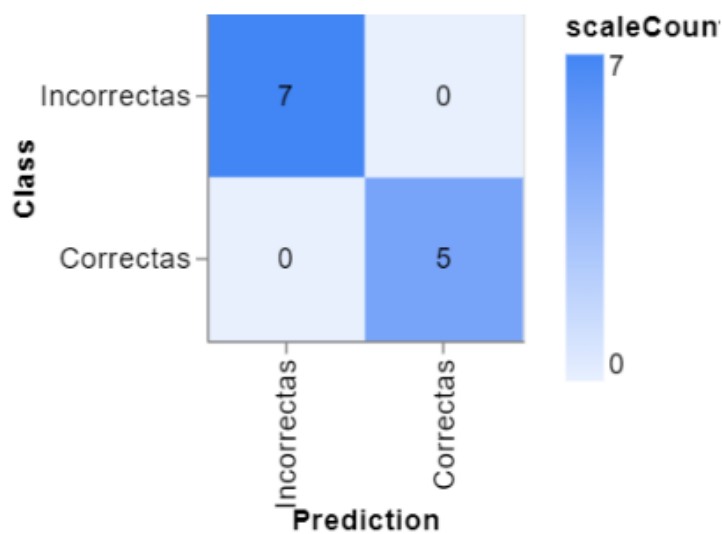
Nota: Elaboración propia, 2022.

Para corroborar los datos obtenidos de los indicadores de predicción y pérdida, se recurre a la matriz de confusión y a la precisión por clase, con lo que se logra visualizar el desempeño del entrenamiento que se emplea en el aprendizaje supervisado del modelo. Además, se logra observar que el modelo no presenta confusión con las clases, significando así que el modelo lleva a cabo la clasificación correctamente, como se puede observar en la Figura 6 y Figura 7.

Figura 6. Precisión por clase.**Precisión por clase**

CLASS	ACCURACY	# SAMPLES
Incorrectas	1.00	7
Correctas	1.00	5

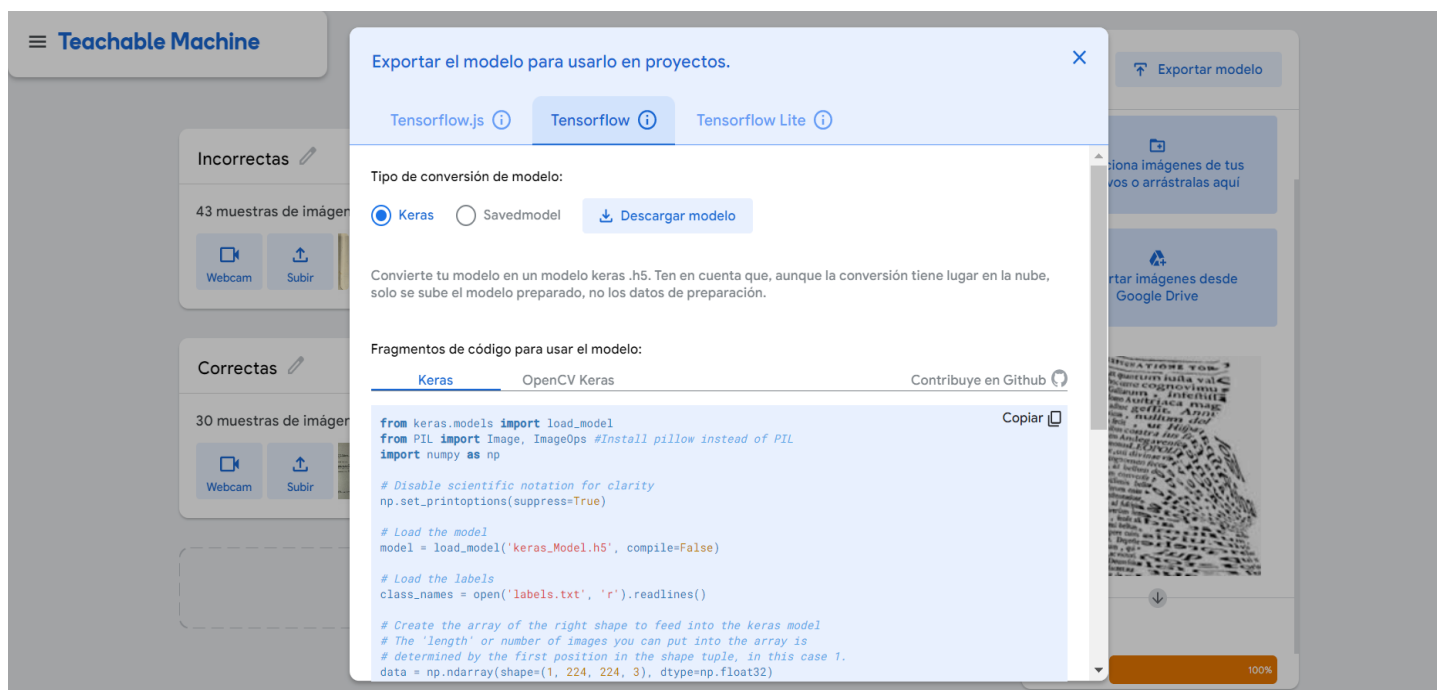
Nota: Elaboración propia, 2022.

Figura 7. Matriz de confusión.**Matriz de confusiones**

Nota: Elaboración propia, 2022.

Una vez generado y entrenado el modelo, se procede a exportarlo. Al respecto, Teachable Machine ofrece tres opciones denominadas Tensorflow, Tensorflow.js y Tensorflow Lite.

Figura 8. Exportación del modelo generado en Teachable Machine.



Nota: Elaboración propia, 2022.

En cuanto a Tensorflow, se puede indicar que es una librería de código libre para Machine Learning, multiplataforma, desarrollada por Google para solventar las necesidades generadas a partir de redes neuronales artificiales, ya que construye y entrena redes neuronales para la detección patrones (Larkin, 2022).

Por otra parte, Tensorflow.js es una biblioteca de aprendizaje automático de código abierto, basada en la biblioteca original de TensorFlow, que se ejecuta en JavaScript (Google, 2022), mientras que Tensorflow Lite es una biblioteca multiplataforma de Machine Learning para ejecutar modelos incluidos en dispositivos móviles iOS y Android (LeViet, 2021).

