

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Posgrados

Desarrollo de un sistema de inteligencia artificial para la personalización predictiva de protocolos de exámenes ocupacionales: Modelo de gestión e implementación para PYMES ecuatorianas

Mecanismo de Titulación: Tesis en torno a una hipótesis o problema de investigación y su contrastación

María Alexandra Guano Zambrano

**Miguel Ángel Moreira García
Director de Trabajo de Titulación**

Trabajo de titulación de posgrado presentado como requisito para la obtención del título de Magister en Gerencia en Salud

Quito, julio del 2025

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

COLEGIO DE POSGRADOS

HOJA DE APROBACIÓN DE TRABAJO DE TITULACIÓN

Desarrollo de un sistema de inteligencia artificial para la personalización predictiva de protocolos de exámenes ocupacionales: Modelo de gestión e implementación para PYMES ecuatorianas

Guano Zambrano María Alexandra

Nombre del Director del Programa:

Miguel Ángel Moreira García

Título académico:

MSc.

Director del programa de:

Maestría en Gerencia en Salud

Nombre del Decano del colegio Académico:

Ximena Patricia Garzón Villalba

Título académico:

Ph.D

Decano del Colegio:

Escuela de Salud Pública y Nutrición

Nombre del Decano del Colegio de Posgrados:

Dario Niebieskwiat

Título académico:

Ph.D

Quito, julio 2025

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombre del estudiante: María Alexandra Guano Zambrano

Código de estudiante: 00343969

C.I.: 1719274316

Lugar y fecha: Quito, 14 de julio de 2025

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETheses>.

UNPUBLISHED DOCUMENT

Note: The following graduation project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETheses>.

DEDICATORIA

A todos los profesionales de la salud ocupacional que trabajan incansablemente por crear ambientes laborales más seguros y saludables, y a las pequeñas y medianas empresas que buscan innovar en el cuidado de sus trabajadores a través de la implementación de tecnologías inteligentes.

AGRADECIMIENTOS

Expreso mi agradecimiento al Dr. Roberto Montenegro, tutor de este trabajo de titulación, por su invaluable orientación metodológica y apoyo constante durante todo el proceso de investigación. A los profesionales de HIRAM por facilitar el acceso a la base de datos de historias clínicas ocupacionales que constituyó el fundamento empírico de esta investigación. Al equipo de médicos ocupacionales que participaron en la validación del sistema, aportando su experiencia clínica para verificar una aplicabilidad práctica del algoritmo con IA. Finalmente, reconozco a las empresas participantes que confiaron en este proceso con mira para aplicar la innovación en la salud ocupacional.

RESUMEN

La medicina ocupacional en Ecuador enfrenta desafíos significativos relacionados con la ausencia de protocolos específicos para la determinación de exámenes médicos ocupacionales, resultando en la aplicación indiscriminada de pruebas diagnósticas que incrementan sustancialmente los costos empresariales sin generar valor clínico proporcional. Este estudio desarrolló un sistema de inteligencia artificial para la personalización predictiva de protocolos de exámenes ocupacionales, implementando un modelo de gestión específicamente diseñado para PYMES ecuatorianas.

La investigación empleó una metodología mixta que integró análisis cuantitativo de 679 historias clínicas ocupacionales de la empresa HIRAM, recopiladas entre febrero 2024 y mayo 2025, con validación cualitativa por expertos médicos ocupacionales. La muestra incluyó trabajadores de seis sectores económicos: construcción, agricultura, sector farmacéutico, manufactura de plásticos, servicios exequiales y ferretería. El sistema desarrollado integra algoritmos de machine learning que procesan variables predictoras incluyendo edad, tipo de exposición a riesgos laborales (ergonómicos, físicos, químicos, psicosociales y mecánicos), nivel de exposición, antecedentes patológicos personales y antecedentes laborales.

El modelo predictivo demostró capacidad para reducir costos asociados a exámenes médicos ocupacionales en un 38.7% promedio, desde un costo real base de \$142.30 USD por trabajador/año hasta \$87.20 USD, generando un ahorro real de \$55.10 USD por trabajador/año. Estos costos fueron validados mediante comparación con tres proveedores ecuatorianos: DISALAB (proveedor primario), MEDIAVANS S.A. (tarifario oficial 2024) y AGNOSIS LAB (proyección 2025) (HIRAM, 2025).

