



Revista Internacional de Investigación e Innovación Tecnológica

Página principal: www.riit.com.mx

Sistema para la Evaluación Automatizada del Test del Dibujo de la Figura Humana a través de un Modelo de Visión Computacional

Automated Human Figure Drawing Test Assessment System through a Computational Vision Model

Jiménez-Fernández, M.C.*, Hernández-Coronado, M., Berlanga-Mendoza, J.A., Carreón-Reyes, J.A., Palomo-Martínez, P.J.

Escuela de Ingeniería y Tecnologías; Universidad de Monterrey; C.P. 66238, San Pedro Garza García, Nuevo León.
consuelo.jimenez@udem.edu*; melania.hernandez@udem.edu; jose.berlangam@udem.edu;
juan.carreonr@udem.edu; pamela.palomo@udem.edu

Innovación tecnológica: Desarrollo de una herramienta tecnológica para la evaluación temprana y precisa de problemas cognitivos basada en el análisis del dibujo de la figura humana.

Área de aplicación industrial: Psicología.

Recibido: 28 noviembre 2023

Aceptado: 06 septiembre 2024

Abstract

Cognitive disorders in childhood, particularly intellectual disability in Mexico and other countries, represent a significant public health challenge. These are often not adequately detected in early stages, limiting timely interventions crucial for the development and well-being of children. The projective technique of Human Figure Drawing (HFD), used by psychologists for the early detection of cognitive problems in children, is limited by a time-consuming process and a subjective manual analysis. In response to these challenges, this paper details the development of an automated system for the evaluation of HFD in children aged 5 to 12 years in the Metropolitan Area of Nuevo León, Mexico, with the aim of improving the early detection of cognitive disorders. The hypothesis proposed is that the developed tool will enable the identification of HFD graphic indicators with an accuracy of over 90%, reducing the evaluation time from 10 minutes to less than 30 seconds. The methodology involves the collection of over 1,000 drawings from the application of HFD in the target population, followed by their processing and labeling to convert them into training data for a machine learning model. As part of the results, a computer vision model was developed, capable of detecting up to 24 graphic indicators in a human figure drawing in less than

1 second. The main findings of the study reveal a reduction in the evaluation task time of an HFD by 99.83%, and a mean Average Precision (mAP50) of 92.8, a precision (P) of 0.912 and a recall (R) of 0.902, highlighting the model's effectiveness in correctly identifying the elements in the drawings. The complete system offers a web application dedicated to evaluating the drawings, thus facilitating the user's interaction with the proposed model. The system integrates technologies such as Python, PyTorch, and YOLO for the processing and analysis of the drawings, and Vue.js, C#, and SQL Server for the web application development. This technological advance presents significant potential applications in hospitals, clinics, mental health centers, and educational settings. It represents a valuable tool for diagnosing and treating cognitive disorders in children, improving not only the efficiency and objectivity in HFD evaluation but also opening new possibilities for addressing challenges in children's mental health.

Key words: Children psychology, Computer vision, Human figure drawings, Machine learning, Technology.

Resumen

Los trastornos cognitivos en la infancia, especialmente la discapacidad intelectual en México y otros países, representan un grave desafío de salud pública. A menudo no se detectan adecuadamente en etapas tempranas, lo que limita las intervenciones oportunas cruciales para el desarrollo y bienestar de los infantes. La técnica proyectiva del Dibujo de la Figura Humana (DFH), usada por psicólogos para la detección temprana de problemas cognitivos en niños y niñas, se ve limitada por un proceso tardado y un subjetivo análisis manual. En respuesta a estos desafíos, este trabajo detalla el desarrollo de un sistema automatizado para la evaluación de DFH en niños y niñas de entre 5 y 12 años de la Zona Metropolitana de Nuevo León, México, con el objetivo de mejorar la detección temprana de trastornos cognitivos. La hipótesis planteada es que la herramienta desarrollada permitirá identificar los indicadores gráficos del DFH con una precisión de más de 90%, reduciendo el tiempo de evaluación de 10 minutos a menos de 30 segundos. La metodología parte de la recolección de más de 1,000 dibujos a partir de la aplicación del DFH en la población objetivo, siguiendo con su tratamiento y etiquetado para convertirlos en datos de entrenamiento de un modelo de aprendizaje automático. Como parte de los resultados, se desarrolló un modelo de visión computacional, capaz de detectar hasta 24 indicadores gráficos en un dibujo de la figura humana en menos de 1 segundo. Los hallazgos principales del estudio revelan una reducción del tiempo de la tarea de evaluación de un DFH en un 99.83% y un mean Average Precision (mAP50) de 92.8, una precisión (P) de 0.912 y una sensibilidad (R) de 0.902, destacando la efectividad del modelo en la identificación correcta de los elementos clave en los dibujos. El sistema completo ofrece una aplicación web dedicada a la evaluación de los dibujos, facilitando así la interacción del usuario con el modelo propuesto. El sistema integra tecnologías como Python, PyTorch y YOLO para el procesamiento y análisis de los dibujos, y Vue.js, C# y SQL Server para el desarrollo de la aplicación web. Este avance tecnológico ofrece aplicaciones potenciales significativas en hospitales, clínicas, centros de salud mental y entornos educativos. Representa una herramienta valiosa para el diagnóstico y tratamiento de trastornos cognitivos en niños, mejorando no solo la

eficiencia y la objetividad en la evaluación de DFH, sino también abriendo nuevas posibilidades para abordar desafíos en la salud mental infantil.

Palabras clave: Dibujos de la figura humana, Machine learning, Psicología infantil, Tecnología, Visión computacional.

1. Introducción

En países como México, los trastornos cognitivos en la infancia son un serio problema de salud pública. No obstante, las limitaciones de recursos y expertos, junto con la falta de consenso en la evaluación del desarrollo infantil, obstaculizan la detección temprana y la recolección de información precisa sobre su incidencia [1]. La discapacidad intelectual en niños y niñas es un desafío notable que, en muchos casos, no se detecta tempranamente, exacerbando su vulnerabilidad [2].

La técnica proyectiva del Dibujo de la Figura Humana (DFH) ha sido una herramienta valiosa para psicólogos en la detección temprana de problemas cognitivos y de comportamiento en niños [3]. Fue creada por Goodenough en 1926 y está enfocada en detectar ciertos indicadores gráficos en dibujos realizados por niños y jóvenes, y realizar inferencias psicológicas a partir de los hallazgos.

El análisis de los dibujos realizados por infantes desde una perspectiva emocional siguió con Machover en 1949 y con Koppitz en 1968. Koppitz ideó la metodología de "Draw a Man" para identificar ciertos desórdenes emocionales en niños y niñas, y propuso ideas para la interpretación de señales determinadas. Dicha metodología está basada en la suposición de que "los dibujos de los niños reflejan aquellas emociones que no pueden expresar verbalmente, sus actitudes, conflictos y autoestima" [4]. En la actualidad, el DFH se ha convertido en una herramienta de detección usada por psicólogos para explorar

posibles problemas cognitivos y de comportamiento que pudieran estar afectando a infantes.

La doctora en Psicología, Angélica Quiroga, es una de las profesionales que ha realizado contribuciones a esta técnica proyectiva en la última década [5, 6]. En sus investigaciones, emplea el DFH para identificar problemáticas que afectan a niños, niñas y adolescentes, como la deficiencia intelectual y situaciones de violencia. La metodología consiste en pedirle a los infantes que realicen un dibujo de una persona, para posteriormente evaluarlo con base en ciertos indicadores gráficos, como la presencia de ciertas partes del cuerpo, el sombreado o la omisión de algunos detalles. A partir de los indicadores gráficos encontrados, se realizan las inferencias psicológicas. Una de las principales áreas de oportunidad identificadas en esta metodología, es que el análisis manual de los dibujos es un proceso que consume tiempo y está sujeto a errores humanos y sesgos, dependiendo de la experiencia del experto con el método.

El objetivo de este trabajo se enfoca en desarrollar, en un periodo de 16 semanas, un sistema para la evaluación automática de dibujos de la figura humana realizados por niños y niñas de entre 5 y 12 años, reconociendo indicadores gráficos esperados y excepcionales, a través de un modelo de visión computacional, con el fin de mostrar la presencia o ausencia de dichos indicadores y su relación con el nivel intelectual mediante reglas establecidas, a través una aplicación web. Su valor radica en que actualmente no existe una herramienta enfocada en la

automatización de la lectura de dibujos generados a partir de técnicas proyectivas, que sirva de apoyo para los psicólogos.

Dicho sistema permite a los usuarios cargar imágenes de los dibujos en diversos formatos en una aplicación web responsiva. Utilizando un modelo de visión computacional, entrenado con más de 1,000 dibujos, el sistema detecta indicadores gráficos y los asocia con el nivel intelectual del niño o niña, siguiendo reglas establecidas en la psicología. Los indicadores se dividen en esperados y excepcionales. Los esperados son: presencia de cabeza, ojos, nariz, boca, cuerpo, piernas, brazos, cabello, pies, brazos (2 dimensiones), piernas (2 dimensiones), cuello, brazos hacia abajo y brazos unidos al hombro. Por otro lado, los excepcionales son: presencia de rodilla, perfil, codo, 2 labios, fosas nasales, brazo unido al hombro, ropa (4 ítems), cinco dedos, pupilas, pies (2 dimensiones) y proporción [5]. Adicionalmente, a través de un dashboard se facilita el análisis de los resultados con el uso de métricas y gráficas.