En cuanto al manejo de la herramienta, no requiere ser instalada en el equipo, porque el modelo se pueda entrenar desde su sitio web (<https://teachablemachine.withgoogle.com/>). En el mismo sitio se explica cómo funciona y cómo exportarlo. Igualmente se indica la existencia de diversos tutoriales en YouTube, lo cual evitaría caer en gastos de capacitación. Finalmente, como ha expuesto antes, es una herramienta, amigable, intuitiva y que no requiere de conocimientos en programación, facilitando de esa forma el proceso de aprendizaje y su manejo.

Capítulo VI: Conclusiones y recomendaciones

A continuación, se desarrollan los principales hallazgos de la investigación. Este apartado aborda, en primera instancia, las conclusiones a partir de los resultados relacionados con los objetivos que orientaron este trabajo, y posteriormente se presentan las recomendaciones surgidas a raíz de la propuesta.

6.1 Conclusiones

De acuerdo con los objetivos específicos de la investigación y los resultados encontrados, se indican a continuación las principales conclusiones para cada uno de ellos.

Del objetivo específico 1: Identificar las fortalezas, oportunidades, debilidades y amenazas del proceso de gestión y evaluación de la calidad de las digitalizaciones, se desprende que:

- Se detectaron elementos positivos como el dominio de las tareas por parte de los funcionarios del ARD y su disposición para trabajar con una metodología que facilite la tarea de la gestión de calidad de las imágenes. De igual modo, que se muestran favorables emplear las TIC para agilizar sus labores diarias.
- Es significativo mencionar que, si bien tienen conocimientos incipientes sobre la Inteligencia Artificial, existe en ellos disposición por aprender, incluso mediante la generación de un documento formal para establecer esta actividad como un procedimiento formal.
- Una de las principales debilidades, según los entrevistados, es la obsolescencia del equipo, pero esta se ve influenciada por una de las amenazas, que es la

restricción presupuestaria por parte del gobierno de Costa Rica, lo que provoca que por el momento no se pueda invertir en recursos tecnológicos.

- Es importante instaurar una coyuntura de trabajo entre el Departamento de Conservación y el Departamento de Tecnologías de la Información, de manera que la implementación de esta propuesta se pueda llevar a cabo lo antes posible.

Con respecto al objetivo específico 2: Diseñar un plan para la automatización de la gestión y evaluación de la calidad de las digitalizaciones realizadas en la Dirección General del Archivo Nacional, se obtuvo que:

- De acuerdo con los resultados se puede concluir que el sistema cuenta con un 100% de efectividad en los procesos de prueba a partir de las imágenes utilizadas, que fueron un total de 73, para alimentar y entrenar el modelo. Se debe destacar que para este proyecto se contó con dos categorías para entrenar el modelo, debido a que se tiene un número limitado de imágenes.
- A partir de las pruebas efectuadas se logró establecer un grado de precisión del 100%, en un total de en un total de 150 imágenes evaluadas. Eso a partir de una validación con imágenes obtenidas de la base de datos del Archivo Nacional, con lo que se garantiza una apreciación integral de las imágenes, diferente de cuando se realizaba de forma manual y no se tenían estadísticas que permitieran conocer el margen de error.
- Al evaluar la eficacia del prototipo en el mejor escenario posible, se consiguió concluir que el modelo propuesto permite alcanzar lo que se deseaba, demostrando el gran aporte que puede brindar el uso de conceptos como Deep Learning en áreas en las que se les podría considerar ajenas a las de informática; por lo que se puede intuir la posibilidad de que más sectores productivos se inclinen por este auge tecnológico, abriendo la posibilidad de mejorar sus procesos operativos y a su vez optimizar sus recursos.

Finalmente, las conclusiones de este apartado hacen referencia al objetivo general de la investigación: Elaborar una propuesta para la automatización del proceso de gestión y evaluación de la calidad de las digitalizaciones realizadas en el Departamento de Conservación de la Dirección General del Archivo Nacional utilizando Deep Learning, destacando lo siguiente:

- En este trabajo se desarrolló un modelo que ayuda a automatizar el control de calidad de las digitalizaciones, mediante la identificación de una muestra de imágenes defectuosas, empleando Deep Learning para el procesamiento digital de imágenes. La finalidad es el detectar aquellas imágenes que presenten defectos, sin necesidad de que una persona las revise una por una, o bien, que aplique un muestreo con el que se puede omitir algún error.
- Los resultados obtenidos con la muestra de prueba en la clasificación de los ejemplares con respecto a las evaluaciones hechas en forma manual por un perito, no presentaron diferencias.

6.2 Recomendaciones

De acuerdo con la propuesta desarrollada en la investigación, se brindan las principales recomendaciones.

- Emplear la propuesta desarrollada en esta investigación, que arrojó resultados positivos. Es fácil de utilizar y no representa un gasto económico para la institución, ya sea por compras de licencias o de capacitación.
- Utilizar este tipo de herramientas como Teachable Machine, porque es de uso libre, lo cual es una ventaja debido a que no se incurriría en gastos financieros, apegándose al momento por el que pasa el sector público del país, además de que su implementación y uso es muy sencillo.

- Establecer un banco de imágenes (dataset) y alimentarlo constantemente con nuevos elementos con diferentes errores para entrenar el modelo y mantener una mejor precisión en el reconocimiento de fallos en las digitalizaciones.
- El modelo es dependiente de valores empíricos, por lo que para aumentar su robustez es necesario expandir el universo de ejemplares que se utilizan para obtenerlos, o bien, determinar técnicas de inteligencia artificial que garanticen el aprendizaje continuo y se ajusten permanentemente.
- A partir de la utilización de este modelo, se deben establecer diferentes categorías de errores detectados en las imágenes digitalizadas, de modo que se generen estadísticas a futuro para un mejor control de esta actividad en el Departamento de Conservación de la Dirección General del Archivo Nacional.