La validación por expertos mostró concordancia del 84% con las recomendaciones del sistema, mientras que las métricas de rendimiento alcanzaron sensibilidad del 87.3% y especificidad del 85.4%. El dashboard gerencial desarrollado facilita la toma de decisiones mediante visualización de indicadores clave de rendimiento (KPIs) específicos para la gestión de salud ocupacional en PYMES.

Los resultados evidencian que la implementación de sistemas de inteligencia artificial en medicina ocupacional representa una oportunidad estratégica para optimizar recursos financieros y humanos, mejorar la detección temprana de enfermedades ocupacionales y establecer protocolos preventivos basados en evidencia científica. Para empresas de 35 empleados típicas en Ecuador, el sistema genera ahorros anuales de \$1,928 USD, permitiendo recuperar la inversión inicial de \$4,800 USD en 2.5 años.

Palabras clave: inteligencia artificial, medicina ocupacional, protocolos personalizados, machine learning, gestión de salud, PYMES Ecuador, optimización de costos, sistemas predictivos

ABSTRACT

Occupational medicine in Ecuador faces significant challenges related to the absence of specific protocols for determining occupational medical examinations, resulting in the indiscriminate application of diagnostic tests that substantially increase business costs without generating proportional clinical value. This study developed an artificial intelligence system for predictive personalization of occupational examination protocols, implementing a management model specifically designed for Ecuadorian SMEs.

The research employed a mixed methodology that integrated quantitative analysis of 679 occupational medical records from HIRAM company, collected between February 2024 and May 2025, with qualitative validation by occupational medical experts. The sample included workers from six economic sectors: construction, agriculture, pharmaceutical sector, plastic manufacturing, funeral services, and hardware stores. The developed system integrates machine learning algorithms that process predictive variables including age, type of exposure to occupational risks (ergonomic, physical, chemical, psychosocial, and mechanical), exposure level, personal pathological history, and occupational history.

The predictive model demonstrated the ability to reduce costs associated with occupational medical examinations by an average of 38.7%, from a real base cost of \$142.30 USD per worker/year to \$87.20 USD, generating a real savings of \$55.10 USD per worker/year. These costs were validated through comparison with three Ecuadorian providers: DISALAB (primary provider), MEDIAVANS S.A. (official 2024 tariff), and AGNOSIS LAB (2025 projection) (HIRAM, 2025).

Expert validation showed 84% concordance with system recommendations, while performance metrics achieved 87.3% sensitivity and 85.4% specificity. The developed management

dashboard facilitates decision-making through visualization of key performance indicators (KPIs) specific to occupational health management in SMEs.

The results demonstrate that implementing artificial intelligence systems in occupational medicine represents a strategic opportunity to optimize financial and human resources, improve early detection of occupational diseases, and establish evidence-based preventive protocols. For typical 35-employee companies in Ecuador, the system generates annual savings of \$1,928 USD, allowing recovery of the initial investment of \$4,800 USD in 2.5 years.

Key words: artificial intelligence, occupational medicine, personalized protocols, machine learning, health management, Ecuador SMEs, cost optimization, predictive systems.

TABLA DE CONTENIDO

PORTADA	1
HOJA DE APROBACIÓN	3
DECLARACIÓN DE AUTORÍA Y DERECHOS	4
DEDICATORIA	6
AGRADECIMIENTOS	7
RESUMEN	8
ABSTRACT	9
CAPÍTULO I. INTRODUCCIÓN	15
CAPÍTULO II. MARCO TEÓRICO	20
CAPÍTULO III. METODOLOGÍA	25
CAPÍTULO IV. DESARROLLO DEL SISTEMA DE IA	36
CAPÍTULO V. ANÁLISIS DE DATOS	62
CAPÍTULO VI. MODELO DE GESTIÓN PARA PYMES	68
CAPÍTULO VII. CONCLUSIONES	76
REFERENCIAS BIBLIOGRÁFICAS	78
ÍNDICE DE TABLAS	13
ÍNDICE DE FIGURAS	14
ÍNDICE DE ANEXOS	83

ÍNDICE DE TABLAS

Tabla 1. Distribución demográfica de la muestra	27
Tabla 2. Distribución por sectores económicos	28
Tabla 3. Características de exposición a riesgos laborales	28
Tabla 4. Frecuencia optimizada de exámenes ocupacionales según IA	66
Tabla 5. Importancia relativa de variables predictoras (SHAP values)	67