La hipótesis que se plantea es que la herramienta desarrollada permitirá identificar los indicadores gráficos del DFH con una precisión de más de 90%, reduciendo el tiempo de evaluación de 10 minutos a menos de 30 segundos.

El trabajo realizado se divide en las secciones que se describen a continuación. En la Sección 2, se realiza una revisión exhaustiva de trabajos similares previos, enfocándose en la interpretación de DFH y la aplicación de la Inteligencia Artificial (IA) en el campo de la psicología, identificando así las necesidades y brechas que este estudio pretende abordar. La Sección 3 describe la adopción del marco de trabajo y la gestión del proyecto. En la Sección 4 se enlistan los materiales de hardware y software necesarios para el desarrollo del sistema. La Sección 5 profundiza en los aspectos prácticos del

estudio, incluyendo la selección y preparación de los datos, el desarrollo y la configuración del modelo YOLO, así como los procedimientos seguidos para el entrenamiento y la evaluación del modelo. En la Sección 6, se presentan los resultados generados, incluyendo un análisis tanto cuantitativo como cualitativo del rendimiento del modelo. En la Sección 7, se expone una interpretación crítica de estos hallazgos, analizando sus implicaciones, las limitaciones del estudio y su relación con investigaciones anteriores. Finalmente, en la Sección 8 se resumen los hallazgos clave, destacando la importancia y el impacto potencial del estudio en la evaluación automatizada de los DFH, y delineando las futuras direcciones y mejoras propuestas para el proyecto.

2. Estado del arte

La visión computacional es un área fundamental de la inteligencia artificial dedicada a emular la capacidad de percepción visual humana. En la literatura, se han destacado múltiples tareas de visión computacional como el reconocimiento facial, la detección de objetos y la segmentación de imágenes, todas ellas beneficiándose significativamente del uso de Python. Este lenguaje de programación es ampliamente utilizado en la visión computacional debido a su simplicidad, legibilidad y versatilidad. Python cuenta con una vasta colección de bibliotecas especializadas, como OpenCV, TensorFlow y PyTorch, que facilitan la implementación y experimentación con modelos de redes neuronales convolucionales (CNN) [6]. Las CNN son especialmente valoradas por su habilidad para procesar grandes conjuntos de datos visuales con un alto grado de precisión y eficiencia.

Para un mayor entendimiento de las distintas arquitecturas de red neuronal, es importante

destacar que existen varios tipos de CNN que se especializan en diferentes aspectos de la visión computacional. Entre las arquitecturas más populares para tareas de clasificación se encuentran VGG-16, ResNets e Inception, conocidas por su profundidad y capacidad para identificar y clasificar con precisión variadas categorías de imágenes. En el campo de la detección de objetos, arquitecturas como YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector) y las variantes de R-CNN (como Fast R-CNN y Faster R-CNN) permiten no solo clasificar objetos dentro de una imagen, sino también determinar su ubicación exacta a través de cuadros delimitadores. Esto es esencial para aplicaciones que requieren una respuesta rápida y precisa, como los sistemas de vigilancia automática y la navegación autónoma de vehículos.

En cuanto a la segmentación de imágenes, técnicas como las redes completamente convolucionales (FCNs) y U-Net ofrecen soluciones robustas para dividir una imagen en segmentos que representan diferentes objetos o regiones, lo que es crucial en aplicaciones médicas para la identificación de tejidos o diagnósticos basados en imágenes.

Este panorama de las arquitecturas de CNN muestra cómo estas redes avanzan continuamente para enfrentar desafíos
















complejos en el campo de la visión computacional, proporcionando herramientas cada vez más sofisticadas y efectivas para el análisis y la interpretación visual automatizada [7].
















De acuerdo con Luxton [8], las tecnologías de inteligencia artificial están siendo ampliamente consideradas como potenciales métodos para el entrenamiento, tratamiento, evaluación psicológica y toma de decisiones clínicas en temas de salud mental. Actualmente, existen diferentes métodos de inteligencia artificial que proveen guía activa a los psicólogos para la toma de decisiones de tratamientos [9] y la detección de padecimientos a través de señales fisiológicas [10], entre otras aplicaciones.

Con el fin de desarrollar una propuesta sólida que tome en cuenta los requerimientos, que incluyen la implementación de aprendizaje automático, específicamente visión por computadora, la detección de más de 20 indicadores gráficos en dibujos, una interfaz de usuario intuitiva, capacidad de almacenamiento de datos, y un alto nivel de precisión, se llevó a cabo una evaluación comparativa de varias soluciones disponibles en el mercado. Esta evaluación se presenta detalladamente en la Tabla 1, donde se contrastan las distintas alternativas actualmente disponibles.

Tabla 1. Comparación entre servicios ofrecidos en el mercado y en la literatura científica.

Solución	Detección de indicadores gráficos	Uso de machine learning	Interfaz de usuario usable y fácil de usar	Almacenamiento en base de datos	Métrica de rendimiento (<i>accuracy</i> /ex actitud)
Sistema para la Evaluación Automatizada de DFH a través de un Modelo de Visión Computacional	●	●	●	●	●

<p>Sistema para detectar abuso sexual infantil [11]</p>	 <p>Detecta si el niño fue víctima de abuso o no.</p>	 <p>Red neuronal convolucional (CNN). Preprocesamiento de imágenes: normalización de escala y posición uniformes, conversión a formato JPG y selección de imágenes basada en tamaño y tipo de dibujo</p>	 <p>Sin interfaz</p>	 <p>No se menciona un sistema específico de almacenamiento en base de datos</p>	 <p>72%</p>
<p>Sistema para calificar pruebas de DFH [12]</p>	 <p>Detecta 9 indicadores en dibujos: presencia de cabeza, piernas, brazos, ojos, boca y cabello, si piernas y brazos están pegados al tronco, y si están pegados en el lugar correcto.</p>	 <p>Red neuronal convolucional (CNN). Preprocesamiento de imágenes: aumento de datos mediante la aplicación de desenfoque gaussiano y adición de ruido gaussiano a cada muestra, redimensionamiento de imágenes y creación de un archivo de verdad del terreno</p>	 <p>Sin interfaz.</p>	 <p>No se menciona un sistema específico de almacenamiento en base de datos.</p>	 <p>86%</p>
<p>IA que detecta emociones en dibujos de niños [13]</p>	 <p>Identifica 5 emociones en dibujos: felicidad, enojo, violencia, ansiedad y depresión.</p>	 <p>Modelo de PyTorch. Preprocesamiento de imágenes: No se menciona un algoritmo específico, pero se habla de la importancia de normalizar las imágenes de los dibujos para</p>	 <p>Sin interfaz.</p>	 <p>No se menciona un sistema específico de almacenamiento en base de datos.</p>	 <p>87%</p>

		asegurar la calidad de los resultados.			
Imomtae [14]	 <p>Reconoce patrones y detecta emociones. Se trata de un servicio de análisis psicológico sin contacto. No se especifica el tipo de diagnóstico.</p>	 <p>Deep learning (no se especifica algoritmo). La solución utiliza técnicas de análisis de imagen y aprendizaje profundo para reconocer patrones y detectar emociones.</p> <p>Preprocesamiento de imágenes: La solución utiliza un algoritmo de preprocesamiento basado en aprendizaje profundo para normalizar las imágenes de los dibujos.</p>	 <p>La solución cuenta con una interfaz de usuario diseñada para ser fácil de usar y accesible para los niños y niñas.</p>	 <p>La solución almacena los datos en una base de datos segura.</p>	 <p>Se presume una alta precisión gracias a la entrada de 7,000 dibujos.</p>
Reconocimiento de emociones en dibujos de la figura humana [15]	 <p>Se utiliza como clasificador para aprender a distinguir entre las cuatro clases de expresiones faciales: feliz, triste, enojado y neutral.</p>	 <p>Máquina de Vectores de Soporte (SVM).</p> <p>Preprocesamiento de imágenes: Binarización y Segmentación de Imágenes, Patrones Binarios Locales, Matrices de Co-ocurrencia de Niveles de Gris, Histogramas de Gradientes Orientados.</p>	 <p>Sin interfaz.</p>	 <p>Los dibujos se almacenan en una base de datos.</p>	 <p>82%</p>
Reconocimiento de cabezas en DFH [16]	 <p>Reconoce y analiza los resultados de la prueba del</p>	 <p>Red neuronal convolucional (CNN).</p>	 <p>Sin interfaz.</p>	 <p>No se menciona un sistema específico de</p>	 <p>66.7%</p>

	DFH, específicamente para detectar 4 clases: cabezas grandes, un poco grandes, normales e incompletas	Preprocesamiento de imágenes: corte de imagen, transformación de perspectiva, redimensionamiento de imagen y normalización de imagen.		almacenamiento en base de datos.	
--	---	---	--	----------------------------------	--

La primera solución se enfoca en el problema global del abuso sexual infantil (ASI). La dificultad principal en tratar el ASI es la renuencia de las víctimas a revelar el abuso y la incapacidad de los profesionales para detectarlo cuando no existen pruebas forenses. El estudio presenta el desarrollo de un sistema de Red Neuronal Convolutiva (CNN, por sus siglas en inglés) en cuatro etapas para detectar casos de abuso en dibujos de figuras realizadas por niños y niñas. Se evaluó el rendimiento del sistema en comparación con expertos humanos y se encontró que, aunque logró un 72% de *accuracy* en la detección de abuso, los expertos superaron al sistema en un 16%. Aunque los resultados preliminares indican que la CNN puede contribuir a la detección de abuso sexual infantil, se necesita un mayor desarrollo [11].