Referencias

- Araya, A. (2019). Plan para la gestión de productos de tecnologías de información y comunicación (tic) elaborados por los estudiantes de la carrera de Bachillerato en Informática Empresarial para el recinto de Turrialba, Universidad de Costa Rica. [Tesis de maestría]. Universidad de Costa Rica. <https://kerwa.ucr.ac.cr/handle/10669/81022>
- Archivo General de la Nación de Colombia. (2018). Fundamentos de preservación digital a largo plazo. https://www.archivogeneral.gov.co/caja_de_herramientas/docs/6.%20preservacion/DOCUMENTOS%20TECNICOS/FUNDAMENTOS%20PRESERVACION%20DIGITAL%20LARGO%20PLAZO.pdf
- Archivo Nacional. (2021a). Comisión Nacional de Selección y Eliminación de Documentos. https://www.archivonacional.go.cr/index.php?option=com_quix&view=page&id=32#qx-text-981903
- Archivo Nacional. (2021b). Departamento de Conservación. <https://www.archivonacional.go.cr/index.php/institucional/content-component-4/conservacion>
- Archivo Nacional. (2021c). Marco Estratégico. https://www.archivonacional.go.cr/index.php?option=com_quix&view=page&id=13#qx-section-5576
- Archivo Nacional. (2021d). Preguntas Frecuentes. https://www.archivonacional.go.cr/index.php?option=com_quix&view=page&id=13#qx-section-5576
- Barbachan, A., Morales, Á., Mosquera, J. y Soto, D. (2015). Calidad en las instituciones y empresas del sector de servicios públicos del Gobierno Central en el Departamento de Lima. [Tesis de maestría]. Pontificia Universidad Católica del Perú.
- Barrantes, R. (2015). Métodos de estudio a distancia e investigación: a la búsqueda del conocimiento científico. San José, C.R.: EUNED
- Bello, D., Murrieta, F. y Peralta, A. (2018). Análisis de los procesos industriales para la generación de una propuesta de mejora continua mediante la planeación y diseño de sistemas de calidad. *Revista Ciencia Administrativa* 3. 107-125.
- Camisón, C., Cruz, S., y González, T. (2007). Gestión de la calidad: conceptos, enfoques, modelos y sistemas. Madrid: Pearson.
- Carneiro, T., Da Nóbrega, R., Nepomuceno, T., Bian, G., De Albuquerque, V. y Reboucas Filho, P. (2018). Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications. *IEEE Access*. 6, 61677-61685.
- Castro, M. (2013). Propuesta de un sistema integrado de control de la gestión para la empresa Constructora Valba S.A. Costa Rica. [Tesis de maestría]. Universidad de Costa Rica. <https://www.kerwa.ucr.ac.cr/handle/10669/76019>
- Cognex Corporation. (2022). Deep Learning para la automatización de fábricas. Combinando la inteligencia artificial con la visión industrial. <https://www.cognex.com/es-mx/resources/white-papers-articles/whitepaperandarticlemain?event=2f4036b3-8746-4f1f-b7b7-f6be37e2645c>
- Corcuera, P. (2017). Introducción a la programación en Python. <https://personales.unican.es/corcuerp/python/Slides/IntroPython.pdf>

- Cuartas, H. (2012). Estandarización de los procesos de producción en la empresa Construcciones Cuartas. [Trabajo de grado]. Universidad Autónoma de Occidente. Colombia. <https://red.uao.edu.co/handle/10614/3028?show=full>
- Delgado, A. (2016). Sistema de protección de los intereses del autor y el empresario en la propiedad intelectual (régimen del empresario en el ámbito de la propiedad intelectual). [Tesis doctoral]. Universitat Ramon Llull. España. <https://www.tdx.cat/handle/10803/400934?show=full>
- Deloitte. (2016). La era de la Automatización Implementación de Robotics en los Centros de Servicios Compartidos. <https://www2.deloitte.com/ni/es/pages/strategy/articles/rpa-en-csc.html>
- Deloitte. (2017). Automatización Robótica de Procesos (RPA). <https://www2.deloitte.com/uy/es/pages/home/articulos/Automatizacion-Robotica-de-Procesos.html>
- Denzin, N. y Lincoln, Y. (2005). The Sage Handbook of Qualitative Research. Londres, Inglaterra: Sage.
- Duarte, V. (2020). Desarrollo de una aplicación web responsiva para la automatización del proceso de gestión de expedientes de información de los locales comerciales en ASP.NET MVC del Área Rectora de Salud en el Ministerio de Salud, [Trabajo final de graduación de licenciatura]. Universidad Estatal a Distancia. Costa Rica.
- Franco, B. y Pérez, R. (2014). Guía de Implementación Gerencial– Política de gestión de documentos y archivos. https://www.archivonacional.go.cr/web/dsae/politica_gestion.pdf
- García, M., Quispe A., Ráez G. (2003). Mejora continua de la calidad en los procesos. *Datos Industriales*. 6(1), 89-94. <https://www.redalyc.org/articulo.oa?id=81606112>
- Glass, G. y Hopkins, K. (1996). Statistical methods in education and psychology. Englewood Cliffs, New Jersey. Prentice-Hall.
- Google Cloud. (2022). AutoML Vision. <https://cloud.google.com/vision/automl/docs?hl=es-419>
- Google Cloud. (2022). TensorFlow.js:Crea tu propia. <https://codelabs.developers.google.com/tensorflowjs-transfer-learning-teachable-machine?hl=es-419#0>
- Güendián, L. (2019). Uso de Deep Learning para la digitalización de documentos históricos. [Trabajo fin de grado]. Universidad Carlos III de Madrid. España. <https://e-archivo.uc3m.es/handle/10016/29329>
- Hernández, R., Fernández, C. y Baptista, M. (2014). Metodología de la investigación. México: McGraw-Hill. <https://www.uca.ac.cr/wp-content/uploads/2017/10/Investigacion.pdf>
- Junta Administrativa del Archivo Nacional. (2021). Norma Técnica Nacional NTN-003 Digitalización de documentos textuales en soporte papel del 13 de diciembre de 2021. Costa Rica. https://www.archivonacional.go.cr/index.php?option=com_content&view=article&id=142
- Junta Administrativa del Archivo Nacional. (2021). Norma Técnica Nacional NTN-004 Lineamientos para la digitalización de documentos textuales con la finalidad de sustituir el soporte original del 16 de diciembre de 2021. Costa Rica. https://www.archivonacional.go.cr/index.php?option=com_content&view=article&id=142
- Junta Administrativa del Archivo Nacional. (2018). Norma técnica para la gestión de documentos electrónicos en el Sistema Nacional de Archivos del 21 de mayo de 2018. Costa Rica. https://www.archivonacional.go.cr/index.php?option=com_content&view=article&id=142