ÍNDICE DE FIGURAS

Figura 1. Arquitectura del sistema de IA aplicado en medicina ocupacional	37
Figura 2. Curva ROC del modelo Random Forest	38
Figura 3. Matriz de confusión del sistema IA	38
Figura 4. Limpieza de datos en el preprocesamiento clínico	39
Figura 5. Codificación de variables categóricas	40
Figura 6. Barras de importancia de variables (SHAP)	41
Figura 7. Visualización del modelo SVM	43
Figura 8. Curva de aprendizaje de Gradient Boosting	44
Figura 9. Esquema de red neuronal (MLP)	46
Figura 10. Sistema de ensamble y selección de algoritmos	48
Figura 11. Sistema de explicabilidad con SHAP	51
Figura 12. Validación de usabilidad con escala SUS	54
Figura 13. Validación clínica A/B	56
Figura 14. Dashboard ejecutivo de salud ocupacional	58

CAPITULO I: INTRODUCCIÓN

La medicina ocupacional constituye una disciplina fundamental para la protección de la salud de los trabajadores, estableciendo mecanismos preventivos que permiten identificar, evaluar y controlar los riesgos derivados de la actividad laboral. En el contexto ecuatoriano, esta especialidad médica falta mayor relevancia con la apertura de posgrados, sin embargo, médicos que se han formado con maestrías en salud ocupacional han adquirido relevancia creciente debido a la evolución del marco normativo y la mayor conciencia empresarial sobre la importancia de implementar programas integrales de salud ocupacional (Ministerio de Salud Pública, 2023).

Este proyecto no implica el desarrollo final de un software clínico operativo, sino una validación investigativa del modelo predictivo con algoritmos de inteligencia artificial, sobre el cual podrá construirse un sistema aplicable posteriormente. Su propósito es ofrecer una base metodológica sólida para una futura herramienta útil tanto para el médico ocupacional como para las PYMES ecuatorianas.

Los exámenes médicos ocupacionales deben representar una herramienta esencial dentro de los sistemas de vigilancia de la salud laboral, permitiendo la detección temprana de alteraciones relacionadas con la exposición a factores de riesgo ocupacional. Sin embargo, la práctica actual en Ecuador evidencia limitaciones significativas en la personalización de protocolos diagnósticos, resultando en la aplicación de enfoques estandarizados que no consideran las particularidades individuales de cada trabajador ni las características específicas de exposición de cada puesto de trabajo (Instituto Ecuatoriano de Seguridad Social, 2024).

Esta problemática se acentúa particularmente en las pequeñas y medianas empresas (PYMES), las cuales representan el 90.8% del tejido empresarial ecuatoriano según datos del

Instituto Nacional de Estadística y Censos (INEC, 2023). Las PYMES enfrentan desafíos únicos en la gestión de salud ocupacional, incluyendo limitaciones presupuestarias, recursos humanos especializados reducidos y dificultades para acceder a tecnologías avanzadas de gestión sanitaria.

La ausencia de protocolos específicos para la determinación de exámenes médicos ocupacionales, establecida en la normativa ecuatoriana vigente, ha generado una práctica clínica caracterizada por la aplicación indiscriminada de pruebas diagnósticas no relacionadas a los riesgos de exposición. Esta situación resulta en incrementos sustanciales de costos para las empresas, sin generar necesariamente un valor clínico proporcional en términos de detección temprana de patologías ocupacionales (Acuerdo Ministerial 135, 2017).

Según el análisis de costos reales validado con tres proveedores ecuatorianos principales (DISALAB, MEDIAVANS y AGNOSIS), el costo promedio actual en exámenes por trabajador es de \$142.30 USD anuales, representando entre el 3% y 8% del presupuesto de recursos humanos en PYMES ecuatorianas. Esta inversión, aunque necesaria desde la perspectiva legal y ética, requiere optimización para maximizar su impacto en la salud de los trabajadores

Justificación del Problema

La justificación de esta investigación se fundamenta en la convergencia de múltiples factores que configuran una oportunidad estratégica para la innovación en medicina ocupacional. Desde la perspectiva epidemiológica, Ecuador presenta indicadores preocupantes en términos de enfermedades ocupacionales, con tasas de notificación que sugieren subregistro significativo debido a limitaciones en los sistemas de detección temprana (Organización Internacional del Trabajo, 2024).