Otra solución se enfoca en utilizar la tecnología para medir la madurez cognitiva de niños a través de la prueba del DFH. Su objetivo es reemplazar la puntuación manual, siendo lenta y propensa a sesgos, por medio de un análisis computarizado de muestras digitalizadas para lograr mayor eficiencia y consistencia. Sin embargo, la variabilidad en los trazos y la falta de datos de entrenamiento presentaron desafíos. La solución contempla el desarrollo de una arquitectura de CNN con dos etapas de afinación. La CNN está diseñada para puntuar el dibujo en tres categorías: Detalles Generales, Adjuntos y Detalles de la Cabeza. Los resultados muestran que esta CNN, afinada en dos

etapas, supera el rendimiento de las CNN pre-entrenadas convencionales, al lograr una exactitud del 81.1% en detalles generales, 99.2% en adjuntos y 79.3% en detalles de la cabeza en las muestras de la prueba del Dibujo de la Figura Humana [12].

Por otro lado, existe un modelo basado en IA para detectar ciertas emociones en dibujos realizados por niños y niñas, no necesariamente de la figura humana. Tras la recopilación de un conjunto de dibujos, se partió de un modelo pre-entrenado y del uso de la biblioteca de aprendizaje automático PyTorch para crear un sistema que identifica categorías emocionales: felicidad, ansiedad y depresión, e ira y violencia. Se consiguió con este una exactitud del 87% [13].

Además de esta clase de estudios y proyectos, se encontró una solución lanzada al mercado de Corea en 2022 que comparte el mismo objetivo de ofrecer apoyo psicológico a los infantes a través de la interpretación de sus dibujos. Se trata de un servicio llamado "Imomtae", que realiza un análisis psicológico sin contacto para el arte infantil mediante IA. El servicio, que utiliza la IA para analizar dibujos infantiles y brindar apoyo emocional, está dirigido a niños y niñas de 7 a 12 años. La empresa que ofrece el servicio destaca que son pocos los servicios comerciales que utilizan IA para asesoría psicológica y resalta la necesidad de datos de dibujos infantiles para comprender mejor la psicología infantil. En este proyecto, se

procesaron más de 7000 dibujos infantiles para análisis de alta precisión [14].

Por otra parte, se propone otra tecnología que se enfoca en el análisis de imágenes faciales dibujadas a mano en pruebas psicológicas [15], bajo la metodología del DFH. El sistema desarrollado automatiza el análisis de imágenes de rostros dibujados, extrayendo expresiones faciales de la imagen. La imagen del rostro dibujado se binariza y segmenta en diferentes componentes faciales. Se extraen características de estos componentes (basadas en patrones binarios locales, matrices de co-ocurrencia de niveles de gris e histogramas de gradientes orientados) para entrenar un clasificador basado en Support Vector Machine (SVM) que aprende a distinguir entre cuatro clases de expresiones: feliz, triste, enojado y neutral. Se consiguió una tasa de clasificación correcta del 82%.

Por último, dos científicos desarrollaron una solución que utiliza aprendizaje profundo con una CNN para reconocer y analizar los resultados de la prueba del DFH. El objetivo fue detectar 4 clases: cabezas grandes, un poco grandes, normales e incompletas. El modelo entrenado logra una exactitud del 99.48% en los datos de entrenamiento, con un 1.74% de pérdida. En los datos de prueba, el modelo alcanza una exactitud del 66.7% [16].

Con base en los criterios de la Tabla 1, se puede identificar que las soluciones actuales en el mercado no cumplen con todos los requerimientos. Varias soluciones carecen de una interfaz de usuario fácil de usar, lo que podría limitar su accesibilidad y practicidad para usuarios no técnicos. Además, no se menciona el uso de un sistema específico de almacenamiento de datos en varias de ellas, lo que plantea preguntas sobre la seguridad y la gestión de los datos recopilados. Aunque algunas soluciones presentan métricas de rendimiento sólidas, con precisión reportada de hasta el 87%, todavía es posible la mejora

en la exactitud y la eficiencia. A su vez, ninguna solución abarca un rango completo de indicadores como el modelo propuesto.

3. Metodología

El diseño del estudio fue experimental para evaluar la eficacia de la herramienta en la detección temprana de problemas cognitivos por medio del dibujo realizado por los participantes. Se comenzó con una etapa de recolección de dibujos. La selección de la muestra fue un muestreo por conveniencia, realizando la aplicación de la metodología del DFH en 3 instituciones educativas por la disponibilidad y disposición para participar por parte de las autoridades administrativas de estas instancias. Se siguió un procedimiento donde se les solicitó a niños y niñas participantes de edades de entre 7 y 12 años dibujar una persona de su mismo sexo y otra del sexo opuesto, proporcionándoles una hoja de papel y un lápiz. Dicha actividad se llevó a cabo durante el horario escolar y de actividades en un entorno controlado, aplicando la misma consigna a todos los grupos de infantes.

Tras la obtención de datos, se comenzó con el desarrollo del modelo, el cual implicó las siguientes etapas: preprocesamiento de los datos (en este caso de los dibujos), etiquetado, división (en conjuntos de datos de entrenamiento, validación y prueba), entrenamiento y validación del modelo. Debido al alto número de indicadores, el proceso se dividió en 5 fases (en cada fase se consideraron de entre 4 a 5 indicadores), de manera que las etapas mencionadas se realizaron de forma iterativa, haciendo el modelo cada vez más robusto.

Al final, se probó dicho modelo en un conjunto de datos de prueba, para así evaluar su rendimiento por medio del cálculo de métricas. Se llevó a cabo una fase de optimización, en la que se experimentó con el

modelo a través de la aplicación de técnicas de aumento de datos y ajuste de hiperparámetros, con el objetivo de mantener una precisión mayor al 90% en el análisis de dibujos.

Paralelo al entrenamiento del modelo, se desarrolló el sistema necesario para la solución, que incluye una base de datos para almacenar la información relevante de estas evaluaciones (datos de los participantes, dibujos e indicadores encontrados), un backend que se encarga del procesamiento de datos, la lógica de negocios, y la interacción entre los componentes, y por último, una aplicación web altamente usable con la que el usuario puede interactuar para aprovechar el potencial de la inteligencia artificial en la evaluación de dibujos.

Para el desarrollo y gestión del software se utilizaron el marco de trabajo Scrum y la metodología Extreme Programming (XP).

Scrum es un marco de trabajo clave en la gestión de proyectos, especialmente en el desarrollo de software, basado en la adaptabilidad para afrontar problemas complejos, favoreciendo un enfoque ágil [17].

Este marco de trabajo consta de 4 eventos que facilitan la planificación, ejecución y revisión del trabajo durante un sprint:

1. **Sprint Planning:** define las tareas del sprint y el Sprint Backlog, que incluye objetivos y elementos del Product Backlog.
2. **Scrum Diario:** reuniones cortas para revisar avances y planificar el día, adaptando el plan según sea necesario.
3. **Sprint Review:** presentación de resultados y discusión de ajustes en el Product Backlog.

4. **Sprint Retrospective:** análisis del desempeño del sprint para identificar mejoras.

Además, presenta 3 entregables clave, que guían y documentan el progreso del desarrollo:

1. **Product Backlog:** lista de elementos necesarios para el desarrollo, orientada a la meta final del producto.
2. **Sprint Backlog:** incluye objetivos y elementos seleccionados para una iteración específica, guiando el trabajo y avance de los desarrolladores.
3. **Incremento:** avance hacia la meta final, con múltiples incrementos posibles en un sprint.

De igual forma, se puede emplear la gráfica de quemado, una herramienta útil para observar el desempeño del equipo durante el sprint. Cada sprint se representa mediante un diagrama, empezando con la suma total de los puntos asignados a las historias por realizar. Diariamente, se determina la cantidad de puntos completados. Al término del sprint, se anticipa que no queden puntos pendientes.

Para el desarrollo de software, se puede integrar XP junto con Scrum. XP enfatiza la calidad del software y la capacidad de respuesta a las necesidades cambiantes del cliente.

A su vez, consta de fases del ciclo de desarrollo que aseguran la entrega continua de software funcional y de alta calidad [18].

1. **Planeación:** se definen las historias de usuario y se planifican las iteraciones.
2. **Análisis:** identificación de requisitos detallados y creación de especificaciones.
3. **Diseño:** diseño simple y arquitectura del sistema.

4. Construcción: implementación del código y desarrollo de funcionalidades.
5. Pruebas: pruebas continuas para asegurar la calidad del software.

4. Materiales y equipos

El equipo requerido para el desarrollo del proyecto consistió en hardware y software descritos a continuación, destinados a tareas de desarrollo de aplicaciones web y entrenamiento de modelos de inteligencia artificial.

En términos de hardware, se emplearon computadoras con capacidad suficiente para ejecutar desarrollo web y modelos de visión computacional. Estas computadoras contaban con un procesador Intel Core i9, 32 GB de RAM, 12 GB de GPU gracias a una tarjeta gráfica NVIDIA GeForce RTX 4070, y SSD de 1 TB.