- KIO Networks. (2021). AI, RPA y Machine Learning – ¿Qué tienen en común y cuáles son sus diferencias? <https://www.kionetworks.com/frontierpost/ai-rpa-y-machine-Learning-que-tienen-en-comun-y-cuales-son-sus-diferencias>
- Larkin, J. (2022). ¿Qué es TensorFlow y para qué sirve? Incentro. <https://www.incentro.com/es-ES/blog/que-es-tensorflow>
- LeViet, K. (2021). Compila e implementa un modelo de detección de objetos personalizado con TensorFlow Lite (Android). <https://developers.google.com/codelabs/tflite-object-detection-android?hl=es-419#0>
- Ley de certificados, firmas digitales y documentos electrónicos N° 8454 de 30 de agosto 2005. Costa Rica. https://www.archivonacional.go.cr/index.php?option=com_content&view=article&id=139
- Ley del Sistema Nacional de Archivos N° 7202 de 24 de octubre de 1990. Costa Rica. https://www.archivonacional.go.cr/index.php?option=com_content&view=article&id=139
- Lombardi, O. (2016). Reconocimiento de una Plataforma de gestión de la calidad sobre la cual se pueda establecer la gestión de la innovación, en una mediana empresa peruana. [Tesis de maestría]. Pontificia Universidad Católica del Perú
- López, K. (2015). Modelo de automatización de procesos para un sistema de gestión a partir de un esquema de documentación basado en Business Process Management (BPM). *Universidad & Empresa*. 17(29), 131-155. <https://www.redalyc.org/articulo.oa?id=187243745006>
- López, R. (2017). Libro online de IAAR. <https://iaarbook.github.io/>
- Microsoft. (2022) ¿Qué es Azure Machine Learning? <https://learn.microsoft.com/es-es/azure/machine-learning/overview-what-is-azure-machine-learning>
- Ministerio de Cultura y Juventud. (2020). Manual de Cargos. Costa Rica.
- Ministerio de Educación, Cultura y Deporte. (2015). Política de Gestión de Documentos Electrónicos del Ministerio. España. <https://www.culturaydeporte.gob.es/cultura/archivos/recursos-profesionales/documentos-electronicos.html>
- Ministerio de Fomento. (2005). Modelos para implantar la mejora continua en la gestión de empresas de transporte por carretera. España.
- Odorica, I. (2020). ¿Qué es Google Cloud y para qué sirve? Incentro. <https://www.incentro.com/es-ES/blog/que-es-google-cloud-platform>
- Parvat, A., Chavan, J., Kadam, S., Dev, S. & Pathak, V. (2017). A survey of deep-Learning frameworks. En 2017 International Conference on Inventive Systems and Control (ICISC) (pp. 1-7). doi:10.1109/ICISC.2017.8068684
- Pensis. (2018). Inteligencia artificial, la revolución que apenas comienza. <https://www.tec.ac.cr/pensis/articulos/inteligencia-artificial-revolucion-apenas-comienza>
- Pereira-Uzal, J. (2018). El control de calidad en los planes de digitalización. *Revista PH*. (95). 63-67. <https://doi.org/10.33349/2018.0.4226>
- Reglamento a la Ley 7202 del Sistema Nacional de Archivos, Decreto Ejecutivo N 24023-C de 30 de enero de 1995. Costa Rica. http://www.pgrweb.go.cr/scij/Busqueda/Normativa/Normas/nrm_texto_completo.aspx?param1=NRTC&nValor1=1&nValor2=84751&nValor3=109465&strTipM=TC

- Riba Campos, C.-E. (2013). La observación participante y no participante en perspectiva cualitativa. Universitat Oberta de Catalunya. <http://hdl.handle.net/10609/69986>
- Rincón, J. (2009). Análisis y mejora de los procesos de gestión en un parque de maquinaria. [Proyecto fin de carrera]. Universidad Carlos III de Madrid. España. <https://e-archivo.uc3m.es/handle/10016/8194>
- Robledano-Arillo, J.; Moreno-Pelayo, V.; Pereira-Uzal, J. M. (2016). Aproximación experimental al uso de métricas objetivas para la estimación de calidad cromática en la digitalización de patrimonio documental gráfico. *Revista Española de Documentación Científica*, 39(2): e128. doi: <http://dx.doi.org/10.3989/redc.2016.2.1249>
- Rodríguez, K. (2019). Propuesta de un plan de utilización de metodologías ágiles para apoyar los procesos de calidad en SisCard de Evertec [Trabajo final de graduación de licenciatura]. Universidad Estatal a Distancia. Costa Rica.
- Romero, D. (2009). Propuesta de automatización de los procesos de verificación y despachos en una empresa panificadora. [Trabajo de grado]. Pontificia Universidad Javeriana. Colombia.
- Romero, N. y Cordero, R. (2020). Propuesta de un modelo para el análisis estratégico de la solidez financiera de la Asociación Solidarista de Empleados de la Caja de Ande y Caja de Ande Seguros S.A. (ASECA), Costa Rica. [Tesis de maestría]. Universidad de Costa Rica. <https://kerwa.ucr.ac.cr/handle/10669/81547>
- RStudio. (2017). Aprendizaje Profundo con Keras: guía rápida. https://raw.githubusercontent.com/rstudio/cheatsheets/main/translations/spanish/keras_es.pdf
- Teachable Machine. (2017) ¿Qué es Teachable Machine? <https://teachablemachine.withgoogle.com/faq>
- Umaña, R. y Marín, K. (2021). Representación digitalizada probatoria de documentos: del mito al hito. *Revista del Archivo Nacional*, 85, 509. <http://www.dgan.go.cr/RAN/index.php/RAN/article/view/509>
- UNE Normalización Española. (2011). Información y documentación. Directrices para la implementación de la digitalización de documentos (UNE-ISO/TR 13028:2011 IN). <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma/?c=N0048666>
- UNE Normalización Española. (2013). Norma para el Proceso de migración y conversión de documentos electrónicos. (ISO 13008-2013). <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma/?c=N0052128>
- UNESCO. (2009). Charter on the Preservation of the Digital Heritage. <https://unesdoc.unesco.org/ark:/48223/pf0000179529.page=2>
- Vásquez, P. (2020). Uso del aprendizaje automatizado y de variables climáticas como herramienta para la predicción del riesgo de dengue en Costa Rica, período 2007-2017. [Tesis de maestría]. Universidad de Costa Rica. <https://kerwa.ucr.ac.cr/handle/10669/82278>
- Vesga, J. (2013). Cultura organizacional y sistemas de gestión de la calidad: una relación clave en la gestión de las instituciones de educación superior. *Revista Guillermo de Ockham* 11(2). pp. 89-100.