La dimensión económica del problema adquiere particular relevancia considerando que los costos asociados a exámenes médicos ocupacionales representan entre el 3% y 8% del

presupuesto total de recursos humanos en PYMES ecuatorianas, según datos del Ministerio de Trabajo (2024). De acuerdo con un análisis de mercado de tres proveedores los costos asociados a exámenes médicos ocupacionales, validados en \$142.30 USD por trabajador/año, representan una inversión alta que requiere optimización sin comprometer la calidad clínica. La revolución tecnológica en el ámbito sanitario, particularmente los avances en inteligencia artificial y machine learning, presenta oportunidades sin precedentes para transformar la práctica de la medicina ocupacional. Los sistemas de soporte a la decisión clínica han demostrado eficacia en múltiples especialidades médicas, evidenciando potencial para mejorar la precisión diagnóstica, optimizar recursos y personalizar tratamientos (Rajkomar et al., 2019).

Planteamiento del Problema

El problema central que aborda esta investigación se articula en la pregunta: ¿Cómo puede un sistema de inteligencia artificial optimizar la personalización de protocolos de exámenes médicos ocupacionales para mejorar la eficiencia diagnóstica y reducir costos en PYMES ecuatorianas, operando con los costos reales del mercado ecuatoriano validados en \$142.30 USD por trabajador/año?

Esta interrogante principal se desglosa en las siguientes preguntas específicas:

1. ¿Cuáles son los patrones de exposición a riesgos laborales y su relación con hallazgos patológicos en exámenes médicos ocupacionales en PYMES ecuatorianas?
2. ¿Qué variables predictoras resultan más relevantes para la personalización de protocolos de exámenes ocupacionales?
3. ¿Cómo puede desarrollarse un modelo predictivo que optimice la selección de exámenes médicos basado en características individuales del trabajador?

4. ¿Qué nivel de precisión diagnóstica y optimización de costos puede alcanzarse mediante la implementación del sistema propuesto?
5. ¿Cuál es el modelo de gestión más apropiado para la implementación del sistema en PYMES ecuatorianas?

Objetivos de la Investigación

Objetivo General: Desarrollar un sistema de inteligencia artificial para la personalización predictiva de protocolos de exámenes médicos ocupacionales, implementando un modelo de gestión específicamente diseñado para PYMES ecuatorianas que optimice la eficiencia diagnóstica y reduzca costos asociados, validado con costos reales del mercado ecuatoriano.

Objetivos Específicos:

1. Validar costos reales de exámenes médicos ocupacionales mediante análisis comparativo de tres proveedores ecuatorianos
2. Analizar patrones de exposición a riesgos laborales y su correlación con hallazgos patológicos en exámenes médicos ocupacionales utilizando una base de datos de 679 historias clínicas de PYMES ecuatorianas.
3. Desarrollar un modelo predictivo basado en algoritmos de machine learning que integre variables como edad, tipo de exposición, nivel de riesgo, antecedentes patológicos y antecedentes laborales.
4. Implementar un sistema de soporte a la decisión clínica que genere recomendaciones personalizadas para la selección de exámenes médicos ocupacionales.
5. Validar la precisión diagnóstica y efectividad del sistema mediante evaluación por expertos médicos ocupacionales y análisis de métricas de rendimiento.

6. Diseñar un modelo de gestión para la implementación del sistema en PYMES ecuatorianas, incluyendo análisis de costo-beneficio y herramientas de presentación gerencial.
7. Desarrollar herramientas de visualización y dashboards gerenciales que faciliten la toma de decisiones en gestión de salud ocupacional.

Hipótesis de Trabajo

La implementación de un sistema de inteligencia artificial para la personalización predictiva de protocolos de exámenes médicos ocupacionales, basado en el análisis de variables individuales del trabajador y características de exposición laboral, permitirá reducir significativamente los costos asociados a exámenes médicos (objetivo: >35%) mientras mantiene o mejora la capacidad de detección temprana de enfermedades ocupacionales, constituyendo un modelo de gestión viable y escalable para PYMES ecuatorianas.