En cuanto al software, el desarrollo del modelo se realizó principalmente en Python, lenguaje utilizado ampliamente en el área de la inteligencia artificial. Se eligió el framework de aprendizaje automático PyTorch para construir y entrenar el modelo. Además, se empleó el programa labelImg, para realizar el etiquetado de los dibujos con las diferentes clases a evaluar, y OpenCV, una biblioteca de visión por computadora, para el procesamiento de imágenes. Google Colab fue utilizada como plataforma interactiva para el desarrollo y el entrenamiento del modelo en sus fases finales, permitiendo la combinación de código y documentación, así como el acceso a GPU de 16 GB en un entorno colaborativo. También se utilizó el algoritmo YOLO (You Only Look Once) para la detección de objetos mediante redes neuronales.

El desarrollo del backend se realizó en C#, utilizando el framework .NET 7. Se

integraron diversas bibliotecas como FluentValidation para definir reglas de validación, AutoMapper para la asignación automática de datos entre objetos, y EntityFramework para interactuar con bases de datos relacionales. Swagger se utilizó para documentar y generar una interfaz de usuario interactiva para las APIs web, mientras que JWT (JSON Web Token) se empleó para la creación de tokens de seguridad. Además, se utilizó Docker para la gestión de contenedores y NextCloud para la creación de servicios de alojamiento de archivos.

El desarrollo del frontend se llevó a cabo con Vue.js, un framework de JavaScript, junto con Quasar Framework para simplificar el desarrollo de aplicaciones web y móviles. Axios se utilizó para facilitar las solicitudes HTTP desde el lado del cliente. Además, se emplearon los lenguajes de programación JavaScript, HTML y CSS para agregar interactividad, estructura y estilo a las páginas web. Nginx se utilizó como servidor web/proxy inverso de alto rendimiento.

Para el control de versiones, se implementó Gitflow, un modelo de flujo de trabajo para Git que mejoró la gestión de proyectos de software. En términos de bases de datos, se utilizó SQL Server 2022, un sistema de gestión de bases de datos relacional de Microsoft. El diseño de interfaces se llevó a cabo en Figma, una plataforma de diseño para la creación de prototipos interactivos y diseños de aplicaciones y sitios web. Finalmente, para la gestión de proyectos, se utilizó ClickUp, una herramienta en línea que ayudó al equipo a planificar, colaborar y realizar un seguimiento eficiente de sus proyectos y tareas.

5. Despliegue del método

El desarrollo del sistema implicó 3 diferentes componentes: el modelo de visión

computacional, la aplicación backend y la aplicación frontend.

Para desarrollar el modelo de visión computacional, inicialmente se consideró crear una CNN por su eficacia en tareas de visión computacional, incluyendo la detección de objetos [19]. No obstante, actualmente existen modelos de detección de objetos validados, que permiten obtener resultados prometedores, por lo que se optó por usar un algoritmo que utiliza redes neuronales convolucionales para proveer detección de objetos. Existen diversas opciones de estos algoritmos como Faster R-CNN, Retina-Net, Single-Shot MultiBox Detector (SSD) y You Only Look Once (YOLO). Estas herramientas resuelven el problema de modelado y limitación de datos. Sin embargo, la principal desventaja de las tres primeras es que no permiten detectar objetos en una sola ejecución del algoritmo. El algoritmo que destaca es YOLO porque ha demostrado tener un desempeño superior en la detección de objetos. Una de las principales ventajas de YOLO radica en su enfoque de detección de objetos en una sola pasada a

través de la red neuronal y por su velocidad en la detección sin sacrificar significativamente la precisión [20]. Generalmente, YOLO se usa para la detección de objetos en fotografías y no en dibujos. Sin embargo, se han realizado propuestas donde emplean YOLO para la clasificación de objetos usando dibujos [21, 22]. Dicho esto, se eligió utilizar YOLO v8 y su implementación en PyTorch, conocida por sus mejoras en eficacia, adaptabilidad y versatilidad.

YOLO utiliza una arquitectura de CNN, la cual se divide en dos partes principales: la columna vertebral (backbone) y la cabeza (head) (Ver Figura 1). La columna vertebral utiliza una versión modificada de CSPDarknet53, que consiste en 53 capas convolucionales y emplea conexiones parciales entre etapas para mejorar el flujo de información entre las diferentes capas de la red. La cabeza consta de múltiples capas de convolución seguidas de capas totalmente conectadas, responsables de predecir las cajas delimitadoras, las puntuaciones de objetividad y las probabilidades de clase [23].

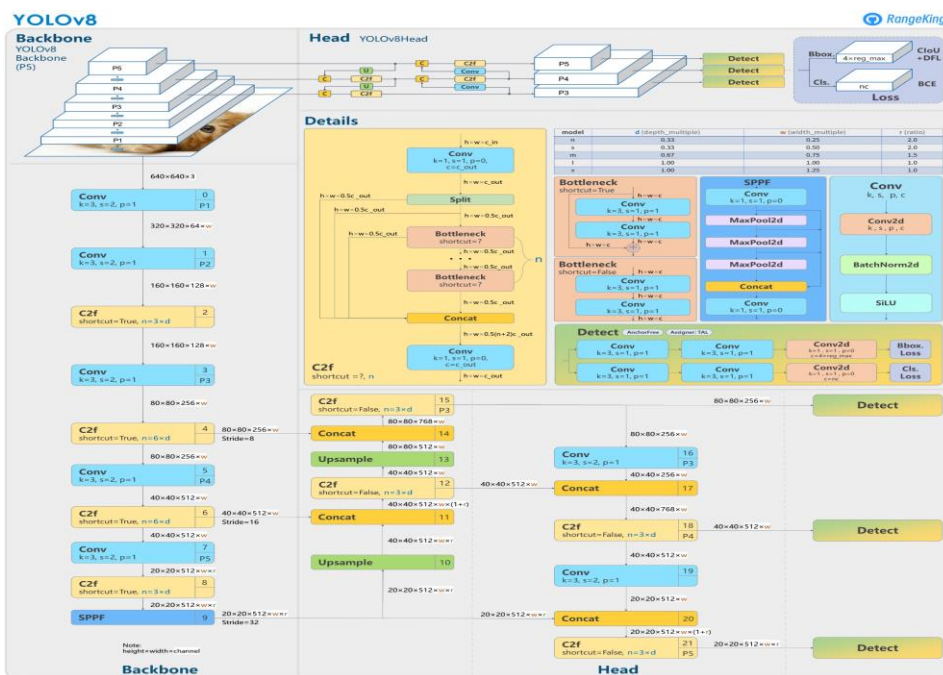


Figura 1. Arquitectura de YOLOv8 [24].

Para la evaluación de los modelos se utilizaron métricas estándar en la detección de objetos con un rango de 0 a 1 [25]. Estas incluyen:

- **Precisión (Precision):** mide la exactitud de las predicciones positivas del modelo.
- **Sensibilidad (Recall):** mide la proporción de instancias positivas reales que el modelo identifica correctamente.
- **Mean Average Precision (mAP):** calcula la precisión promedio para cada clase y toma la media de todas las clases, considerando tanto precisión como sensibilidad. Se empleó mAP50 para este proyecto, lo que significa medir la precisión media promedio con un umbral de Intersección sobre Unión (IoU) de 0.5.

Para los datos de entrada, se recopilaron un total de 1,214 dibujos realizados por niños y niñas de 5 a 12 años, provenientes de diversas escuelas y centros comunitarios ubicados en la Zona Metropolitana de Nuevo León, México.

Se etiquetaron los dibujos con labelImg. Cada indicador gráfico en las imágenes representa una clase específica (Ver Figura 2).

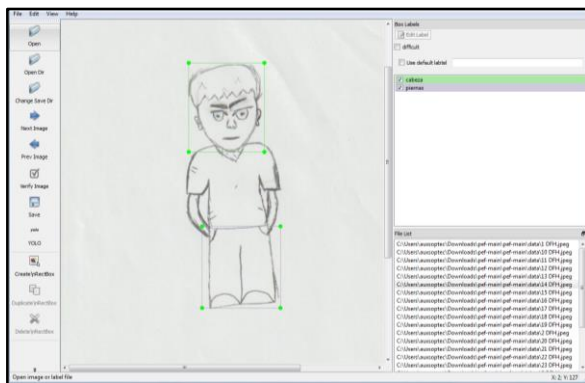


Figura 2. Etiquetado de dibujos.

Después de etiquetar las imágenes, se generan archivos .txt que contienen las clases marcadas y sus coordenadas en la imagen.

Antes de procesar los dibujos para su uso en el modelo, se llevó a cabo una etapa de preprocesamiento (Ver Figura 3) para garantizar minimizar cualquier diferencia que pudiera afectar al modelo. Este preprocesamiento incluye escala de grises, asegurar formatos correctos (JPG, JPEG o PNG), orientación vertical y técnicas de aumento de datos para compensar la limitación de dibujos de entrada: voltear horizontalmente, rotaciones entre -8° y $+8^\circ$, cambiar la exposición entre -5% y $+5\%$ y un desenfoque de hasta 0.25px. Estas técnicas sirven para aumentar artificialmente la variabilidad del conjunto de datos de entrenamiento mediante la generación de imágenes alteradas [26]. Después de la etapa de procesamiento, se llegó a la cantidad de 5,257 imágenes en total.

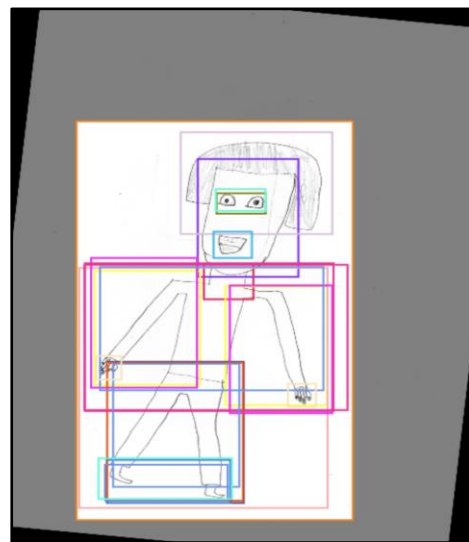


Figura 3. Imagen preprocesada.

Además, YOLO como otros modelos de inteligencia artificial requieren de la división del conjunto de datos en datos de entrenamiento, validación y prueba [27]. Dicho esto, se realizó una división de datos en la siguiente proporción: 80% para

entrenamiento, 10% para validación y 10% para pruebas.

Para entrenar el modelo YOLOv8, se creó primeramente un archivo data.yaml que contiene detalles como las ubicaciones de los datos de entrenamiento y validación, y las clases involucradas. El entrenamiento se llevó a cabo usando un comando específico en PyTorch, configurando YOLO para detectar objetos. Este proceso incluyó definir la arquitectura del modelo que es conocida por su alta precisión y numerosas capas, así como especificar otros parámetros como el número de épocas (60) y el tamaño de las imágenes (1024x1024 píxeles). La etapa de entrenamiento tuvo una duración de entre 1 a 4 horas y se aceleró significativamente usando una GPU con alta capacidad de memoria, como la proporcionada por Google Colab.

Durante el entrenamiento, se monitorearon varias métricas como la pérdida de los cuadros delimitadores (box_loss), la pérdida de clasificación (cls_loss), la pérdida de deformación (dfl_loss), la precisión (P), el recall (R), y el mean average precision (mAP50). Al final del entrenamiento, se obtuvo un archivo best.pt, el cual contiene los pesos del modelo con el mejor rendimiento en el conjunto de validación, indicando que esta versión del modelo es la más adecuada para inferencias o predicciones futuras.

El desarrollo del frontend y backend se llevó a cabo de forma paralela, utilizando la metodología de XP, asegurando una integración iterativa de los entregables generados hasta obtener el sistema final.

6. Resultados

El proyecto tuvo un periodo de duración de 16 semanas para cumplir con los requerimientos definidos.

Desarrollo de Product Backlog

A partir del objetivo del proyecto, se construyeron las historias de usuario, descripciones que proporcionaron una base sólida para definir y priorizar el trabajo a realizar en cada iteración y, en conjunto, conforman el Product Backlog. Las historias de usuario (HU) ayudan a capturar los requisitos y objetivos del usuario en un formato fácil de entender. Además, representan un entregable de la programación extrema (XP) en su primera fase. Se determinó tanto la prioridad como los requisitos específicos de cada historia, y posteriormente, el equipo de desarrollo asignó una puntuación a cada una, basándose en su complejidad y el tiempo estimado para su realización. Las HU generadas, en orden de prioridad, se enlistan a continuación:

1. Modelo de visión computacional para reconocimiento de indicadores gráficos.
 - a. Fase 1: indicadores esperados (cabeza, piernas, brazos, pies, cuello).
 - b. Fase 2: indicadores esperados (ojos, nariz, boca, cabello, cuerpo).
 - c. Fase 3: indicadores esperados (brazos 2 dimensiones, piernas 2 dimensiones, brazos hacia abajo, brazos unidos al hombro).
 - d. Fase 4: indicadores excepcionales (rodilla, perfil, codo, dos labios, fosas nasales).
 - e. Fase 5: indicadores excepcionales (4 ítems de ropa, cinco dedos, pupilas, pies 2 dimensiones, proporción).
 - f. Fase X: optimización del modelo.
2. Carga de imagen del Dibujo de la Figura Humana (DFH).

3. Visualización de resultados de la evaluación del dibujo en una interfaz gráfica.
4. Control de acceso.
5. Visualización de dashboard con resumen de la información.
6. Exportación de resultados.
7. Seguridad de datos.

Con ello se estableció el product backlog, reflejando tanto los objetivos como su importancia relativa. Además, se identificaron y planificaron las tareas necesarias para completar cada historia, resultando en la creación de las correspondientes tarjetas de tarea.

Antes de comenzar cada sprint, se eligieron las historias de usuario a desarrollar conformando el sprint backlog. Cada una de estas iteraciones incluyó fases de análisis, diseño, construcción y pruebas. A continuación, se presentan los resultados generados por sprint:

I. Sprint 1

Como parte del Sprint Planning, se definieron las HU a completar:

- Modelo de visión computacional para reconocimiento de indicadores gráficos (Fase 1).

A. Análisis

- Recolección de muestras de datos.
- Estudio de modelos existentes y librerías de visión computacional.
- Definición de los parámetros de entrada/salida y métricas para evaluación del modelo de visión computacional.

B. Diseño

- Arquitectura de software (Ver Figura 4).

Esta se basa en un enfoque progresivo y escalable y utiliza la arquitectura MVC para la interfaz web. A su vez, integra una

estructura basada en componentes, facilitando la segmentación y el abordaje concurrente y colaborativo de distintos aspectos del proyecto.

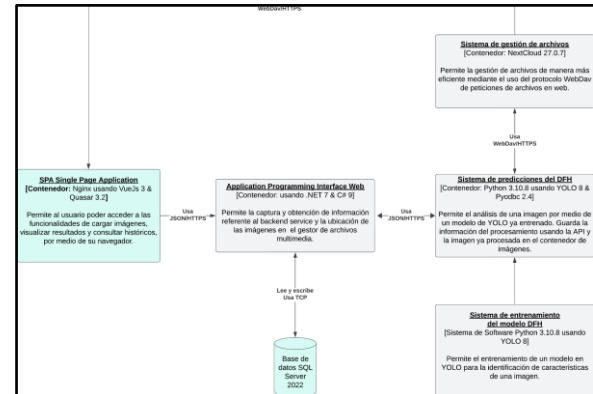


Figura 4. Arquitectura de software por componentes.

La Single Page Application (SPA) interactúa con la API Web mediante JSON/HTTPS para el intercambio de información. La API Web gestiona imágenes y archivos multimedia a través de WebDav/HTTPS y analiza imágenes con el sistema de predicciones DFH usando JSON/HTTPS. También se conecta con la base de datos SQL Server para leer y escribir datos.

- Estándares de diseño.
- Diseño de la interfaz de usuario.
 - Inicio de sesión (Ver Figura 5).

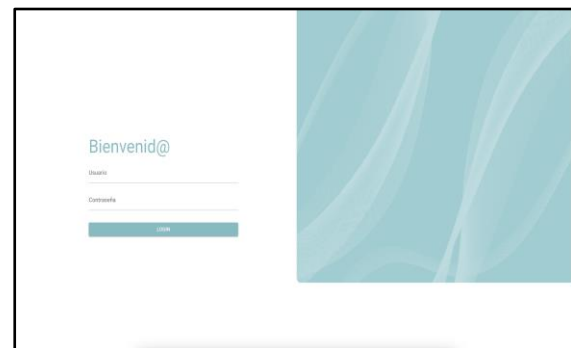


Figura 5. Pantalla de inicio de sesión.

- Carga de dibujo (Ver Figura 6).

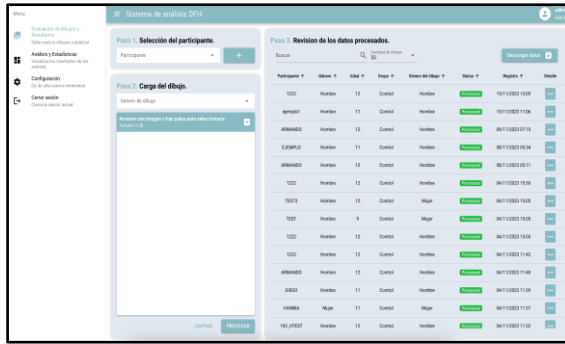


Figura 6. Pantalla de carga de imagen.

- Resultados del participante (Ver Figura 7).

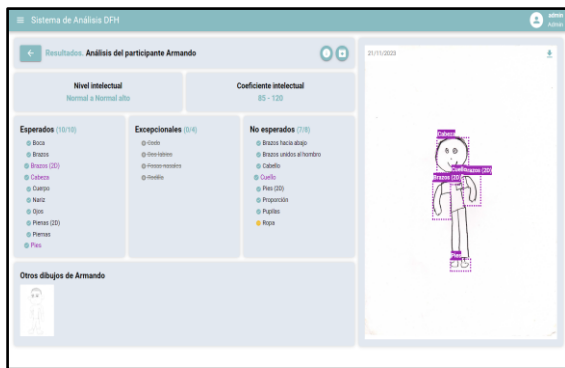


Figura 7. Pantalla de resultados.

- Dashboard estadístico (Ver Figura 8).



Figura 8. Pantalla de dashboard.

- Modelo entidad-relación
- **C. Construcción**
- Estándares de programación
- Preparación de ambiente

- Instalación de librerías de visión computacional
- Desarrollo del modelo de visión computacional (Fase 1)
 - Preprocesamiento de los dibujos
 - Etiquetado de los dibujos
 - Separación en datos de entrenamiento, validación y prueba
 - Entrenamiento del modelo

D. Pruebas

- Validación del modelo

Cabe agregar que a lo largo de los sprints se realizó el Scrum Diario para monitorear el desempeño del equipo. Adicionalmente, cada sprint finalizó con los eventos de Sprint Review y Sprint Retrospective, para presentar cada incremento e identificar áreas de mejora para los sprints posteriores.