Apéndices

Apéndice A. Guía de entrevista semiestructurada para las personas funcionarias del Área de Reprografía Documental del Archivo Nacional.



PPTICGO

Programa de Posgrado en
**Tecnologías de Información y
Comunicación para la Gestión
Organizacional**

UNIVERSIDAD DE COSTA RICA

Posgrado en Tecnologías de Información y Comunicación para la Gestión
Organizacional

Guía de entrevista semiestructurada para las personas funcionarias del Área de Reprografía Documental del Archivo Nacional.

Fecha: _____
Lugar: _____
Entrevistador: _____
Hora de inicio: _____
Hora de finalización: _____

El presente instrumento tiene como finalidad la recolección de datos para la realización del Trabajo Final de Investigación Aplicada titulado “*Automatización del proceso de gestión y evaluación de la calidad de las digitalizaciones realizadas en el Departamento de Conservación de la Dirección General del Archivo Nacional utilizando Deep Learning*”. Toda la información aquí brindada es confidencial y únicamente será utilizada con fines académicos.

El instrumento está conformado de 3 partes. La primera parte es de datos personales básicos, la segunda parte pretende determinar la situación del proceso de gestión de la calidad y la tercera y última parte se centra en los parámetros requeridos para evaluar la calidad.

I Parte. Datos personales

1. ¿Cuál es su grado académico? _____
2. ¿Cuántos años tiene de laborar en la institución? _____
3. Edad: _____

II Parte. Categoría I: Gestión de la calidad.

4. ¿Conoce los procedimientos establecidos para la digitalización de documentos en el Archivo Nacional? Descríbalo.
5. ¿Cuál es el protocolo de digitalización de documentos? Descríbalo.
6. ¿Cuáles son las actividades para llevar a cabo la digitalización de documentos históricos? Enumérelas y descríbalas y ¿Cuánto se dura digitalizando un documento?
7. ¿Cuáles son los recursos tecnológicos con los que cuentan para el proceso de digitalización de documentos históricos?
8. ¿En qué formato se digitalizan los documentos históricos? Revise su es singular o plural, en caso de que sean varios los usados
9. ¿Existe algún manual o procedimiento sobre el control de calidad de las imágenes digitalizadas?
10. ¿Cuántas personas realizan el control de calidad de las imágenes digitalizadas? Y ¿Cuáles son sus puestos?
11. ¿Cuáles son las actividades que se desarrollan durante el control de calidad de las imágenes?
12. Aparte de usted, ¿existe otro filtro de control de calidad de las imágenes? ¿Cuál es?
13. ¿Llevan estadísticas del tiempo empleado en el proceso completo de la digitalización de documentos?
14. Del tiempo empleado en la digitalización, ¿Cuánto corresponde al control de calidad?
15. ¿Quién corrige los errores encontrados?

III Parte. Categoría II: Evaluación de la calidad

16. ¿Conoce alguna norma sobre la digitalización de documentos?
17. ¿Sabe qué es una copia maestra o derivada?
18. ¿Cuál es el nivel de precisión o margen de error en el control de calidad que realizan?

Muchas gracias

Apéndice B. Guía de observación al control de calidad de las imágenes digitalizadas.



PPTICGO

Programa de Posgrado en
**Tecnologías de Información y
Comunicación para la Gestión
Organizacional**

UNIVERSIDAD DE COSTA RICA

Posgrado en Tecnologías de Información y Comunicación para la Gestión
Organizacional

Guía de observación al control de calidad de las imágenes digitalizadas

Fecha: _____

Lugar: _____

Observador: _____

Hora de inicio: _____

Hora de finalización: _____

Introducción

El objetivo de la observación es conocer cómo se lleva a cabo el control de calidad de las imágenes digitalizadas en el Archivo Nacional. La información recolectada se utilizará como parte del Trabajo Final de Investigación Aplicada titulado “*Automatización del proceso de gestión y evaluación de la calidad de las digitalizaciones realizadas en el Departamento de Conservación de la Dirección General del Archivo Nacional utilizando Deep Learning*”.

Aspectos por observar	Sí	No	No aplica	Observaciones
I Parte. Categoría I: Gestión de la calidad.				
1. El personal aplica el protocolo de digitalización de documentos				
2. Se determina si es una copia maestra o derivada, antes de hacer el control de calidad				
3. Tiempo durado en la digitalización				
4. Tiempo durado en control de calidad				