Alcance y Limitaciones

Alcance: Este estudio abarca el desarrollo completo de un sistema de inteligencia artificial aplicado a medicina ocupacional, incluyendo la fase de análisis de datos, desarrollo del modelo predictivo, validación clínica y diseño del modelo de gestión. El ámbito geográfico se circunscribe a Ecuador, con enfoque específico en PYMES de seis sectores económicos: construcción, agricultura, sector farmacéutico, manufactura de plásticos, servicios exequiales y ferretería.

Limitaciones: Las limitaciones identificadas incluyen: (1) base de datos proveniente de una sola empresa de asesoría, lo que podría afectar la generalización de resultados; (2) período de recolección de datos de 15 meses, que puede no capturar variaciones estacionales; (3) ausencia de seguimiento longitudinal para evaluar la evolución de salud a largo plazo; (4) validación limitada a opinión de expertos sin implementación prospectiva completa.

CAPITULO II: MARCO TEÓRICO DE LA MEDICINA OCUPACIONAL

La medicina ocupacional, definida como la especialidad médica dedicada a la prevención, diagnóstico, tratamiento y rehabilitación de enfermedades y lesiones relacionadas con el trabajo, ha experimentado una evolución significativa en los últimos años. Según la Organización Mundial de la Salud (2021), esta disciplina trasciende el enfoque clínico tradicional para incorporar perspectivas epidemiológicas, toxicológicas y de gestión de riesgos. Sin dejar a un lado el reconocimiento de factores individuales como edad, género, susceptibilidad genética y antecedentes médicos que influyen significativamente en la respuesta a exposiciones laborales (Occupational Safety and Health Administration, 2023). Esta perspectiva individualizada contrasta con enfoques tradicionales basados en protocolos estandarizados aplicados indiscriminadamente a poblaciones trabajadoras, llegando a costos promedio de \$142.30 USD por trabajador/año.

En Ecuador, el marco normativo establece la obligatoriedad de exámenes médicos ocupacionales a través del Acuerdo Ministerial 135 (2017) y sus reformas posteriores, mostrando limitaciones en la especificación de criterios a aplicar, dejando a la interpretación a los profesionales de la salud ocupacional sobre los tipos de exámenes médicos ocupacionales de pre-ingreso, y post-ocupacionales (Ministerio de Trabajo, 2024).

La evidencia científica respalda la efectividad de enfoques personalizados en medicina ocupacional. Krishnan et al. (2022) demostraron que protocolos individualizados mejoran la detección temprana de enfermedades ocupacionales en un 34% comparado con protocolos estandarizados. Similarmente, la investigación de García-Molina et al. (2023) evidenció reducciones del 28% en costos asociados a exámenes médicos ocupacionales mediante la implementación de sistemas de decisión basados en perfiles de riesgo individual.

Inteligencia Artificial en Medicina Ocupacional

La aplicación de inteligencia artificial en medicina ocupacional representa un área emergente con potencia con la innovación. Los sistemas de machine learning han demostrado capacidad superior para procesar grandes volúmenes de datos médicos y identificar patrones complejos no evidentes mediante análisis tradicionales (Chen et al., 2023).

Los algoritmos de aprendizaje automático, particularmente random forests y support vector machines, han mostrado eficacia en la predicción de riesgos ocupacionales. La investigación de Thompson et al. (2024) implementó algoritmos de deep learning para predecir la probabilidad de desarrollo de neumoconiosis en trabajadores expuestos a sílice, alcanzando precisión del 92.3% en la identificación de casos de alto riesgo.

El análisis de big data en medicina ocupacional ha revelado asociaciones previamente desconocidas entre factores de exposición y resultados de salud. Kumar et al. (2023) utilizaron técnicas de natural language processing para analizar 50,000 historias clínicas ocupacionales, identificando 23 nuevos factores de riesgo para enfermedades respiratorias ocupacionales.

La implementación de sistemas de inteligencia artificial en medicina ocupacional enfrenta desafíos relacionados con la calidad de datos, interpretabilidad de algoritmos y aceptación por parte de profesionales médicos. Sin embargo, la evidencia sugiere que estos sistemas pueden complementar efectivamente el juicio clínico, mejorando la precisión diagnóstica y optimizando recursos (Rodriguez-Martinez et al., 2024).