II. Sprint 2

Como parte del Sprint Planning, se definieron las HU a completar:

- Carga de imagen del Dibujo de la Figura Humana (DFH)
- Modelo de visión computacional para reconocimiento de indicadores gráficos (Fase 2 y 3)
- Control de acceso

A. Análisis

- Recolección de más datos

B. Diseño

- Diagramas de proceso de las funcionalidades a implementar en el sitio web

C. Construcción

- Creación de la base de datos
- Creación de aplicación backend
- Creación de aplicación web base
- Programación de la carga de imagen
- Programación del inicio de sesión
- Desarrollo del modelo de visión computacional (Fase 2 y 3)

D. Pruebas

- Validación del modelo

- Pruebas unitarias
- Pruebas de integración

III. Sprint 3

Como parte del Sprint Planning, se definieron las HU a completar:

- Modelo de visión computacional para reconocimiento de indicadores gráficos (Fase 4 y 5)
- Visualización de resultados de la evaluación del dibujo en una interfaz gráfica
- Exportación de resultados
- Visualización de dashboard con resumen de la información
- Seguridad de datos

A. Análisis

- Limitaciones del modelo

B. Diseño

- Diagramas de proceso de las funcionalidades a implementar en el sitio web

C. Construcción

- Programación de la visualización de resultados de la evaluación del dibujo.

Se desarrolló un algoritmo para hacer la relación de los indicadores esperados y excepcionales con el puntaje estableciendo la categoría diagnóstica y el coeficiente intelectual, reflejando los resultados en el sistema (Ver Figura 9).



Figura 9. Pantalla de resultados.

- Programación de la exportación de resultados
- Programación del dashboard
- Implementación de seguridad
- Desarrollo del modelo de visión computacional (Fase 4 y 5)

D. Pruebas

- Validación del modelo
- Pruebas unitarias
- Pruebas de integración
- Prueba de calidad del código con SonarQube.

IV. Sprint 4

Este último sprint estuvo dedicado principalmente a pruebas del sistema y sus respectivas mejoras.

A. Análisis

B. Ajuste de hiperparámetros del modelo de visión computacional

- Diseño de prueba de usabilidad

Se plantearon escenarios y tareas a completar por usuarios potenciales, con el objetivo de evaluar la usabilidad del sitio.

- Arquetipos para pruebas de usabilidad

Se diseñaron arquetipos de usuarios basados en la población objetivo. Esto aseguró que el sistema fuera evaluado por una muestra representativa, tal como se muestra en la Tabla 2.

Tabla 2. Tabla demográfica de participantes.

Ocupación	Género	Rango de Edad	Cantidad
Psicólogos	Mujer	30-50 años	2
	Mujer	50+ años	1
	Hombre	18-29	1
Profesores	Mujer	50+ años	3
	Mujer	30-50 años	1
	Hombre	30-50 años	1
	Hombre	50+ años	1
Estudiantes	Mujer	18-29 años	3
	Hombre	18-29 años	2

C. Construcción

- Optimización del modelo

D. Pruebas

- Prueba de sistema
- Prueba de usabilidad

Los usuarios completaron las tareas de manera eficiente y encontraron la interfaz intuitiva y el diseño adecuado. La organización del sitio permitió una fácil localización de información y se destacó la rapidez del proceso de evaluación y la utilidad del sistema para analizar datos. A pesar de ello, se hicieron ajustes basados en la retroalimentación recibida, como aumentar la compatibilidad con dispositivos y la legibilidad del contenido, simplificar la

interfaz y mejorar la visibilidad de elementos clave. También se refinó la presentación de datos y se clarificó la navegación, todo para asegurar un sistema eficiente y fácil de usar. Después de finalizar estos ajustes, se desplegó el sistema.

- Prueba final del modelo

En la etapa de pruebas del modelo, se llevaron a cabo 60 épocas del modelo, las cuales se completaron en 2.661 horas utilizando el GPU de Google Colab. El modelo final cuenta con 268 capas, 68,146,680 parámetros y 0 gradientes. Un aspecto destacable es que el tiempo promedio para realizar la detección en dibujos de prueba fue de menos de un segundo. Los resultados de la validación se presentan en la Tabla 3.

Tabla 3. Métricas de validación obtenidas en la optimización del modelo.

Clase	Imágenes	Instancias	P	R	mAP50
Todas	215	4088	0.912	0.902	0.928
cabeza	215	222	0.965	1	0.993
piernas	215	222	0.948	0.982	0.985
brazo	215	436	0.963	0.956	0.975
cuello	215	137	0.949	0.942	0.958
pies	215	179	0.938	0.933	0.944
ojos	215	204	0.943	0.971	0.969
nariz	215	90	0.942	0.9	0.945
boca	215	200	0.944	0.93	0.953
cuerpo	215	224	0.978	1	0.995
cabello	215	203	0.929	0.961	0.964
brazo2d	215	373	0.966	0.979	0.983
piernas2d	215	184	0.938	0.984	0.984
brazosabajo	215	153	0.954	0.941	0.955
brazoshombro	215	128	0.845	0.938	0.931
rodilla	215	9	0.818	1	0.995
perfil	215	2	1	0.5	0.75
codo	215	20	0.9	0.9	0.942
doslabios	215	12	0.833	0.833	0.888
fosasnasales	215	6	0.8	0.667	0.766
itemropa	215	550	0.897	0.9	0.925
cincodedos	215	93	0.768	0.785	0.807
pupilas	215	125	0.896	0.896	0.885
pies2d	215	174	0.96	0.96	0.97
proporcion	215	142	0.818	0.789	0.811

El modelo fue evaluado en un conjunto de validación de 215 imágenes y ha detectado un

total de 4088 instancias de objetos. Este modelo demuestra un desempeño

sobresaliente, evidenciado por una precisión (P) de 0.912 y una sensibilidad (R) de 0.902. Estos resultados implican que el modelo logra identificar correctamente más del 90% de las instancias de los objetos presentes y es capaz de detectar aproximadamente el 90% de los indicadores relevantes dentro del conjunto de imágenes evaluadas. Es capaz de identificar con éxito clases como cabeza, piernas, brazos, cuello, pies, ojos, boca, cuerpo, cabello, entre otros, con un tiempo de detección promedio de 77.4 milisegundos.

Además, se observa una eficacia notable en la detección de clases específicas como "cabeza", "piernas", "brazo" y "cuerpo". Las métricas elevadas asociadas a estas categorías reflejan la habilidad del modelo para identificar con precisión dichas partes del cuerpo, debido a su definición clara y

prevalencia significativa en el conjunto de datos.

Por otra parte, las clases que cuentan con un número reducido de instancias, como "fosas nasales", "perfil" y "dos labios", registraron un rendimiento menor en comparación con las demás. Esto resalta la potencial necesidad de enriquecer el conjunto de datos con una mayor diversidad y cantidad de ejemplos en estas categorías específicas para mejorar la capacidad predictiva del modelo.

En las Figuras 10, 11 y 12, se muestra la detección de indicadores por el modelo final utilizando una serie de nuevos dibujos con el objetivo de evaluar su precisión y eficacia en la detección de una amplia gama de indicadores gráficos.

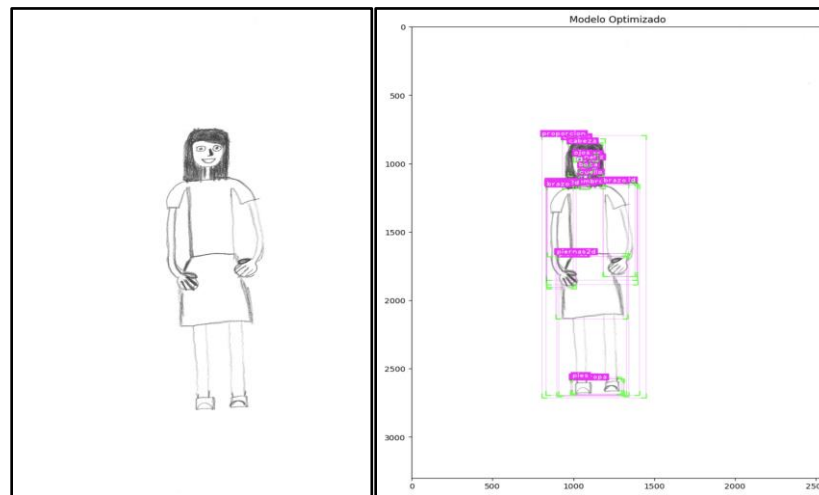


Figura 10. Prueba 1 del modelo final.

El modelo identificó con éxito una variedad de clases, incluyendo boca, brazos, brazo2ds, brazosabajo, brazoshombro, cabello, cabeza, cuello, cuerpo, itemropas, nariz, ojos, piernas, piernas2d, pies, pies2d, proporción y pupilas. El único indicador faltante fueron los cinco

dedos. Sin embargo, en el dibujo la cantidad de dedos de una mano es superior a 5 y en la otra no se puede apreciar con claridad la división de dedos, por lo que aspectos como este podrían ser desafiantes para el modelo.

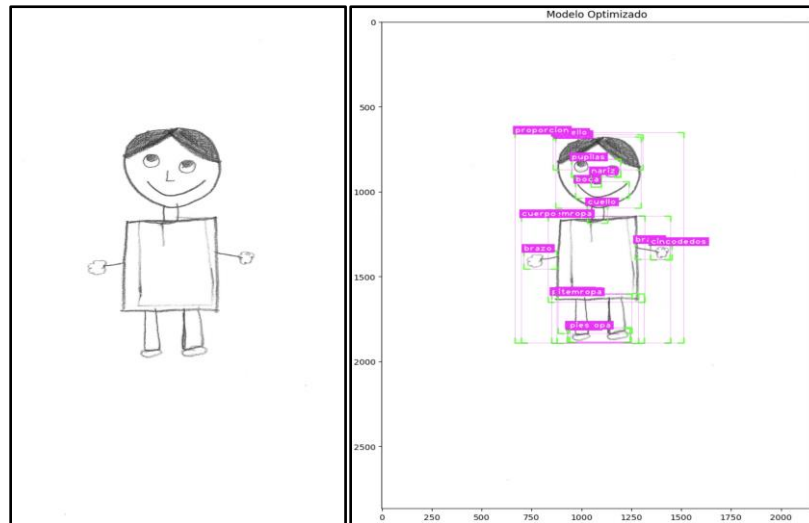


Figura 11. Prueba 2 del modelo final.

En la segunda prueba, el modelo logró identificar todas las clases presentes en el dibujo sin omitir ningún indicador,

demonstrando una notable mejora y capacidad de adaptación del modelo.



Figura 12. Prueba 3 del modelo final.

El modelo continuó mostrando un rendimiento sobresaliente, detectando todas las clases sin errores. Un detalle notable es la capacidad del modelo para detectar la 'nariz', a pesar de su pequeño tamaño y sutileza en el dibujo.

se muestra un dibujo con todos los 24 indicadores, con excepción del perfil.

Posteriormente, se probó el modelo con dibujos de prueba adicionales. En la Figura 13

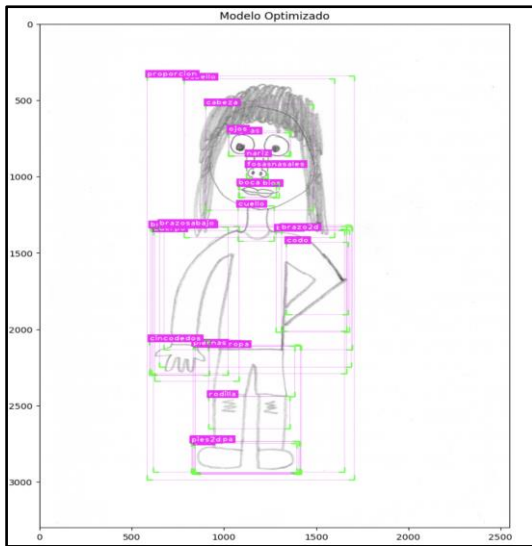


Figura 13. Prueba 4 del modelo final.

En esta prueba, el modelo fue capaz de identificar correctamente todos los indicadores gráficos presentes en un dibujo complejo, incluyendo aquellos con menos representación en las fases anteriores, como 'doslabios', 'fosasnasales' y 'rodilla'.

En la Figura 14 se probó con un dibujo más simple.

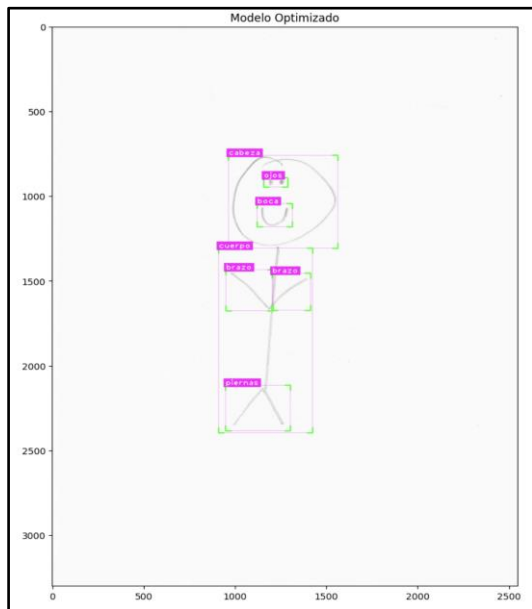


Figura 14. Prueba 5 del modelo final.

Con esta prueba, el modelo demostró su versatilidad al detectar correctamente los

indicadores gráficos presentes, lo que resalta su capacidad para adaptarse a diferentes niveles de complejidad y detalle.

7. Discusión de resultados

En comparación con versiones anteriores y otros modelos en el campo [11, 12, 13, 14, 15, 16], el modelo YOLO sobresale en su rapidez y en la amplitud de clases que puede identificar con precisión. Los modelos existentes reportaban una exactitud del 86% para la identificación de 9 indicadores de DFH [12], 87% para la detección de 4 expresiones faciales [15], y 66.7% para la clasificación de tamaños de cabeza [16]. El modelo presentado en este documento superó estas cifras, alcanzando un valor de precisión de 91.2% y un valor de mAP50 de 92.8%, demostrando una superioridad significativa. Además, es el primer modelo que es capaz de identificar hasta 24 indicadores gráficos. Otra diferencia significativa es que el sistema actual no solo ofrece un modelo de IA, sino una aplicación web que facilita la manipulación y la visualización de los resultados de la evaluación de DFH.

Es relevante mencionar que el modelo YOLO, basado en redes neuronales convolucionales, no se había aplicado antes al análisis de DFH. Si bien se encontraron algunas soluciones prometedoras que utilizaban YOLO para identificar objetos en dibujos, el enfoque de este proyecto fue distinto, ya que se centró en la detección de partes del cuerpo.

La etapa de optimización del modelo fue crucial para alcanzar estos resultados. Se implementaron técnicas de aumento de datos como el flip horizontal, rotación, variaciones en la exposición y desenfoque, incrementando el conjunto de datos de 1,214 a 5,257 imágenes de entrenamiento. Además, se realizaron cambios en los hiperparámetros del modelo, como el aumento de épocas, el

tamaño de la imagen y la confianza, basados en experimentos previos, lo que contribuyó significativamente a la mejora del desempeño del modelo.

A pesar de que se lograron los objetivos planteados, existen áreas significativas de mejora. Específicamente, el modelo necesita un conjunto de datos más robusto y diverso para clases menos representadas como "rodilla", "perfil" y "fosas nasales". Este enriquecimiento del conjunto de datos podría mejorar la capacidad del modelo para manejar variaciones sutiles y detalles finos.

Un desafío notable del modelo es su rendimiento en la detección de detalles pequeños o finos. La implementación de técnicas de aumento de datos y algoritmos de procesamiento de imágenes especializados podría ayudar a mejorar la precisión en estas áreas.

En futuras mejoras del sistema, se planea mejorar el desempeño del modelo de inteligencia artificial aumentando su precisión utilizando un mayor conjunto de dibujos de entrenamiento para las clases sub-representadas como el perfil, fosas nasales y rodillas. Además, se propone enriquecer el modelo con indicadores gráficos adicionales que detecten emociones y problemáticas en niños, como el abuso, y expandir su uso a otras técnicas proyectivas, como el test de la persona bajo la lluvia, para realizar diagnósticos psicológicos más amplios. También se contempla adaptar y validar el sistema para diferentes contextos culturales, trabajando con expertos internacionales y usando un conjunto más variado de datos, para hacer el modelo más inclusivo y global.

8. Conclusiones

Este proyecto ha culminado en el desarrollo de un sistema automatizado para la evaluación del DFH, representando un

avance significativo en la intersección de la psicología y la tecnología. Se obtuvo un modelo entrenado con 1,214 dibujos, que es capaz de detectar hasta 24 indicadores gráficos en el DFH con una precisión de (P) de 0.912, una sensibilidad (R) de 0.902 y un mAP50 (Mean Average Precision a IoU 0.5) de 0.928; lo cual significa que el modelo tiene un rendimiento promedio del 92.8% en la precisión de detección de objetos en el rango de indicadores. Además, se consiguió reducción del tiempo de la tarea de evaluación de dibujos en un 99.83%, aceptando así la hipótesis planteada.

La mayoría de los indicadores gráficos poseen un mAP50 mayor a 0.9 (Ver Tabla 3). Clases como "cabeza", "piernas", "brazo" y "cuerpo", presentan la precisión y sensibilidad más altas, con un mAP50 superior a 0.975, lo que sugiere que el modelo es particularmente hábil para detectar y reconocer estos elementos en imágenes. Esto puede deberse a la baja variabilidad de estos indicadores en los dibujos, pero principalmente al alto número de instancias de estos elementos en los dibujos recopilados, al ser indicadores esperados. Otras clases, como "fosas nasales", "perfil" y "dos labios", que forman parte de los indicadores excepcionales, tienen métricas de rendimiento más bajas, con un mAP50 de entre 0.750 y 0.888. La principal razón de este comportamiento es el bajo número de instancias para estas clases. Una línea de trabajo futura consiste en aumentar el número de instancias de dichos indicadores, con el objetivo de mejorar la precisión del modelo.

La automatización de la evaluación psicológica presenta una serie de beneficios. Primeramente, se ha logrado reducir el tiempo requerido para evaluar cada dibujo, pasando de 10 minutos a menos de 30 segundos, permitiendo a los profesionales enfocarse en análisis más complejos. Además, al emplear técnicas de inteligencia

artificial, se consigue una mayor precisión en la detección de patrones y características en los dibujos, lo que conduce a diagnósticos más certeros. Esta precisión mejorada también contribuye a minimizar los errores humanos, los cuales pueden ser comunes en evaluaciones manuales debido a factores como la fatiga o el sesgo. Por último, el acceso histórico a datos proporciona una valiosa información para el seguimiento de tendencias a lo largo del tiempo y la mejora continua del proceso de evaluación.

En términos de impacto social, la herramienta es crucial para la detección temprana de problemas de desarrollo, contribuyendo a una mejor calidad en la educación y los servicios de salud psicológica. La protección de datos y la ética han sido prioridades, asegurando la privacidad y seguridad de la información mediante protocolos robustos. Profesionales de la psicología y la educación se benefician de un método más eficiente y objetivo para los diagnósticos, facilitando intervenciones tempranas y precisas en el desarrollo emocional y cognitivo de los niños. Para los padres y tutores, proporciona una comprensión más clara de la situación emocional y cognitiva de sus hijos, mejorando el apoyo en el hogar. Instituciones educativas y de salud pueden realizar evaluaciones más eficientes y precisas, mejorando sus servicios. Además, el proyecto abre nuevas áreas de innovación para empresas de tecnología y desarrolladores de software.

La viabilidad a largo plazo del proyecto se evidencia en su diseño pensado para integrarse de manera sencilla con sistemas ya existentes, así como en su capacidad para actualizaciones y mantenimiento continuo.

En resumen, este proyecto no solo es un logro tecnológico, sino también un compromiso con el bienestar social y la ética, alineándose con objetivos globales de salud y bienestar,

educación de calidad, e industria e innovación. El software ha sido debidamente registrado ante el instituto nacional del derecho de autor (INDAUTOR) bajo el número 03-2024-042413223900-01. Su implementación promete ser un cambio transformador en la salud mental y el desarrollo cognitivo infantil, abriendo un camino hacia un futuro más inclusivo y empático.

9. Agradecimientos

Agradecemos a la Dra. Angélica Quiroga quien con su conocimiento en la técnica DFH estableció los requerimientos para esta tecnología desarrollada, validó el desarrollo y los resultados logrados con las pruebas realizadas. También se agradece a la escuela Simón Bolívar II cuya directora es la profesora María de los Ángeles Espinoza Márquez, y a la Dra. Laura Zúñiga Valero del Centro Comunitario Nuevo León San Bernabé y Centro Comunitario Nuevo León La Alianza, quienes dieron acceso para aplicar la técnica del dibujo de la figura humana y de esa manera lograr recabar los dibujos que fueron indispensables para el entrenamiento del modelo.

10. Referencias

- [1] Gaytán, E., Rosales, M., Reyes, H., Díaz-Barriga, F., & Calderón, J. (2015). Prevalencia de dificultades emocionales, conductuales y cognitivas en niños de escenarios urbanos con diferente grado de marginación. *Revista de psicología y ciencias del comportamiento de la Unidad Académica de Ciencias Jurídicas y Sociales*, 6(1), 57-74. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-18332015000100004&lng=es&tlng=es.
- [2] Delahunty, L., O'Hare, A., Marryat, L., Stewart, T. M., McKenzie, K., Murray, G., & Kaza, N. (2022). Short Report: Exploring the extent to which Intellectual Disability is undiagnosed within children attending

developmental pediatric clinics. *Research in Developmental Disabilities*, 131, 104359. <https://doi.org/10.1016/j.ridd.2022.104359>

[3] Chollat, C., Joly, A., Houivet, E., Bénichou, J., & Marret, S. (2019). School-age human figure drawings by very preterm infants: Validity of the Draw-a-Man test to detect behavioral and cognitive disorders. *Archives de Pédiatrie*, 26(4), 220-225. <https://doi.org/10.1016/j.arcped.2019.02.015>

[4] Dağlioğlu, H. E., Deniz, Ü., & Kan, A. (2010). A study on the emotional indicators in 5-6 year-old girls' and boys' human figure drawings. *Procedia - Social and Behavioral Sciences*, 2(2), 1503-1510. <https://doi.org/10.1016/j.sbspro.2010.03.226>

[5] Quiroga, A., Fuentes, M., Saucedo, E., Briones, E. N., Lozano, A. P., Montegudo, D., & González, N. (2022). *Sistematización de la Evaluación Psicológica de Recursos Personales a través del Dibujo de la Persona*. Universidad de Monterrey, Vicerrectoría de Ciencias de la Salud, Escuela de Psicología. Programa de Evaluación Final del Programa Académico de Licenciatura en Psicología/Licenciatura en Psicología Clínica. San Pedro Garza García, Nuevo León.

[6] Raschka, S., & Mirjalili, V. (2017). *Python Machine Learning* (2nd ed.). Packt Publishing.

[7] Taye, M. M. (2023). Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation*, 11(3), 52. <https://doi.org/10.3390/computation11030052>

[8] Luxton, D. D. (2014). Artificial intelligence in psychological practice: Current and future applications and implications. *Professional Psychology: Research and Practice*, 45(5), 332-339. <https://doi.org/10.1037/a0034559>

[9] Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism*, 69(Supplement), 36-40. <https://doi.org/10.1016/j.metabol.2017.01.011>

[10] Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine*, 161, 1-13. <https://doi.org/10.1016/j.cmpb.2018.04.005>

[11] Kissos, L., Goldner, L., Butman, M., Eliyahu, N., & Lev-Wiesel, R. (2020). Can artificial intelligence achieve human-level performance? A pilot study of childhood sexual abuse detection in self-figure drawings. *Child Abuse & Neglect*, 109, 104755. <https://doi.org/10.1016/j.chiabu.2020.104755>

[12] Zeeshan, M.O., Siddiqi, I., & Moetesum, M. (2021). Two-Step Fine-Tuned Convolutional Neural Networks for Multi-label Classification of Children's Drawings. In J. Lladós, D. Lopresti, & S. Uchida (Eds.), *Document Analysis and Recognition – ICDAR 2021*, 12822, 321-334. Springer, Cham. https://doi.org/10.1007/978-3-030-86331-9_21

[13] Romero, V. (2019). *How Artificial Intelligence can detect emotions in children's drawings*. <https://medium.datadriveninvestor.com/how-artificial-intelligence-can-detect-emotions-in-childrens-drawings-4359cf51ab3d>

[14] Taeuck, L. (2022). *Understanding Children's Psychology with Drawings Drawn by Children Using AI*. <https://www.24-7pressrelease.com/press-release/497168/korea-contact-free-insigher->

understanding-childrens-psychology-with-drawings-drawn-by-children-using-ai

[15] Moetesum, M., Aslam, T., Saeed, H., Siddiqi, I., & Masroor, U. (2017). Sketch-based Facial Expression Recognition for Human Figure Drawing Psychological Test. In *2017 International Conference on Frontiers of Information Technology (FIT)*, 258-263. Islamabad, Pakistan. <https://doi.org/10.1109/FIT.2017.00053>

[16] Widiyanto, S., & Abuhasan, J. W. (2020). Implementation The Convolutional Neural Network Method For Classification The Draw-A-Person Test. In *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 1-6. Gorontalo, Indonesia. <https://doi.org/10.1109/ICIC50835.2020.9288651>

[17] Schwaber, K. & Sutherland, J. (2020). *The 2020 Scrum Guide*. <https://scrumguides.org/scrum-guide.html>

[18] Beck, K., & Andres, C. (2004). *Extreme Programming Explained: Embrace Change* (2nd ed.). Addison-Wesley Professional.

[19] Terra, J. (2023). *Keras vs Tensorflow vs Pytorch: Key Differences Among Deep Learning*. <https://www.simplilearn.com/keras-vs-tensorflow-vs-pytorch-article>

[20] Boesch, G. (2023). *Object Detection in 2023: The Definitive Guide*. <https://viso.ai/deep-learning/object-detection/>

[21] Elyan, E., Jamieson, L. and Ali-Gombe, A. (2020). Deep learning for symbols detection and classification in engineering drawings. *Neural networks*, 129, 91-102. <https://doi.org/10.1016/j.neunet.2020.05.025>

[22] Zhao, Y. & Deng, X., & Huahui, L. (2020). A YOLO-Based Method to Recognize Structural Components from 2D

Drawings. In *Construction Research Congress 2020: Computer Application*, 753-762.

<https://doi.org/10.1061/9780784482865.080>

[23] Schcolnik-Elias, A., Martinez, S., Luna-Taylor, J., & Castro, I. (2023). Detección de armas tipo pistola mediante el uso de redes convolucionales con una arquitectura tipo YOLO y estereoscopia. *Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI*, 11(2), 196-204. <https://doi.org/10.29057/icbi.v11iEspecial2.10727>

[24] King, R. (2023). *Brief summary of YOLOv8 model structure*. <https://github.com/ultralytics/ultralytics/issues/189>

[25] Casas, E., Ramos, L., Bendek, E., & Rivas-Echeverría, F. (2023). Assessing the Effectiveness of YOLO Architectures for Smoke and Wildfire Detection. *IEEE Access*, 11, 96554-96583. <https://doi.org/10.1109/ACCESS.2023.3312217>

[26] Shorten, C. & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(60). <https://doi.org/10.1186/s40537-019-0197-0>

[27] Kearns, M. & Nenkova, A. (2019). *Training, Validation and Test sets*. University of Pennsylvania. https://www.cis.upenn.edu/~mkearns/teaching/ScienceDataEthics/files/lecture_15.pdf