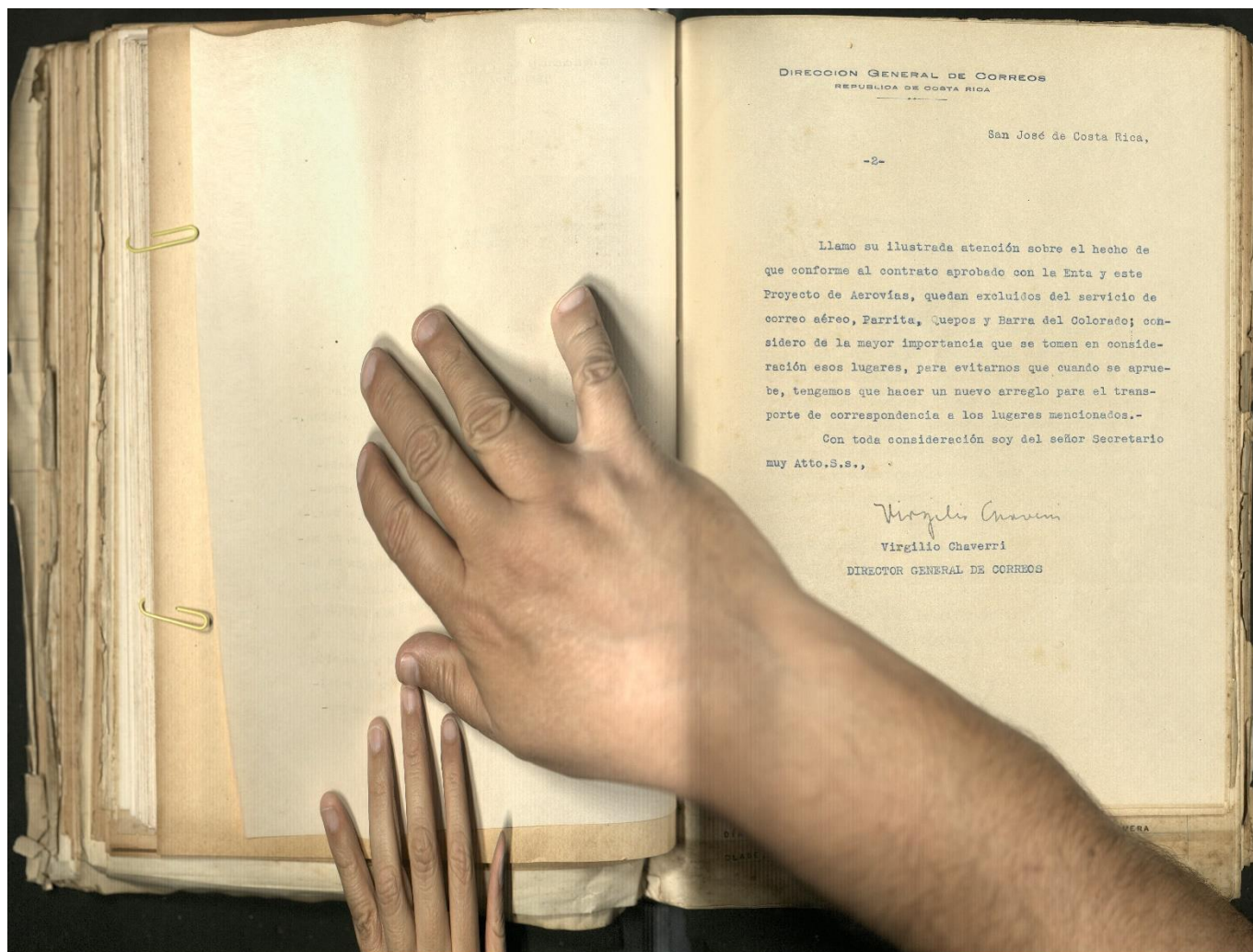
5. Filtros en el control de calidad				
6. Cantidad de folios revisados				
7. Cantidad de imágenes revisadas				
II Parte. Categoría II: Evaluación de la calidad.				
8. Es una copia maestra				
9. Es una copia derivada				
10. Existen errores				
11. Solventan los errores				
12. Cantidad de correcciones realizadas				
13. Existen informes sobre el control de calidad				
14. Existen estadísticas sobre el control de calidad.				

Anexos

Anexo 1. Instrumento para el control de calidad de imágenes en Microsoft Excel.