Sistemas de Soporte a la Decisión Clínica

Los sistemas de soporte a la decisión clínica (CDSS) han evolucionado desde herramientas básicas de recordatorio hasta sistemas inteligentes capaces de generar recomendaciones personalizadas basadas en análisis de datos complejos. La definición contemporánea de CDSS incluye sistemas que proporcionan información específica del

paciente, filtrada inteligentemente, para mejorar la toma de decisiones clínicas (Sutton et al., 2020).

La efectividad de los CDSS en medicina ocupacional ha sido documentada en estudios. Patel et al. (2023) implementaron un sistema de soporte a la decisión para la selección de exámenes respiratorios en trabajadores expuestos a asbestos, demostrando mejoras del 41% en la detección temprana de fibrosis pulmonar y reducción del 33% en exámenes innecesarios.

Los componentes esenciales de un CDSS efectivo incluyen: (1) base de conocimiento médico actualizada, (2) motor de inferencia capaz de procesar reglas complejas, (3) interfaz de usuario intuitiva, y (4) sistema de retroalimentación para aprendizaje continuo. La integración de estos componentes determina la utilidad clínica del sistema (Johnson et al., 2024).

La aceptación de CDSS por parte de profesionales médicos depende de factores como facilidad de uso, transparencia en el proceso de decisión, y capacidad de personalización según preferencias individuales. La investigación de Williams et al. (2023) identificó que el 78% de médicos ocupacionales prefieren sistemas que explican el razonamiento detrás de las recomendaciones, mientras que el 65% valora la capacidad de ajustar parámetros según criterio clínico.

Personalización en Protocolos Médicos

La personalización de protocolos médicos representa una evolución natural de la medicina de precisión, adaptando intervenciones diagnósticas y terapéuticas a características individuales del paciente. En medicina ocupacional, esta personalización considera factores como susceptibilidad individual, patrón de exposición, edad, género y antecedentes médicos (National Institute for Occupational Safety and Health, 2024).

La medicina de precisión ocupacional ha mostrado resultados prometedores en la prevención de enfermedades relacionadas con el trabajo. Anderson et al. (2023) desarrollaron un modelo de estratificación de riesgo personalizado para trabajadores de la industria química, logrando reducir la incidencia de dermatitis ocupacional en un 47% mediante la implementación de medidas preventivas específicas.

Los biomarcadores de susceptibilidad individual han emergido como herramientas valiosas para la personalización de protocolos ocupacionales. La investigación de Liu et al. (2024) identificó polimorfismos genéticos asociados con mayor susceptibilidad a toxicidad por metales pesados, permitiendo la implementación de programas de vigilancia intensificada en trabajadores de alto riesgo.

La personalización efectiva requiere sistemas capaces de integrar múltiples fuentes de información, incluyendo datos clínicos, exposicionales, genéticos y ambientales. Los algoritmos de machine learning han demostrado capacidad superior para procesar esta información multidimensional y generar recomendaciones personalizadas (Zhang et al., 2023).

Gestión de Salud Ocupacional en PYMES

Las pequeñas y medianas empresas enfrentan desafíos únicos en la gestión de salud ocupacional, incluyendo limitaciones de recursos, acceso restringido a servicios especializados y dificultades para implementar programas integrales de vigilancia de la salud (European Agency for Safety and Health at Work, 2024).

La investigación de Morales et al. (2023) analizó la gestión de salud ocupacional en 500 PYMES latinoamericanas, identificando que el 67% carece de médicos ocupacionales dedicados, mientras que el 82% depende de servicios externos para exámenes médicos. Esta situación genera oportunidades para la implementación de sistemas automatizados de soporte a la decisión.

Los costos asociados a exámenes médicos ocupacionales representan una proporción significativa del presupuesto de salud en PYMES. El estudio de Fernández et al. (2024) documentó que empresas con menos de 50 empleados destinan entre 4% y 12% de su presupuesto de recursos humanos a exámenes médicos ocupacionales, con alta variabilidad según el sector económico.

La implementación de tecnologías digitales en PYMES ecuatorianas ha mostrado crecimiento sostenido, con el 34% de empresas utilizando algún tipo de sistema de gestión digital según datos del Ministerio de Telecomunicaciones (2024). Esta tendencia sugiere receptividad creciente hacia soluciones tecnológicas innovadoras en gestión de salud ocupacional.

Contexto Ecuatoriano

En el ámbito ecuatoriano presenta características específicas que influyen en la implementación de sistemas de inteligencia artificial en medicina ocupacional. La estructura empresarial del país, dominada por PYMES que representan el 90.8% del ámbito empresarial, configura un escenario particular para la gestión de salud ocupacional (Instituto Nacional de Estadística y Censos, 2023).

Los indicadores epidemiológicos revelan subestimación significativa de enfermedades ocupacionales. Según datos del Instituto Ecuatoriano de Seguridad Social (2024), la tasa de notificación de enfermedades ocupacionales es de 2.3 por 10,000 trabajadores, significativamente inferior a promedios internacionales que oscilan entre 8.7 y 15.4 por 10,000 trabajadores.

La infraestructura tecnológica ecuatoriana presenta oportunidades y desafíos para la implementación de sistemas de inteligencia artificial. El índice de desarrollo digital del país alcanzó 6.1 puntos en 2023, ubicándose en posición intermedia regional según el Banco Interamericano de Desarrollo (2024).

CAPITULO III: METODOLOGÍA Y DISEÑO DE LA INVESTIGACIÓN

Modelo de Investigación

Esta investigación se fundamenta en un modelo post-positivista que reconoce la complejidad propia de los acontecimientos de salud ocupacional, adoptando un enfoque práctico que prioriza la utilidad práctica de los hallazgos científicos. El diseño metodológico integra elementos cuantitativos para el análisis de datos médicos y componentes cualitativos para la validación clínica, configurando una aproximación metodológica mixta que maximiza la robustez de los resultados (Creswell & Plano Clark, 2017).

La selección de este modelo se justifica por la naturaleza aplicada de la investigación, que busca desarrollar una solución tecnológica específica para un problema real en el contexto de la salud ocupacional ecuatoriana. El enfoque práctico permite flexibilidad metodológica para adaptar técnicas de investigación según las necesidades específicas de cada fase del estudio (Tashakkori & Teddlie, 2010).

Diseño de Investigación

El estudio adopta un diseño descriptivo-analítico con componente tecnológico, estructurado en cuatro fases secuenciales:

1. Análisis retrospectivo de datos médicos ocupacionales,
2. Validación de costos de mercado de proveedores de exámenes
3. Desarrollo del sistema de inteligencia artificial,
4. Validación clínica del sistema,
5. Diseño del modelo de gestión

Esta estructura secuencial permite la construcción progresiva de conocimiento, desde la comprensión del problema hasta la propuesta de solución (Hernández-Sampieri et al., 2014).

La investigación emplea un enfoque transversal para el análisis de datos existentes, complementado con elementos prospectivos para la validación del sistema desarrollado. El período de estudio abarca desde febrero 2024 hasta mayo 2025, proporcionando una ventana temporal suficiente para capturar variaciones estacionales y patrones epidemiológicos relevantes.

El estudio se circunscribe al contexto ecuatoriano, específicamente enfocado en PYMES ubicadas en la región Sierra, con representación de seis sectores económicos: construcción, agricultura, sector farmacéutico, manufactura de plásticos, servicios exequiales y ferretería.

Población y Muestra

La población objetivo comprende trabajadores de PYMES ecuatorianas sometidos a exámenes médicos ocupacionales periódicos. La población accesible incluye trabajadores atendidos por la empresa HIRAM durante el período de estudio, representando una muestra diversificada del panorama laboral de PYMES ecuatorianas.

La muestra total consiste en 679 historias clínicas ocupacionales completas, determinada por la disponibilidad de registros en la base de datos de HIRAM. Este tamaño muestral supera los requerimientos mínimos para análisis estadístico multivariado, proporcionando poder estadístico adecuado para detectar asociaciones significativas entre variables predictoras y resultados de interés.

Criterios de Selección:

Criterios de Inclusión:

- Historias clínicas ocupacionales completas con información sobre variables de interés
- Trabajadores de PYMES con actividad económica continua durante el período de estudio
- Registros que incluyan resultados de exámenes médicos ocupacionales específicos

- Documentación completa de exposición a riesgos laborales según metodología estandarizada

Criterios de Exclusión:

- Historias clínicas con datos faltantes superiores al 10% en variables críticas
- Trabajadores con patologías preexistentes autoinmunes que interfieran con el análisis de causalidad ocupacional
- Registros duplicados o con inconsistencias evidentes en la información
- Trabajadores con menos de 12 meses de antigüedad en el puesto de trabajo

Caracterización de la Muestra

Distribución Demográfica: La muestra de 679 trabajadores presenta características demográficas representativas del panorama laboral ecuatoriano. La edad promedio fue 34.2 años ($DE \pm 8.7$), con rango etario de 18 a 65 años. La distribución por género mostró predominio masculino (64.2%, n=436) sobre femenino (35.8%, n=243), consistente con la composición típica de los sectores económicos analizados.

Tabla 1

Distribución demográfica de la muestra

Género	Porcentaje	n
<i>masculino</i>	64,2%	n=436
<i>femenino</i>	35.8%,	n=243

Nota. Datos obtenidos de la base de historias clínicas ocupacionales de 679 trabajadores recolectadas entre febrero 2024 y mayo 2025 por HIRAM.

Distribución por Sectores Económicos:

Tabla 2*Distribución por sectores económicos*

<i>Sector económico</i>	<i>Porcentaje</i>	<i>N</i>
<i>Construcción</i>	23.0%	n=156
<i>Agricultura</i>	18.0%	n=122
<i>Sector farmacéutico (visitadores médicos)</i>	15.0%	n=102
<i>Manufactura de productos plásticos</i>	16.0%	n=109
<i>Servicios exequiales</i>	14.0%	n=95
<i>Ferretería/comercio</i>	14.0%	n=95

Nota. Elaboración propia con base en datos de trabajadores atendidos por la empresa HIRAM durante el período de estudio.

Características de Exposición: La evaluación de riesgos laborales reveló distribución heterogénea de exposiciones en los puestos de trabajo: riesgos ergonómicos (42.3%), riesgos físicos (28.7%), riesgos químicos (18.9%), riesgos psicosociales (16.2%) y riesgos mecánicos (14.7%). El 63.4% de trabajadores presentó exposición de nivel alto, 28.6% nivel medio y 8.0% nivel bajo.

Tabla 3*Características de exposición a riesgos laborales*

<i>Riesgos Laborales</i>	<i>Porcentaje</i>
<i>Riesgos ergonómicos</i>	16.2%
<i>Riesgos físicos</i>	28.7%

Riesgos químicos	18.9%
Riesgos psicosociales	16.2%
Riesgos mecánicos	14.7%

Nota. Clasificación realizada según riesgos laborales aplicada por el equipo médico de HIRAM en los seis sectores económicos analizados.

Variables del Estudio

Variables Independientes:

1. **Edad:** Variable numérica continua expresada en años completos
2. **Género:** Variable categórica dicotómica (masculino/femenino)
3. **Tipo de exposición a riesgos laborales:** Variable categórica múltiple con cinco categorías
4. **Nivel de exposición:** Variable ordinal (bajo/medio/alto)
5. **Antecedentes patológicos personales:** Variable categórica binaria
6. **Antecedentes laborales:** Variable categórica que incluye tiempo de exposición y trabajos previos
7. **Tiempo en puesto actual:** Variable numérica continua expresada en meses
8. **Nivel educativo:** Variable ordinal (primaria/secundaria/superior)

Variables Dependientes:

1. **Exámenes médicos realizados:** Variable categórica múltiple incluyendo:
 - Exámenes de laboratorio básicos (biometría, glucosa, perfil lipídico)
 - Exámenes especializados (optometría, audiometría, espirometría)
 - Estudios radiológicos (radiografías de columna, extremidades)

- Estudios ultrasonográficos (ecografía de muñeca)

- 2. Hallazgos patológicos: Variable categórica binaria (normal/anormal)**
- 3. Costos asociados:** Variable numérica continua expresada en USD
- 4. Tiempo de procesamiento:** Variable numérica continua expresada en minutos

Instrumentos de Recolección de Datos

Base de Datos Primaria: Se utilizó la base de datos electrónica de HIRAM, estructurada en formato Excel con campos codificados de clasificación médica. La base incluye 23 variables principales con subvariables específicas para cada tipo de examen médico.

Protocolo de Extracción: Se desarrolló un protocolo estandarizado de extracción de datos que incluye:

- Verificación de integridad de registros
- Codificación de variables categóricas
- Normalización de valores numéricos
- Identificación y manejo de valores faltantes
- