



Impacto de la IA en la Predicción y Detección Oportuna de Enfermedades: Una Revisión Sistemática

Impact of AI on Disease Prediction and Early Detection: A Systematic Review

Carlos Zepeda-Lugo^{a, *}, Andrea Insfran-Rivarola^b, Ana Arévalos^b

^aInstituto de Servicios de Salud Pública del Estado de Baja California, Departamento de Conservación y Mantenimiento de la Jurisdicción de Servicios de Salud, Ensenada, México

^bLaboratorio de Producción y Métodos del Departamento de Ingeniería Industrial, Facultad de Ingeniería, Universidad Nacional de Asunción, Paraguay

*Corresponding author: carlos.zplg@gmail.com

Abstract—Artificial Intelligence (IA) has emerged as a transformative tool in the healthcare environment. This review aims to evaluate the effects of AI-based technologies on the early prediction and detection of diseases. A systematic review of the literature was conducted following the PRISMA methodology guidelines. Out of 27,359 studies, 28 were included in the analysis as they provided evidence supporting the effectiveness of machine learning algorithms. The findings indicated that the algorithms were predominantly utilized for the prediction of cardiovascular (n = 9), neurological (n = 6), oncological (n = 5), hepatic (n = 3), pulmonary (n = 3), and infectious diseases (n = 2). In addition, the accuracy of the main algorithms was measured in 23 studies, and 19 reported values greater than 90%. The application of AI techniques has demonstrated significant potential to improve the diagnosis and prediction of a wide range of diseases.

Keywords—Machine learning, medical diagnosis, predictive algorithms, healthcare 4.0.

Resumen—La Inteligencia Artificial (IA) ha surgido como una herramienta transformadora en el entorno de la salud. Esta revisión tiene como objetivo evaluar los efectos de las tecnologías con base en IA en la predicción y detección oportuna de enfermedades. Se realizó una revisión sistemática de la literatura con base en los lineamientos de la metodología PRISMA. A partir de un total de 27,359 estudios identificados, 28 fueron incluidos debido a que demostraron la efectividad de los algoritmos de machine learning. Los resultados revelaron que los algoritmos se aplicaron principalmente para predecir enfermedades cardiovasculares (n = 9), neurológicas (n = 6), oncológicas (n = 5), hepáticas (n = 3), pulmonares (n = 3) e infecciosas (n = 2). Además, se midió la precisión de los principales algoritmos en 23 estudios y 19 reportaron valores superiores al 90%. La aplicación de técnicas de IA, ha demostrado un potencial significativo para mejorar el diagnóstico y la predicción de una amplia gama de enfermedades.

Palabras Claves—Machine learning, diagnóstico médico, algoritmos predictivos, salud 4.0.

INTRODUCCIÓN

EN los últimos años la Inteligencia Artificial (IA), ha surgido como una herramienta transformadora del sector de la salud, ofreciendo nuevas oportunidades para la predicción y detección oportuna de enfermedades [1-2]. Por ejemplo, machine learning (ML) y deep learning (DL) han

demostrado su eficacia en el análisis de datos clínicos, imágenes médicas y señales biológicas, con el objetivo de identificar patrones y predecir el desarrollo de enfermedades como el cáncer, enfermedades cardiovasculares, enfermedades pulmonares y enfermedades neurológicas [3-7]. Estos modelos, funcionan a través de algoritmos computacionales altamente sofisticados que les permiten extraer información de los datos clínicos de manera

autónoma y mejorar su rendimiento sin la intervención humana. Los algoritmos de predicción se clasifican generalmente en dos grupos: supervisados y no supervisados [8]. Los algoritmos de aprendizaje supervisado, son utilizados para predecir variables objetivo específicas, a partir de distintos datos de entrada [9]. Entre los algoritmos de aprendizaje supervisado más utilizados se encuentran las Redes Neuronales Artificiales (*Artificial Neural Networks*, ANN), Máquinas de Vectores de Soporte (*Support Vector Machines*, VSM), Árboles de Decisión (*Random Forest*, RF) y Regresión Logística (*Logistic Regression*, LR) [10].

Por ejemplo, las ANN han sido ampliamente utilizadas en la predicción de enfermedades cardíacas, cáncer de mama y diabetes [11-12]. Las VSM se han empleado para la detección oportuna de enfermedades como el Alzheimer y Parkinson [12].

Por otro lado, los algoritmos de aprendizaje no supervisado son utilizados para identificar patrones ocultos y agrupaciones dentro de los datos, sin un objetivo específico predefinido [13]. Ejemplos de estos algoritmos son K-Medias, Agrupamiento Jerárquico y Análisis de Componentes Principales, los cuales han sido utilizados para identificar subtipos de enfermedades o agrupaciones de pacientes con características similares [14-16].

Independientemente del algoritmo utilizado, los investigadores evalúan sus modelos través de diversas métricas de desempeño. Estas evaluaciones, están relacionadas con la precisión, es decir, el porcentaje de predicciones correctas realizadas por el modelo [17-18]. La sensibilidad, para identificar correctamente a todos los pacientes enfermos [19]. En contraste, también se evalúa la capacidad para identificar correctamente a los pacientes sanos, conocido como especificidad [17-18]. Otra métrica importante y comúnmente reportada en la literatura, es el área bajo la curva (*Area Under the Curve*, AUC), la cual permite una correcta clasificación entre los pacientes [18-19]. De estos principales indicadores, se pueden derivar otros importantes como la Puntuación-F1 [18-20].

De este modo, la integración entre la IA y los sistemas de salud se denomina Salud 4.0 (S4.0) y representa una evolución significativa en el cuidado de la salud. El enfoque de S4.0 emplea tecnologías emergentes con base en datos sin procesar, como Salud Inteligente, Salud Móvil, Salud Inalámbrica, Registros Electrónicos de pacientes y Telemedicina, para crear sistemas interconectados [21-22]. Estas tecnologías, permiten elaborar tratamientos personalizados y en tiempo real para el cuidado de los pacientes, mejorando la precisión de los diagnósticos y reducir el error humano [22-23]. En consecuencia, los profesionales de la salud pueden tomar decisiones informadas, mejorando los resultados de los pacientes y agilizando los procedimientos clínicos.

Por lo tanto, la intervención oportuna en las distintas enfermedades a través de modelos computacionales, puede

impactar de forma significativa la trayectoria de la enfermedad, de modo que se puede fortalecer la eficacia del tratamiento y mejorar la calidad de vida de los individuos afectados [1-2].

No obstante, a pesar de los avances prometedores en SH4.0, los desafíos asociados en la adopción de la IA persisten. Estos retos plantean obstáculos relacionados con la calidad de los datos, la capacidad de interpretación de los modelos y la integración en los flujos de trabajo clínicos [23-24]. Además, la literatura sobre S4.0 permanece dispersa y carece de directrices prácticas. A pesar de que previas revisiones han intentado condensar la información sobre la transformación digital en salud, el tema en común ha sido las tendencias, oportunidades y desafíos. Por consiguiente, existe una necesidad de evaluar los efectos actuales de la IA en la predicción y detección temprana de enfermedades e identificar lagunas en la literatura existente.

Esta revisión tiene como objetivo, primero, identificar las principales técnicas de IA utilizadas en la detección temprana y predicción de enfermedades. Segundo, evaluar la eficacia de las diversas tecnologías utilizadas de IA y ML. Tercero, identificar los principales factores críticos de éxito, así como las barreras asociadas con la implementación de las tecnologías de S4.0. Adicionalmente, esta revisión sistemática examina la implementación de tecnologías de IA en la predicción y detección temprana de enfermedades en diferentes países y entornos de atención médica.

II. METODOLOGÍA

La revisión sistemática se realizó de acuerdo con los elementos de informe preferidos para revisiones sistemáticas y metaanálisis (PRISMA) [25]. La Fig. 1 presenta el diagrama de flujo de las etapas involucradas en el proceso de selección. Las siguientes subsecciones describen la metodología completa.

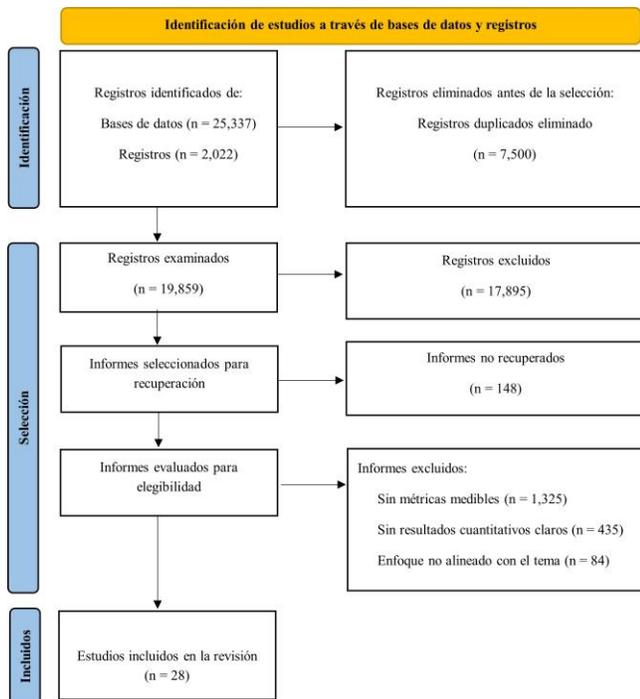


Fig. 1. Revisión sistemática de la literatura utilizando la metodología PRISMA.

2.1 Criterios de elegibilidad, fuentes de información y estrategia de búsqueda

Se incluyeron estudios que reportaron la aplicación de tecnologías como IA y ML para la predicción y detección temprana de enfermedades. Además, se incluyeron artículos originales publicados entre 2020 y 2024, escritos en inglés y español centrados en el desarrollo y validación de modelos de IA. Los estudios incluyeron grupos de participantes de todas las edades y géneros. Se excluyeron todos los artículos que contenían solamente el resumen, editoriales, presentaciones en poster, tesis académicas, artículos de divulgación y estudios que no involucraron sujetos humanos.

Se realizó una búsqueda exhaustiva de la literatura comprendiendo el periodo de diciembre de 2023 a julio de 2024, en distintas de bases de datos orientadas al área de la salud, incluyendo PubMed y la librería de Cochrane. Además, se consultaron otras bases de datos interdisciplinarias como EBSCO, Google Scholar, Science Direct, Emerald insight e IEEE Xplore, para recopilar toda la información relevante relacionada con los efectos de la IA en la predicción y detección temprana de enfermedades. Adicionalmente artículos potenciales se identificaron a través de una búsqueda manual en la lista de referencias de los estudios seleccionados y algunas plataformas de registros de ensayos clínicos como el Registro de Ensayos Clínicos del Gobierno de los Estados Unidos (NIH) y de la

Organización Mundial de la Salud (ICTRP).

La estrategia de búsqueda involucro la utilización de una serie de términos relacionados a la IA y la detección y predicción de enfermedades para abordar todos los estudios relevantes. Las palabras clave utilizadas en la búsqueda fueron inteligencia artificial, machine learning, deep learning, redes neuronales, modelos predictivos, predicción de enfermedades y detección oportuna. Operadores booleanos como AND, OR y NOT, se utilizaron para combinar las palabras clave.

2.2 Proceso de selección y extracción de datos

Dos investigadores independientes revisaron el título y resumen de cada artículo seleccionado, con el propósito de evaluar si los estudios eran consistentes con los criterios de inclusión y exclusión, además todos los trabajos duplicados se eliminaron. Los estudios que se alinearon con los criterios de inclusión se sometieron a una revisión exhaustiva. Cualquier desacuerdo entre los investigadores se resolvió a través de un consenso con un tercer investigador.

La extracción se realizó utilizando un formato de Microsoft Excel estandarizado para asegurar la consistencia y precisión de los datos. La información que se recuperó de los estudios incluyó, metadatos bibliográficos, características clave, y un resumen de los hallazgos. El formato generado fue verificado por un tercer investigador para validar la extracción de los datos.

Los datos extraídos incluyeron el título, el año y país de publicación, el apellido del primer autor, la enfermedad objetivo de estudio, el entorno de la intervención, el tamaño de muestra, la principal intervención, los factores críticos de éxito, la arquitectura de IA, ML y las métricas de rendimiento para la clasificación de modelos. Adicionalmente, se recolectaron los datos de las características de los participantes como la edad, el género, información demográfica, estados de salud. Respecto a los datos faltantes o ambiguos, se contactó a los autores para especificar la información.

2.3 Evaluación del riesgo de sesgo y medidas del efecto

El riesgo de sesgo se evaluó utilizando la herramienta de Cochrane (RoB 2) [26] para ensayos clínicos aleatorizados y ROBINS-I [27] para intervenciones en estudios no aleatorizados y observacionales. Para evaluar la aplicabilidad y calidad de los modelos se utilizó una versión ajustada de la herramienta para evaluar el riesgo de sesgo en modelos predictivos (PROBAST), integrada por cuatro dominios, la selección de participantes, predictores, resultados y análisis. La validación del riesgo de sesgo de los estudios incluidos se llevó a cabo por dos investigadores de manera independiente. En todos los casos, los estudios se calificaron como bajo, intermedio o alto riesgo de sesgo con base en el juicio de los investigadores, y ante una diferencia,

un tercer revisor evaluó el estudio para llegar a un acuerdo.

Para cada uno de los resultados se incluyeron las principales características de rendimiento de los modelos como precisión, sensibilidad, especificidad, así como métricas con base en curvas como, área bajo la curva y métricas balanceadas como la Puntuación-F1. Estos indicadores fueron indispensables para evaluar la capacidad de los modelos para la detección y predicción oportuna de enfermedades.

2.4 Métodos de síntesis

Los datos extraídos se organizaron manualmente y se tabularon en formatos estandarizados. Sin embargo, debido a la heterogeneidad prevista entre los estudios, en términos de diseño, contexto y resultados, no fue posible realizar un metaanálisis. Por lo tanto, se realizó una síntesis descriptiva de los resultados con gráficos, figuras y tablas que resumen los hallazgos.

III. RESULTADOS

La búsqueda de literatura en las bases de datos resultó en un total de 25,337 estudios, los cuales fueron identificados como destacables en la primera exploración. De manera similar, 2,022 registros fueron recuperados de plataformas especializadas en registros de ensayos clínicos (NIH, $n = 987$ y ICTRP, $n = 1,035$). Después, se eliminaron 7,500 estudios duplicados, de esta manera la cantidad total se redujo a 19,859 registros únicos. Posteriormente, con la revisión de la información de primera mano cómo el título y el resumen, se excluyeron 17,895 estudios. En la siguiente etapa 1,964 registros se evaluaron detalladamente, sin embargo, por falta de acceso o datos incompletos, 148 quedaron fuera dejando así 1,816 reportes preliminarmente elegibles. Sin embargo, 1,844 no fueron incluidos debido a que no cumplieron con aspectos críticos de inclusión como por ejemplo métricas de medición ($n = 1,325$), resultados poco claros ($n = 435$) o el enfoque del estudio era irrelevante para la revisión ($n = 84$).

Esta depuración final dejó únicamente 28 estudios que cumplieron con todos los criterios y, por lo tanto, fueron incluidos en la revisión.

3.1. Atributos de los Estudios Incluidos

La distribución de las publicaciones presentó una variabilidad considerable. La Fig. 2 indica que los años con mayor cantidad de publicaciones fue en 2023 ($n = 10$), mientras que 2024 tuvo la menor participación ($n = 1$). No obstante, existe un aumento considerable desde 2021 ($n = 2$) hasta 2023.

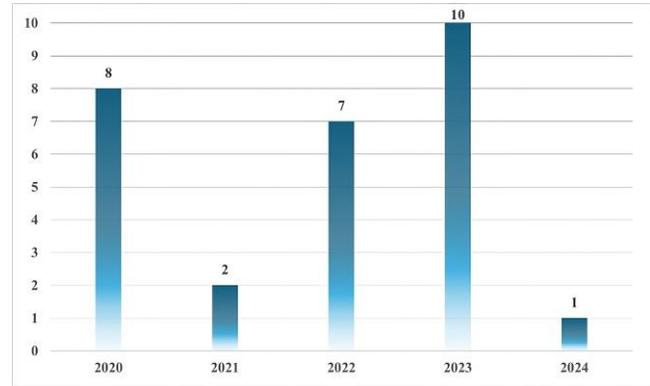


Fig. 2. Cantidad de publicaciones por año.

De manera similar, en la Fig. 3 se observa una diversificación de países, entornos y diseños de estudio de las investigaciones. Específicamente, Estados Unidos de América (EE.UU.) contribuyó con el mayor número de publicaciones incluidas ($n = 7$), seguido de China ($n = 6$), India ($n = 5$) y Alemania ($n = 2$), el resto de los países se distribuyeron en diversos países.

Posteriormente, la tabla 1 describe los atributos clave de los estudios incluidos. De esta manera, los datos clínicos de pacientes junto al sistema hospitalario en general, representaron los entornos con más menciones, y con menor medida se encontraron los entornos como poblaciones rurales y laboratorios de análisis de voz y de movimiento.

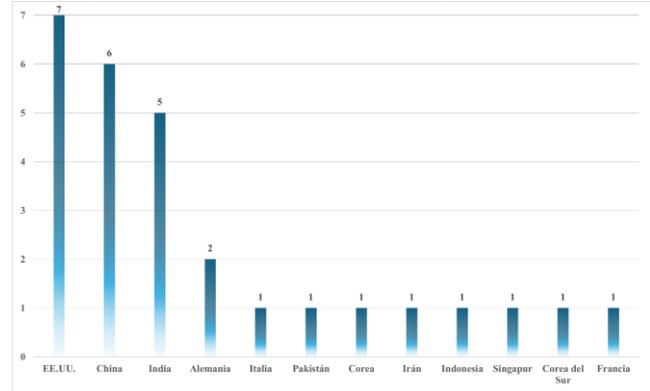


Fig. 3. Cantidad de publicaciones por país.

Por otro lado, se logró identificar que las enfermedades con mayor frecuencia fueron la enfermedad de Alzheimer y la cardiopatía, ambas mencionadas en 4 estudios.

Respecto a los algoritmos, las VSM surgieron como el algoritmo más frecuente, empleado en 13 estudios seguido de RF, utilizado en 12 estudios. La LR y XGBoost tuvieron notables apariciones entre los estudios. En términos de rendimiento del algoritmo, SVM demostró mejoras sustanciales en múltiples métricas. Por ejemplo, en estudios de la enfermedad de Alzheimer, las aplicaciones de SVM lograron un aumento de la sensibilidad a 89% y mejoras de especificidad de 95% a 96% ($p < 0.0001$). De manera similar,

en otro estudio sobre el Alzheimer en el que se utilizó SVM con mejoras de kernel, la precisión alcanzó 92.31 %, la sensibilidad mejoró hasta un 96.33 % y la especificidad aumentó hasta 88.62 % ($p < 0.0001$).

Por su parte, RF también mostró mejoras significativas, con estudios que informaron aumentos del AUC de 99.97 % en el síndrome coronario agudo y mejoras de precisión hasta 93.83% en los casos de cardiopatía. La LR, utilizada en ocho estudios, mejoró de manera constante sus métricas de precisión, como la mejora 89.42 % a 91.6 % en cardiopatía y un logro de una precisión del 93.15% en otro estudio de cardiopatía ($p < 0.05$).

XGBoost también demostró mejoras significativas en el rendimiento, con un aumento de las AUC de hasta 98% en

las enfermedades hepáticas y una precisión que alcanzó 94.41 % en los casos de cardiopatía. Otros algoritmos notables incluyeron DT, utilizados en cuatro estudios, ANN en tres y NB en tres. Estos algoritmos contribuyeron a diversas mejoras en la precisión, la sensibilidad y la especificidad.

TABLA 1. Resumen de Estudios sobre el Uso de Algoritmos en el Diagnóstico de Enfermedades.

Primer Autor Año, País	Contexto, Diseño del Estudio, n, Período	Enfermedad	Algoritmo	Resumen de Resultados
[28], 2020, Italia	Imágenes clínicas; Estudio transversal; (n= 80); No especificado	Alzheimer	SVM, RF, ANN	Sensibilidad incrementó a 89% y Especificidad mejoró de 95% a 96% ($p < 0.0001$)
[29], 2020, Pakistán	Datos clínicos; Estudio experimental; (n= 304, 1025); No especificado	Cardiopatía	LR, DT, NB, RF, ANN, SVM, AB, GB	AUC incrementó de 92% a 94.24%, Precisión mejoró de 92.09% a 94.41%, Sensibilidad aumentó a 95%, y Especificidad incrementó a 96% ($p < 0.00001$)
[30], 2020, EE.UU.	UCI; Estudio de cohorte; (n= 5,691); 12 meses	Hemorragia gastrointestinal	XGBoost	AUC incrementó a 85% y Especificidad mejoró a 27% ($p < 0.001$)
[31], 2020, China	Datos clínicos; Estudio longitudinal; (n= 1,131); 24 meses	Predicción de Accidentes cerebrovasculares	LRR, SVM, RF	AUC incrementó a 72%, Sensibilidad mejoró a 75%, y Especificidad aumentó a 80% ($p < 0.01$)
[32], 2020, China	Datos clínicos; Estudio transversal; (n= 124); No especificado	Acromegalia	LDA	Precisión mejoró a 92.86% en mujeres y a 75% en hombres ($p < 0.0001$, $p < 0.001$)
[33], 2020, Corea	Sistema hospitalario; Estudio observacional prospectivo; (n= 70); 12 meses	Problemas de autocuidado en niños con discapacidades motoras y físicas	GA-XGB, DT, NB, LR, SVM	Precisión incrementó de 84.29% a 90% y Puntuación-F1 mejoró de 92.86% a 98.57% ($p < 0.05$)
[34], 2020, Irán	Laboratorio de análisis; Estudio comparativo y de clasificación; (n=60); No especificado	Alzheimer	SVM (+Kernels)	Precisión incrementó a 92.31%, Sensibilidad mejoró a 96.33%, y Especificidad aumentó a 88.62% ($p < 0.0001$)
[35], 2020, EE.UU.	Sistema hospitalario; (n= 677); No especificado	Tuberculosis	CheXaid	Precisión mejoró a 79% y Sensibilidad aumentó en 5% ($p = 0.002$)
[36], 2021, EE.UU.	Datos clínicos de UCI; Estudio de cohorte; (n= 615); No especificado	Enfermedades del hígado (hepatitis, fibrosis, cirrosis)	ANN, RF, SVM	AUC incrementó a 98% y Precisión mejoró a 98.14% ($p < 0.001$)
[37], 2021, Alemania	Imágenes clínicas; Estudio observacional transversal; (n= 2,110); No especificado	Alzheimer	RF, CNN	AUC incrementó a 92.6% y Precisión mejoró a 91.68% ($p < 0.001$)
[38], 2022, China	Población rural; Estudio de cohorte; (n=12,692); 59.3 meses	Enfermedad Cardiovascular	L1-LR, RF, SVM, Lasso-AdaBoost	AUC incrementó a 81.7%, Sensibilidad mejoró a 89%, y Especificidad aumentó a 84% ($p < 0.05$)
[39], 2022, China	Sistema hospitalario; Estudio retrospectivo; (n= 1,374); 4 meses	COVID-19	RF, AB, GBDT, XGB	AUC incrementó a 86.4%, Precisión mejoró a 94%, y Sensibilidad aumentó a 74% ($p < 0.01$)
[40], 2022, India	Sistema hospitalario; Estudio comparativo; (n= 92); No especificado	Cardiopatía	DKNN, SVM	Precisión mejoró de 67.21% a 84.44% ($p < 0.001$)
[41], 2022, India	Datos clínicos; Estudio transversal; (n= 70,000); No especificado	Enfermedad Cardiovascular	RFE-GB	Precisión incrementó a 88.84% y Puntuación-F1 mejoró a 93.83% ($p < 0.05$)
[42], 2022, EE.UU.	Sistema hospitalario; Estudio longitudinal prospectivo; (n= 4,664); No especificado	Alzheimer	CNN, GB	AUC incrementó a 87% y Precisión mejoró a 85.3% ($p < 0.001$)

[43], 2022, China	Datos clínicos; Estudio retrospectivo; (n= 4,887); 120 meses	Carcinoma Laríngeo	LR, KNN, SVM, RF, XGB, LGBM	AUC incrementó a 82.9%, Precisión mejoró a 77%, y Puntuación-F1 aumentó a 79.1% (p < 0.001)
[44], 2022, India	Sistema hospitalario; Estudio experimental; (n= 20); No especificado	Cardiopatía	LR, DT	Precisión mejoró de 89.42% a 91.6% (p < 0.05)
[45], 2023, Indonesia	Datos clínicos; Estudio transversal; (n= 606); No especificado	Cardiopatía	RF, LR, KNN	Precisión incrementó a 93.15% y Puntuación-F1 mejoró a 93.83% (p < 0.05)
[46], 2023, India	Datos clínicos; Estudio longitudinal retrospectivo; (n= 500); No especificado	Enfermedad de la Arteria Coronaria	PABC, LR, DT, RF, SVM	AUC incrementó a 90%, Precisión mejoró a 85%, Sensibilidad aumentó a 94%, y Especificidad mejoró a 88% (p < 0.001)
[47], 2023, Singapur	Sistema hospitalario; Estudio de cohorte retrospectivo; (n= 3,710); No especificado	Artritis Reumatoide	RFE-RF, LR, XGB, SVM, NB	AUC incrementó a 90%, Sensibilidad mejoró a 94%, y Especificidad aumentó a 88% (p < 0.001)
[48], 2023, India	Sistema hospitalario; Estudio diagnóstico prospectivo; (n= 300); No especificado	Enfermedad Pulmonar Obstructiva Crónica	F-QNN, GBB-BC	Precisión incrementó a 96%, Sensibilidad mejoró a 95%, y Especificidad aumentó a 92% (p < 0.05)
[49], 2023, China	UCI; Estudio longitudinal prospectivo; (n= 7,149); 144 meses	Insuficiencia Cardíaca	tBNA-PR	AUC incrementó a 71.98%, Precisión mejoró a 78%, y Puntuación-F1 aumentó a 76.7% (p < 0.01)
[50], 2023, Corea del Sur	Datos clínicos; Estudio observacional retrospectivo; (n= 13,104); 179 meses	Síndrome Coronario Agudo	RF, ET, GBM	AUC incrementó a 99.97%, Precisión mejoró a 99.07%, Sensibilidad aumentó a 98.2%, y Puntuación-F1 mejoró a 99.97% (p < 0.001)
[51], 2023, EE.UU.	Sistema hospitalario; Estudio de Cohorte retrospectivo; (n= 109,490); 120 meses	Enfermedad Cardiovascular	LR, RF, GBT, LSTM	AUC incrementó a 78.1%, Sensibilidad mejoró a 87%, y Especificidad aumentó a 85% (p < 0.05)
[52], 2023, Alemania	Análisis de voz; Estudio experimental; (n= 376); No especificado	Parkinson	Wav2Vec 2.0, FCN-4	AUC incrementó a 88%, Precisión mejoró a 83.2% y a 76%, y Sensibilidad mejoró a 85% (p < 0.001)
[53], 2023, EE.UU.	Imágenes clínicas; Estudio diagnóstico prospectivo; (n= 7,113); No especificado	Enfermedad Cardíaca Reumática	ResNet-50, LinkNet-VGG16, SVM	Precisión incrementó a 86%, Sensibilidad mejoró a 92%, y Especificidad aumentó a 79% (p < 0.05)
[54], 2023, Francia	Imágenes clínicas; Estudio longitudinal; (n= 299 videos); 60 meses	Parkinson	XGB, VDFF	AUC incrementó a 83.1% y Precisión mejoró a 77.3% (p < 0.003)
[55], 2024, EE.UU.	Datos clínicos; Estudio de cohorte; (n= 511); No especificado	Enfermedad Cardíaca Reumática	SVM, 3D-CNN	AUC incrementó a 84%, Sensibilidad mejoró a 98%, y Precisión aumentó a 85% (p < 0.05)

Los estudios emplearon distintos diseños de investigación, incluyendo estudios transversales, de cohorte y longitudinales, con tamaños de muestra que oscilaban entre 20 hasta 109, 490 participantes y períodos de estudio que se extendieron hasta un máximo de 179 meses.

De manera complementaria, la Tabla 2 presenta información sobre el rendimiento comparativo de los algoritmos de ML en la predicción y el diagnóstico de enfermedades en tres categorías, imágenes médicas, parámetros clínicos y perfiles genéticos.

En el campo de las imágenes médicas, las VSM y las ANN han demostrado un rendimiento notable. El estudio de [34] que utilizó VSM logró una precisión del 92.3%, una sensibilidad del 96.3% y una especificidad del 90.8%, aunque no se informó el AUC.

En contraste, la aplicación de SVM por [53] arrojó una precisión del 86%, una sensibilidad del 92% y una especificidad del 79%, lo que resalta la solidez de las SVM

en diferentes contextos. El modelo ANN en el estudio de [28] exhibió un rendimiento moderado con una precisión del 75%, una sensibilidad del 80% y una especificidad del 70%, junto con un AUC del 83%.

Además, la combinación de CNN-3D y transformadores en la investigación de [55] demostró una alta sensibilidad y una precisión del 84%, con un AUC del 84%, lo que indica su fortaleza para identificar eficazmente los casos positivos verdaderos.

Otros algoritmos, como CheXaid y FastSurfer, informaron valores de precisión del 78% y el 88.2%, respectivamente, con métricas de sensibilidad y especificidad variables, lo que refleja diversos niveles de rendimiento en diferentes técnicas de imágenes.

Con respecto a la predicción y el diagnóstico con base en parámetros clínicos, los métodos de conjunto como RF y XGB ofrecieron constantemente un alto rendimiento. La aplicación RF de [36] demostró una precisión, sensibilidad y

autores en la literatura [56-57]. La integración de modelos avanzados de ML, incluidas las VSM, RF, XGBoost y ANN, no solo ha mejorado la precisión del diagnóstico, sino que también trazan el camino para intervenciones médicas más personalizadas y oportunas [56-58].

En concordancia con otros autores, se identificó que las enfermedades cardiovasculares fueron la categoría más estudiada [59-61]. El predominio de los algoritmos RF y XGBoost en este dominio subraya su solidez en el manejo de datos clínicos complejos y de alta dimensión. Por ejemplo, en [36] demostraron el desempeño superior de RF, con una precisión del 98.1%, una sensibilidad del 99%, una especificidad del 97.3% y un área bajo la curva del 98%, lo que destaca su potencial para la predicción confiable de enfermedades cardiovasculares. De manera similar, en [45] optimizaron RF utilizando Optuna, logrando una precisión del 93.2%, una sensibilidad del 95%, una especificidad del 90.9% y un AUC del 92,9%.

IV. DISCUSIÓN

Esta revisión sistemática de literatura destaca la eficacia y versatilidad de los algoritmos de aprendizaje automático en múltiples entornos de salud, que comprenden enfermedades cardiovasculares, neurológicas, oncológicas, hepáticas, pulmonares e infecciosas, así como lo proponen diversos

Estos hallazgos corroboran la afirmación de [2] en relación con los métodos de conjunto como RF brindan una estabilidad y precisión mejoradas, lo que los convierte en herramientas esenciales en el diagnóstico cardiovascular.

Sin embargo, la variabilidad en las métricas de desempeño en los estudios sugiere que el ajuste de algoritmos y el preprocesamiento de datos juegan un papel crítico en la optimización de los resultados. Por ejemplo, en [38] informaron un AUC del 81.7% utilizando una combinación de L1-LR, RF, SVM y Lasso-AdaBoost, lo que indica que los enfoques híbridos pueden refinar aún más las capacidades predictivas.

TABLA 2. Valores de desempeño de los modelos utilizados.

Algoritmos para la predicción y diagnóstico de enfermedades mediante imágenes							
<i>Ref.</i>	<i>Técnica</i>	<i>Prec. (%)</i>	<i>Sens. (%)</i>	<i>Esp. (%)</i>	<i>F1 (%)</i>	<i>Pred. (%)</i>	<i>AUC (%)</i>
[28]	ANN	75	80	70	NM	NM	83
[32]	LDA	92.9	NM	NM	NM	NM	NM
[35]	CheXaid	78	67	87	NM	NM	83
[37]	FastSurf	88.2	NM	NM	NM	NM	NM
[42]	3D CNN	NM	NM	NM	NM	NM	85.0
[55]	3D-CNN +Transf.	84	98	NM	87	NM	84
[53]	SVM	86	92	79	NM	NM	NM
[54]	XGBoost	NM	NM	NM	NM	NM	83.1
[34]	SVM	92.3	96.3	90.8	NM	NM	NM

Algoritmos utilizados en la predicción y diagnóstico de enfermedades usando parámetros clínicos							
<i>Ref.</i>	<i>Técnica</i>	<i>Prec. (%)</i>	<i>Sens. (%)</i>	<i>Esp. (%)</i>	<i>F1 (%)</i>	<i>Pred. (%)</i>	<i>AUC (%)</i>
[29]	ET	94.4	94.9	94.8	95	NM	94.2
[30]	XGBoost	NM	100	27	NM	NM	85
[31]	LRR/SVM/RF	70/95	75.0	69.0	NM	NM	72.0
[33]	GA-XGBoost	90.0	84.8	NM	81.2	NM	NM
[36]	RF	98.1	99	97.3	NM	NM	98
[38]	FLR-L1-LR	79.5	75.0	79.0	NM	NM	81.7
[39]	GBDT	80.4	74	74	82.8	80.3	86.4
[40]	DKNN	84.4	NM	NM	NM	NM	NM
[41]	RFE-GB	88.8	NM	NM	83	NM	NM
[44]	LR	91.6	90	92	NM	NM	NM
[51]	GBT	78.1	68.0	74.0	NM	NM	78.1
[52]	FL	83.2	77.2	89.2	NM	NM	83.6
[45]	Optuna RF	93.2	95	90.9	NM	NM	92.9
[46]	PABC	85	87	82	NM	NM	90
[48]	F-QNN	96	95	98	NM	NM	91
[49]	tBNA-PR	78	NM	NM	76.7	NM	71.9

Algoritmos utilizados en la predicción y diagnóstico de enfermedades mediante análisis de perfiles genéticos							
<i>Ref.</i>	<i>Técnica</i>	<i>Prec. (%)</i>	<i>Sens. (%)</i>	<i>Esp. (%)</i>	<i>F1 (%)</i>	<i>Pred. (%)</i>	<i>AUC (%)</i>
[47]	OPRS	94	91	95	90	NM	97

Algoritmos utilizados en la predicción y prognosis de enfermedades							
<i>Ref.</i>	<i>Técnica</i>	<i>Prec. (%)</i>	<i>Sens. (%)</i>	<i>Esp. (%)</i>	<i>F1 (%)</i>	<i>Pred. (%)</i>	<i>AUC (%)</i>
[43]	XGBoost	77	67.7	91.2	70.6	NM	82.9

Esto coincide con las perspectivas de [5], quienes promueven modelos híbridos para aprovechar las fortalezas de múltiples algoritmos, mitigando así las limitaciones individuales y mejorando el rendimiento general.

En el dominio de los trastornos neurológicos, seis estudios examinaron afecciones como la enfermedad de Alzheimer y Parkinson, empleando principalmente VSM y ANN. Por ejemplo, en [34] utilizaron un modelo SVM, logrando una precisión del 92.3 %, una sensibilidad del 96.3%, una especificidad del 90.8 % y un área bajo la curva de aproximadamente el 85%. Estas métricas de rendimiento destacan la formidable capacidad de SVM para identificar enfermedades neurodegenerativas, particularmente en sus primeras etapas donde la intervención oportuna es crucial.

Por el contrario, en [42] integraron 3D-CNN con transformadores, logrando una sensibilidad del 98% y una precisión del 84%, aunque el AUC se mantuvo en el 84%. Esto sugiere que, si bien las arquitecturas de aprendizaje profundo híbridas pueden aumentar la sensibilidad, equilibrar esto con la especificidad y el poder discriminatorio general sigue siendo un desafío. El análisis comparativo de las VSM y las ANN en el diagnóstico neurológico refleja las ventajas y limitaciones inherentes de estos algoritmos. Como se menciona en [11],

las SVM se destacan en escenarios con límites de decisión claros y espacios de características de alta dimensión, lo que

las hace adecuadas para datos neurológicos complejos. Por el contrario, la variabilidad del rendimiento de las ANN, como observaron en [18] sugiere que las ANN pueden requerir datos de entrenamiento más extensos y un ajuste fino meticuloso para lograr resultados óptimos en este dominio.

Se examinaron enfermedades oncológicas en cinco estudios, empleando un conjunto diverso de algoritmos que incluían LR, KNN, SVM y XGBoost. En [43] utilizaron una combinación de estos algoritmos para el diagnóstico de carcinoma laríngeo, logrando un AUC del 82.9%, una precisión del 77% y una puntuación F1 del 79.1%. Estos resultados demuestran el potencial de los enfoques de múltiples algoritmos para mejorar la precisión diagnóstica en la detección del cáncer como lo afirman otros autores [62]. Sin embargo, la precisión moderada indica que hay margen de mejora, posiblemente mediante la integración de modalidades de datos adicionales o técnicas de ingeniería de características más sofisticadas.

La variabilidad observada en las métricas de rendimiento en los estudios oncológicos evidencia la complejidad inherente de los diagnósticos del cáncer. Factores como la heterogeneidad de las fuentes de datos y la naturaleza diversa de los subtipos de enfermedades pueden afectar significativamente la eficacia de los enfoques algorítmicos. Este hallazgo se alinea con las perspectivas de [63-65], quienes enfatizan la necesidad de desarrollar modelos de aprendizaje automático personalizados que tengan en cuenta las características únicas de los diferentes tipos de cáncer para lograr una mayor precisión predictiva. Asimismo, en esta revisión se logró identificar de manera general, los factores clave que permiten la implementación exitosa de estas aplicaciones de IA.

Algunos de los más representativos incluyeron la óptima integración de los algoritmos con los sistemas de información hospitalaria existentes y la interoperabilidad con los registros médicos electrónicos.

En este mismo sentido, los estudios que demostraron resultados óptimos generalmente emplearon métodos que se evaluaron dentro de infraestructuras clínicas capaces de manejar conjuntos de datos complejos. Esto sugiere que cuando los datos se capturan, curan y formatean de manera confiable, los modelos de ML se pueden entrenar y validar de manera más efectiva [56-58]. Otro factor predominante entre los estudios fue la presencia de personal bien capacitado, que comprenda los resultados algorítmicos y pueda interpretarlos dentro de un contexto clínico. Esto sugiere que sin profesionales de la salud que confíen en que un modelo que arroja una precisión del 98.14% y un AUC del 98% en el diagnóstico de una enfermedad hepática es confiable, el potencial tecnológico permanece infrutilizado [66-67].

La interoperabilidad resulto ser particularmente notable, así como lo aseguran otros autores [68-69]. Los

estudios que emplean datos multimodales se beneficiaron de terminologías y estructuras de datos estandarizadas, lo que permitió una integración más fluida y redujo el riesgo de degradación del rendimiento cuando los modelos se implementan más allá del entorno de prueba original.

A pesar de los avances significativos observados y las prometedoras tecnologías, está claro que existen desafíos y barreras persistentes en la implementación de estas aplicaciones de IA. Por ejemplo, resulto ser común entre algunos autores distintas preocupaciones en el tema de la privacidad y la seguridad de los datos [70-71]. Esto es debido a que la complejidad de entrenar modelos grandes con datos confidenciales de pacientes plantea una serie de riesgos potenciales. No obstante, los estudios incluidos en esta revisión no reportaron explícitamente alguna problemática de seguridad.

Por otra parte, un constante desafío es el de garantizar el cumplimiento de las normas regulatorias, esto plantea un obstáculo importante. Además, los altos costos asociados con la implementación de la infraestructura necesaria, la concesión de licencias a las plataformas de ML y la financiación de la capacitación del equipo pueden requerir muchos recursos, lo que presenta otra barrera para entornos subdesarrollados. Eventualmente, la resistencia al cambio dentro de las jerarquías clínicas también puede frenar la adopción de esta tecnología.

V. CONCLUSIÓN

Los hallazgos presentados en esta revisión sistemática subrayan el potencial de la IA para reforzar la detección temprana de enfermedades y mejorar la precisión diagnóstica en diversos entornos clínicos. Las técnicas con base en IA, incluyendo tanto métodos tradicionales de aprendizaje automático (ML) como de aprendizaje profundo (DL), han demostrado un impacto significativo, especialmente en áreas cardiovasculares, neurológicas y oncológicas, alcanzando en muchos casos métricas de precisión superiores al 90%.

Al integrarse en ecosistemas de salud 4.0, estas tecnologías pueden optimizar la eficiencia operativa, orientar la asignación de recursos y fomentar una atención continua y personalizada. Con su capacidad para acelerar la identificación de individuos de alto riesgo y mejorar la clasificación de pacientes, la IA contribuye a disminuir procedimientos innecesarios, reducir retrasos en el diagnóstico y, en última instancia, aliviar la carga clínica y económica asociada a enfermedades crónicas y agudas.

El objetivo del estudio se ha cumplido al evidenciar que las tecnologías basadas en IA ofrecen avances considerables en la detección temprana de enfermedades. Sin embargo, es indispensable seguir investigando y perfeccionando su implementación para superar las limitaciones actuales, asegurar la interpretación responsable de los datos, abordar las consideraciones éticas y prácticas, y fomentar su

adopción en entornos diversos. Al aprovechar el poder de la IA de manera ética, responsable y centrada en el paciente, la atención médica puede avanzar hacia nuevas fronteras en la prevención, intervención temprana y tratamiento personalizado, lo que repercutirá positivamente en los resultados clínicos y en el uso eficiente de los recursos de salud.

AGRADECIMIENTOS

Los autores agradecen al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) por el apoyo otorgado, y a la Universidad Autónoma de Baja California (UABC) por el acceso a las bases de datos que posibilitaron este estudio.

REFERENCIAS

- [1] D. Houfani, S. Slatnia, O. Kazar, H. Saouli, and A. Merizig, "Artificial intelligence in healthcare: a review on predicting clinical needs," *International Journal of Healthcare Management*, vol. 15, no. 3. Taylor & Francis, p. 267, Feb. 28, 2021. doi: 10.1080/20479700.2021.1886478.
- [2] N. Caballé, J. L. Castillo, J. A. Gómez-Pulido, J. M. Gómez-Pulido, and M. L. Polo-Luque, "Machine Learning Applied to Diagnosis of Human Diseases: A Systematic Review," *Applied Sciences*, vol. 10, no. 15. Multidisciplinary Digital Publishing Institute, p. 5135, Jul. 26, 2020. doi: 10.3390/app10155135.
- [3] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine Learning for Medical Imaging," *Radiographics*, vol. 37, no. 2. Radiological Society of North America, p. 505, Feb. 17, 2017. doi: 10.1148/rg.2017160130.
- [4] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis.," Feb. 11, 2007, National Institutes of Health. Accessed: Dec. 2024. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/19458758>
- [5] R. Chalapathy and S. Chawla, "Deep Learning for Anomaly Detection: A Survey," Jan. 01, 2019, Cornell University. doi: 10.48550/arXiv.1901.1901.
- [6] M. Mann, R. P. Badoni, H. Soni, M. Al-Shehri, A. C. Kaushik, and D. Wei, "Utilization of Deep Convolutional Neural Networks for Accurate Chest X-Ray Diagnosis and Disease Detection," Mar. 26, 2023, Springer Science+Business Media. doi: 10.1007/s12539-023-00562-2.
- [7] J. Ker, L. Wang, J. P. Rao, and C. C. T. Lim, "Deep Learning Applications in Medical Image Analysis," Dec. 29, 2017, Institute of Electrical and Electronics Engineers. doi: 10.1109/access.2017.2788044.
- [8] A. Vellido, V. Ribas, C. Morales, A. R. Sanmartín, and J. C. Ruiz-Rodríguez, "Machine learning in critical care: state-of-the-art and a sepsis case study," Nov. 01, 2018, BioMed Central. doi: 10.1186/s12938-018-0569-2.
- [9] S. M. Varnosfaderani and M. Forouzanfar, "The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century," Mar. 29, 2024, Multidisciplinary Digital Publishing Institute. doi: 10.3390/bioengineering11040337.
- [10] K. Schenkelberg, U. Seidenberg, and F. Ansari, "Supervised Machine Learning for Knowledge-Based Analysis of Maintenance Impact on Profitability," Jan. 01, 2020, Elsevier BV. doi: 10.1016/j.ifacol.2020.12.2830.
- [11] J. Chen, H. H. Cheong, and S. W. I. Siu, "xDeep-AcPEP: Deep Learning Method for Anticancer Peptide Activity Prediction Based on Convolutional Neural Network and Multitask Learning," Jul. 30, 2021, American Chemical Society. doi: 10.1021/acs.jcim.1c00181.
- [12] F. Fernandes, H. Vicente, A. Abelha, J. Machado, P. Novais, and J. Neves, "Artificial neural networks in diabetes control," Jul. 01, 2015. doi: 10.1109/sai.2015.7237169.
- [13] M. L. Wang, "Algoritmos de minería de datos en la industria sanitaria," Jan. 01, 2021, Cornell University. doi: 10.48550/arxiv.2104.09395.
- [14] B. Lund, "Four Categories of Academic Libraries: A Cluster Analysis Based on Collections, Expenditures, and Circulation per Student Data," Oct. 01, 2017, Taylor & Francis. doi: 10.1080/14649055.2020.1794748.
- [15] G. Lorca, J. Arzola, y O. Pereira, "Segmentación de imágenes médicas digitales mediante técnicas de clustering," *Rev. Aporte Santiaguino*, vol. 3, no. 2, 2010.
- [16] M. C. de Souto, I. G. Costa, D. S. de Araujo, T. B. Ludermitz, and A. Schliep, "Clustering cancer gene expression data: a comparative study," Nov. 27, 2008, BioMed Central. doi: 10.1186/1471-2105-9-497.
- [17] J. Gardner and C. Brooks, "Evaluating Predictive Models of Student Success: Closing the Methodological Gap," Aug. 05, 2018, Society for Learning Analytics Research. doi: 10.18608/jla.2018.52.7.
- [18] M. C. M. Delgado et al., "Indicadores de calidad en el enfermo crítico," Jan. 01, 2008, Elsevier BV. doi: 10.1016/s0210-5691(08)70899-4.
- [19] D. J. Hand, "Assessing the Performance of Classification Methods," Aug. 22, 2012, Wiley. doi: 10.1111/j.1751-5823.2012.00183.x.
- [20] D. Powers, "Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation," Dec. 15, 2011, Cornell University. Accessed: Nov. 2024. [Online]. Available:
- [21] G. Dicuonzo, G. Galeone, M. Shini, and A. Massari, "Towards the Use of Big Data in Healthcare: A Literature Review," *Healthcare*, vol. 10, no. 7. Multidisciplinary Digital Publishing Institute, p. 1232, 2022. doi: 10.3390/healthcare10071232.
- [22] I. R. Herrera and F. Gascón, "Nuevas herramientas en salud," 2012, Elsevier BV. doi: 10.1016/j.medcli.2012.05.010.
- [23] J. F. Á. de Tomás, M. Á. Mayer, and V. J. Q. Varela, "La inteligencia artificial y sus aplicaciones en medicina II: importancia actual y aplicaciones prácticas," *Atención Primaria*, vol. 53, no. 1. Elsevier BV, p. 81, Jun. 19, 2020. doi: 10.1016/j.aprim.2020.04.014.
- [24] C. Kelly, A. Karthikesalingam, M. Suleyman, G. S. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," Oct. 29, 2019, BioMed Central. doi: 10.1186/s12916-019-1426-2.
- [25] M. J. Page et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," Mar. 29, 2021. doi: 10.1136/bmj.n71.
- [26] J. P. T. Higgins, J. Savović, M. J. Page, R. G. Elbers, and J. A. C. Sterne, "Assessing risk of bias in a randomized trial," Sep. 20, 2019. doi: 10.1002/9781119536604.ch8.
- [27] K. G. M. Moons et al., "PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration," Dec. 31, 2018, American College of Physicians. doi: 10.7326/m18-1377.
- [28] E. Lella et al., "Machine Learning and DWI Brain Communicability Networks for Alzheimer's Disease Detection," Jan. 31, 2020, Multidisciplinary Digital Publishing Institute. doi: 10.3390/app10030934.
- [29] Y. Muhammad, M. Tahir, M. Hayat, and K. T. Chong, "Early and accurate detection and diagnosis of heart disease using intelligent computational model," Nov. 12, 2020, Nature Portfolio. doi: 10.1038/s41598-020-76635-9.

- [30] F. Deshmukh and S. S. Merchant, "Explainable Machine Learning Model for Predicting GI Bleed Mortality in the Intensive Care Unit," Apr. 25, 2020, Lippincott Williams & Wilkins. doi: 10.14309/ajg.0000000000000632.
- [31] Y. Wu and Y. Fang, "Stroke Prediction with Machine Learning Methods among Older Chinese," Mar. 12, 2020, Multidisciplinary Digital Publishing Institute. doi: 10.3390/ijerph17061828.
- [32] T. Meng et al., "Identifying Facial Features and Predicting Patients of Acromegaly Using Three-Dimensional Imaging Techniques and Machine Learning," Jul. 29, 2020, Frontiers Media. doi: 10.3389/fendo.2020.00492.
- [33] M. Syafrudin et al., "A Self-Care Prediction Model for Children with Disability Based on Genetic Algorithm and Extreme Gradient Boosting," Sep. 15, 2020, Multidisciplinary Digital Publishing Institute. doi: 10.3390/math8091590.
- [34] M. Seifollahi, H. Soltanizadeh, A. H. Mehraban, and F. Khamseh, "Alzheimer's disease detection using skeleton data recorded with Kinect camera," Nov. 16, 2019, Springer Science+Business Media. doi: 10.1007/s10586-019-03014-z.
- [35] P. Rajpurkar et al., "CheXaid: deep learning assistance for physician diagnosis of tuberculosis using chest x-rays in patients with HIV," Sep. 09, 2020, Nature Portfolio. doi: 10.1038/s41746-020-00322-2.
- [36] F. Mostafa, E. Hasan, M. Williamson, and H. Khan, "Statistical Machine Learning Approaches to Liver Disease Prediction," Dec. 01, 2021, Multidisciplinary Digital Publishing Institute. doi: 10.3390/livers1040023.
- [37] L. Bloch and C. M. Friedrich, "Comparison of Automated Volume Extraction with FreeSurfer and FastSurfer for Early Alzheimer's Disease Detection with Machine Learning," Jun. 01, 2021. doi: 10.1109/cbms52027.2021.00096.
- [38] X. Qian et al., "A Cardiovascular Disease Prediction Model Based on Routine Physical Examination Indicators Using Machine Learning Methods: A Cohort Study," Jun. 17, 2022, Frontiers Media. doi: 10.3389/fcvm.2022.854287.
- [39] H. Gong, M. Wang, H. Zhang, M. F. Elahe, and M. Jin, "An Explainable AI Approach for the Rapid Diagnosis of COVID-19 Using Ensemble Learning Algorithms," Jun. 21, 2022, Frontiers Media. doi: 10.3389/fpubh.2022.874455.
- [40] K. S. K. Reddy and K. V. Kanimozhi, "Novel Intelligent Model for Heart Disease Prediction using Dynamic KNN (DKNN) with improved accuracy over SVM," Feb. 16, 2022. doi: 10.1109/icbats54253.2022.9758996.
- [41] P. Theerthagiri, "Predictive analysis of cardiovascular disease using gradient boosting based learning and recursive feature elimination technique," Sep. 06, 2022, Elsevier BV. doi: 10.1016/j.iswa.2022.200121.
- [42] S. Liu et al., "Generalizable deep learning model for early Alzheimer's disease detection from structural MRIs," Oct. 17, 2022, Nature Portfolio. doi: 10.1038/s41598-022-20674-x.
- [43] M. Feng et al., "Application of an Interpretable Machine Learning Model to Predict Lymph Node Metastasis in Patients with Laryngeal Carcinoma," Nov. 12, 2022, Hindawi Publishing Corporation. doi: 10.1155/2022/6356399.
- [44] C. B. M. Karthi and A. Kalaivani, "Heart Disease Prediction Based on Age Detection using Novel Logistic Regression over Decision Tree," Feb. 14, 2023, Russian New University. doi: 10.18137/cardiometry.2022.25.17181724.
- [45] S. Hadianti, "Optimization of The Machine Learning Approach using Optuna in Heart Disease Prediction," Sep. 07, 2023. doi: 10.37034/medinftech.v1i3.15.
- [46] J. James P, C. B. E, and R. M, "Improved Coronary Artery Disease Prediction Using Persistence-Based Artificial Bee Colony Optimization Algorithm," May 01, 2023, IJRAR-International Journal of Research and Analytical Reviews (IJRAR).
- [47] A. J. W. Lim et al., "Robust SNP-based prediction of rheumatoid arthritis through machine-learning-optimized polygenic risk score," Feb. 07, 2023, BioMed Central. doi: 10.1186/s12967-023-03939-5.
- [48] G. S. Karthick and P. B. Pankajavalli, "Chronic obstructive pulmonary disease prediction using Internet of things-spiro system and fuzzy-based quantum neural network classifier," Aug. 27, 2022, Elsevier BV. doi: 10.1016/j.tcs.2022.08.021.
- [49] L. Ye and C. Guo, "Heart failure disease prediction and stratification with temporal electronic health records data using patient representation," Jan. 01, 2023, Elsevier BV. doi: 10.1016/j.bbe.2022.12.008.
- [50] S. W. A. Sherazi, H. Zheng, and J. Y. Lee, "A Machine Learning-Based Applied Prediction Model for Identification of Acute Coronary Syndrome (ACS) Outcomes and Mortality in Patients during the Hospital Stay," Jan. 25, 2023, Multidisciplinary Digital Publishing Institute. doi: 10.3390/s23031351.
- [51] F. Li, P. Wu, H. H. Ong, J. F. Peterson, W. Wei, and J. Zhao, "Evaluating and mitigating bias in machine learning models for cardiovascular disease prediction," Jan. 25, 2023, Elsevier BV. doi: 10.1016/j.jbi.2023.104294.
- [52] S. T. Arasteh et al., "Federated Learning for Secure Development of AI Models for Parkinson's Disease Detection Using Speech from Different Languages," Aug. 14, 2023. doi: 10.21437/interspeech.2023-2108.
- [53] P. Roshanibrizi et al., "Analysis of Mitral Regurgitation Jet for Rheumatic Heart Disease Detection in Doppler Echocardiography," Apr. 18, 2023. doi: 10.1109/isbi53787.2023.10230680.
- [54] F. R. Anas et al., "Early-stage parkinson's disease detection based on action unit derivatives," Sep. 28, 2023, olloque en TélésANTé et dispositifs biomédicaux (JETSAN).
- [55] K. Brown et al., "Using Artificial Intelligence for Rheumatic Heart Disease Detection by Echocardiography: Focus on Mitral Regurgitation," Jan. 16, 2024, Wiley. doi: 10.1161/jaha.123.031257.
- [56] M. Vázquez-Marrufo, E. Sarrias-Arrabal, M. Garcia-Torres, R. Martín-Clemente, and G. Izquierdo, "A systematic review of the application of machine-learning algorithms in multiple sclerosis," *Neurología (English Edition)*, vol. 38, no. 8. Elsevier BV, p. 577, Jul. 14, 2022. doi: 10.1016/j.nrleng.2020.10.013.
- [57] N. Kaieski, C. A. da Costa, R. da R. Righi, P. S. Lora, and B. M. Eskofier, "Application of artificial intelligence methods in vital signs analysis of hospitalized patients: A systematic literature review," Aug. 06, 2020, Elsevier BV. doi: 10.1016/j.asoc.2020.106612.
- [58] A. N. Reiz, M. Á. A. de la Hoz, and M. S. García, "Big Data Analysis y Machine Learning en medicina intensiva," *Medicina Intensiva*, vol. 43, no. 7. Elsevier BV, p. 416, Dec. 24, 2018. doi: 10.1016/j.medin.2018.10.007.
- [59] S. Romiti, M. Vinciguerra, W. Saade, I. A. Cortajarena, and E. Greco, "Artificial Intelligence (AI) and Cardiovascular Diseases: An Unexpected Alliance," *Cardiology Research and Practice*, vol. 2020. Hindawi Publishing Corporation, p. 1, Jun. 27, 2020. doi: 10.1155/2020/4972346.
- [60] S. Sanchez-Martinez et al., "Machine Learning for Clinical Decision-Making: Challenges and Opportunities in Cardiovascular Imaging," *Frontiers in Cardiovascular Medicine*, vol. 8. Frontiers Media, Jan. 04, 2022. doi: 10.3389/fcvm.2021.765693.
- [61] S. Subramani et al., "Cardiovascular diseases prediction by machine learning incorporation with deep learning," Apr. 17, 2023, Frontiers Media. doi: 10.3389/fmed.2023.1150933.
- [62] Á. J. Soto et al., "Diagnóstico de metástasis en pacientes con cáncer de próstata resistente a la castración: algoritmo de decisión en pruebas

- de imagen,” *Actas Urológicas Españolas*, vol. 43, no. 2. Elsevier BV, p. 55, Aug. 03, 2018. doi: 10.1016/j.acuro.2018.05.011.
- [64] J. A. Cruz and D. S. Wishart, “Applications of Machine Learning in Cancer Prediction and Prognosis,” Jan. 01, 2006, SAGE Publishing. doi: 10.1177/117693510600200030.
- [65] L. B. Cardoso et al., “Machine learning for predicting survival of colorectal cancer patients,” Jun. 01, 2023, *Nature Portfolio*. doi: 10.1038/s41598-023-35649-9.
- [66] R. Alloro and E. Sinagra, “Artificial intelligence and colorectal cancer: How far can you go?,” Apr. 26, 2021. doi: 10.35713/aic.v2.i2.7.
- [67] N. Jahan, Md. S. Anower, and R. Hassan, “Automated Diagnosis of Pneumonia from Classification of Chest X-Ray Images using EfficientNet,” Feb. 27, 2021. doi: 10.1109/iciict4sd50815.2021.9397055.
- [68] Sato; Masaya; Morimoto; Kentaro; Kajihara; Shigeki; Tateishi; Ryosuke; Shiina; Shuichiro; Koike; Kazuhiko; Yatomi; Yutaka, “Machine-learning Approach for the Development of a Novel Predictive Model for the Diagnosis of Hepatocellular Carcinoma - Scientific Reports.” Feb. 2023. Accessed: Dec. 10, 2024. [Online]. Available: <https://www.nature.com/articles/s41598-019-44022-8>
- [69] R. A. Mancisidor, “Using multimodal learning and deep generative models for corporate bankruptcy prediction,” Jan. 01, 2022, Cornell University. doi: 10.48550/arxiv.2211.08405.
- [70] Y. Liu et al., “On the Hidden Mystery of OCR in Large Multimodal Models,” Jan. 01, 2023, Cornell University. doi: 10.48550/arXiv.2305.
- [71] R. M. Ferrer, “Valoraciones éticas para una inteligencia artificial adecuada a la privacidad,” Dec. 30, 2021, Spanish National Research Council. doi: 10.3989/arbor.2021.802006.