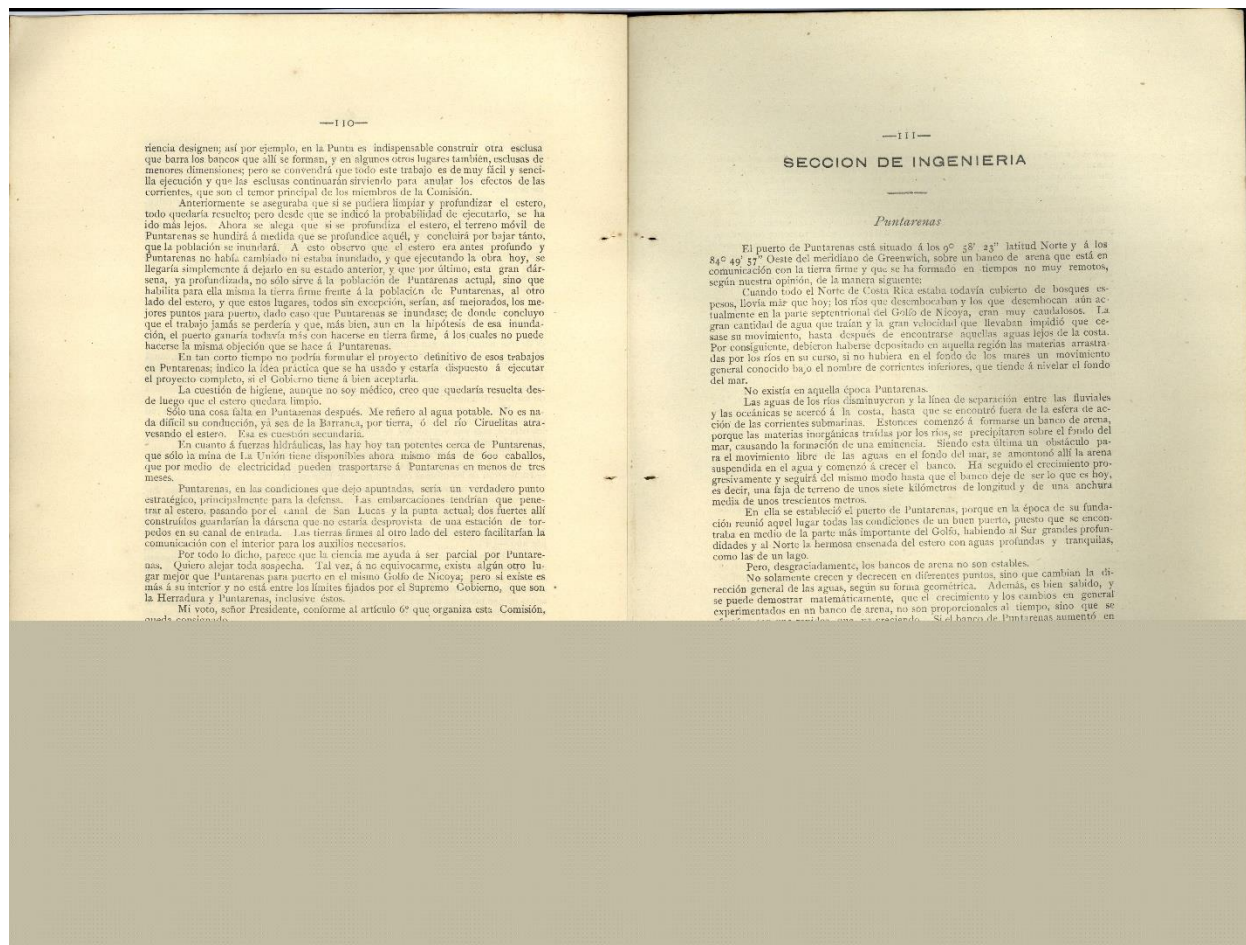
Excel Compartido de control de digitalizaciones 2022 (2) - Guardado				
Buscar (Alt + Q)				
Archivo Inicio Insertar Dibujo Diseño de página Fórmulas Datos Revisar Vista Ayuda				
Edición Compartir				
Cortar Copiar Pegar Copiar formato Fuente Ajustar texto Combinar y centrar Alineación Formato condicional Dar formato como tabla Estilos Insertar Eliminar Formato Autosuma Borrar Ordenar y filtrar Buscar y seleccionar				
B29				
	B	C	D	E
	Números extremos de signatura	Cantidad de imágenes	Nombre persona que hizo el control de calidad	Inconsistencias encontradas
2				
3	#8872-4 / #8872-6	426		Varias imágenes vueltas pero sin correcciones
5	#8873-1 / #8873-15	770		#8873-3 (Imágenes #10 y #12 cortadas repetir) / #8873-15 #253 hoja al revés repetir o eliminar si no tiene información. Varias imágenes vueltas.
6	#8873-16 / #8875	565		#8873-29 (Imágenes #18 y #19, #23 y #24, #27 y #28 repetidas revisar o eliminar). Varias imágenes vueltas.
7	#8876-1 / #8877	946		#8876-1 (Imagen # 156 desenfocada repetir) #8876-2 (Imagen #236 con mano/#266 y # 267 , #291 y #292 repetidas) #8876-3 (#39 y#40 repetidas SON COPIAS) #8877 (Imágenes #11 doblada repetir/#5 y #6 repetidas). Varias imágenes vueltas.
8	#8878-1	485		#8878-1 (Imágenes #205 y # 206 repetidas). Algunas imágenes vueltas.
9	#8878-2	330		Ninguna
10	#8886-14 - #8886-15	278		#8886-14 (Imagen #168 repetir está movida)
11	#8886-16 - #8886-17	270		Ninguna
12	#8886-18 - #8886-19	176		#8886-19 (Imagen #16 repetir está tapada) imágenes #64 y #65 están repetidas.
13	#8887, #8888, #8889-1	52		#8877 (#5 y #6 repetidas)
14	#8889-2 - #8889-3	116		Ambos está con la costura muy ajustada y se cortan algunos textos internos. Varias imágenes vueltas.
15	#8889-4, #8889-5, #8889-6	133		Se cortan algunos textos internos
16	#8870-3	320		Varias imágenes vueltas pero sin correcciones. Tiene la palabra image al inicio de cada signatura.
17	#8889-7 - #8889-13	508		#8889-7 (Imágenes #26 y #27 repetidas en su lugar en la #26 falta folio 4 / imágenes #63 y # 64 repetidas) #8889-8 (Falta portada / imagen #28 doblada repetir/ algunas costura muy ajustada y se cortan algunos textos internos) 8889-10, #8889-12. (algunas costura muy ajustada y se cortan algunos textos internos)
18	#8889-14 - #8889-16	373		8889-16 (Imágenes #88 #89 #90) Algunas costura muy ajustada y se cortan algunos textos internos en las tres signaturas.
19	#8912-1 - #8914-2	608		Documentos con grapas y pines quedan sueltos al quitarlas.
20	#8914-3 - #8917	1536		Documentos con grapas y pines quedan sueltos al quitarlas.
21	#8889-17 - #8890-3	416		8889-17 (eliminar imagen #81 y #82 / #112, #113, #114 / #115, # 116 # 117/ #118, #119, #120, #121 / #136, #137 / #196 y #197 repetidas revisar) 8890-1 (#81, #83, #84, #92, #93, #94, #105, #108, #114, #125 no están legibles, revisar papel complicado letra borrosas) 8890-3 Costura muy ajustada y se cortan algunos textos internos en las tres signaturas.
22	#8890-4 - #8890-7	166		8890-6 / 8890-7 Costura muy ajustada y se cortan algunos textos internos
23	#8917-3 - #8917-4	971		Documentos con grapas y pines quedan sueltos al quitarlas.
24	#8918-1 - #8918-4	2004		Ninguna
25	#008890-8 -#008890-21	758		008890-8(Texto corado en el lomo en la imagen # 64) / 008890-9(Texto corado en el lomo en las imágenes # 32,34,37,43,62,68) / 008890-11(Texto corado en el lomo en las imágenes # 37,38,45, 46, 79,80) repetidas las imágenes 51 y 52 (corregido) / 008890-12 (Texto corado en el lomo en las imágenes # 11, 16, 37, 82, 83) / 008890-13(papel que tapa el texto en imagen # 34(es un orificio en el papel que deja ver el papel blanco de base), posible firma cortada en imagen # 37) (Firma en el lomo parcialmente cortada por la costura de los documentos) / 008890-14(Texto cortado en el lomo en las imágenes # 16, 19, 20, 35, 41, 42, 54, 58, 59, 62, 66, 69, 78,79, 81, 84, 85, 88) / 008890-16(Texto cortado en el lomo en las imágenes # 3,5,6,7,8,10,11,12,13,16,17,18,20,22,23,24,25,26,28,30,31,32,33) / 008890-21(Texto cortado en el lomo en las imágenes # 8,10,22,25,27,28,31,33,34,35,37,42,44,48,57,59,60,61,66,68,70,71) (Los textos cortados en el lomo solo se pueden corregir descosiendo el documento por tener la costura dentro del mismo texto)

Nota: ARD, DCONS, AN, 2022.

Anexo 2. Imagen defectuosa, imagen con mano.

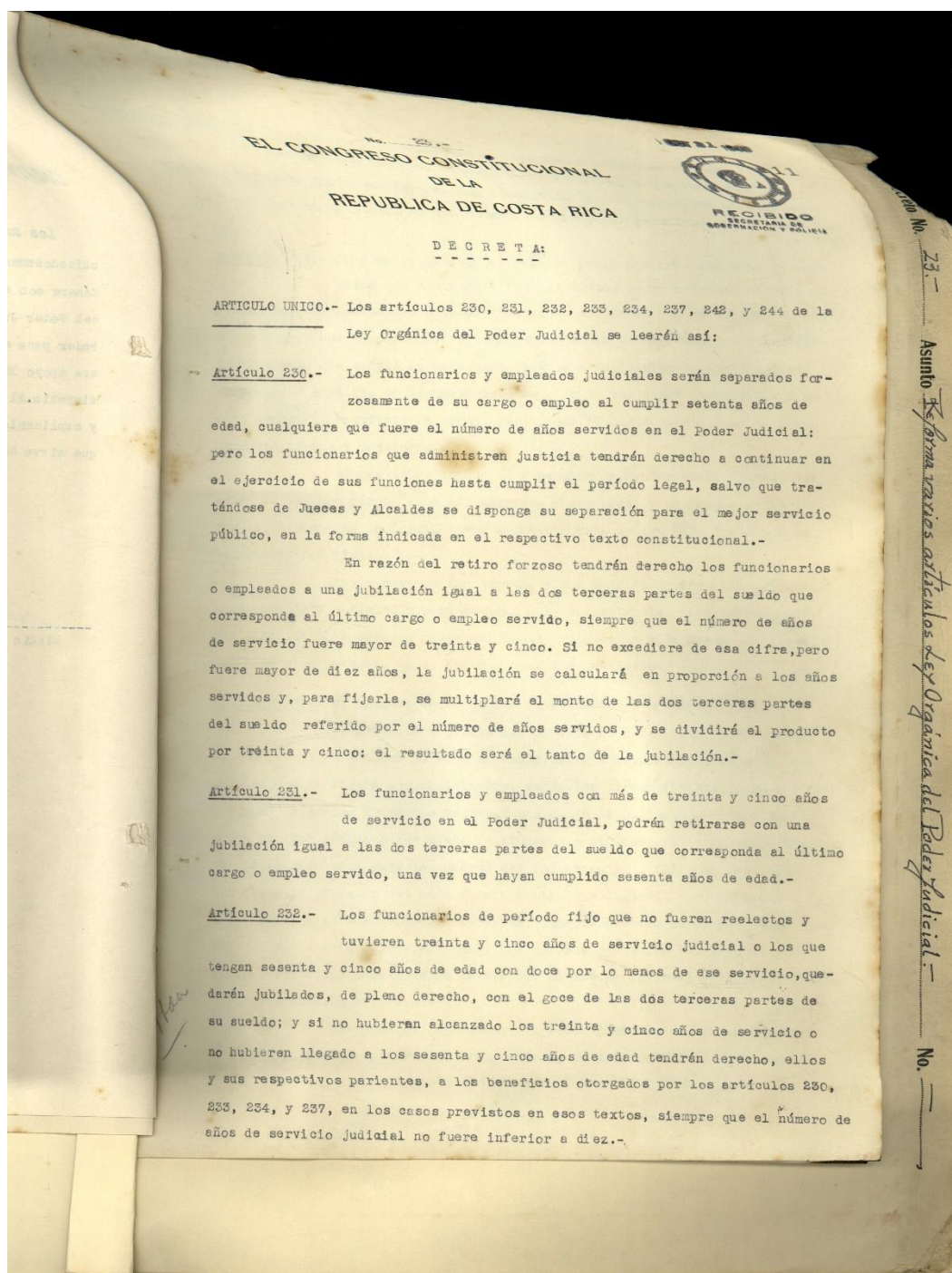
Nota: ARD, DCONS, AN, 2022.

Anexo 3. Imagen defectuosa, imagen cortada.



Nota: ARD, DCONS, AN, 2022.

Anexo 4. Imagen defectuosa, imagen movida.



Nota: ARD, DCONS, AN, 2022.

Anexo 5. Certificación de revisión filológica.

Gestión de Negocios
Servicios Educativos Profesionales



San José, 01 de junio de 2023

SF-067-2023

Señores
Sistema de Estudios de Posgrado
Universidad de Costa Rica
S. D.

Estimados señores:

En mi condición de profesional en el Área de la Filología y Letras, doy fe de haber leído y revisado el trabajo final de graduación titulado **“Automatización del proceso de gestión y evaluación de la calidad de las digitalizaciones realizadas en el Departamento de Conservación de la Dirección General del Archivo Nacional utilizando Deep Learning”**, escrito por Max Zúñiga Fallas, cédula de identidad número 1-1409-0369, para optar por la Maestría Profesional en Tecnologías de la Información y Comunicación para la Gestión Organizacional.

Se han revisado y hecho las modificaciones pertinentes en distintos niveles textuales, macro y microestructura, aspectos gramaticales, de puntuación y ortografía, vicios del lenguaje, intención comunicativa, coherencia y otros aspectos relacionados en el campo filológico, que se manifestaron en el documento escrito.

De ustedes, atentamente,

Lic. Gustavo Castro Miranda
Carné COLYPRO #29873
Cédula 2-0463-0329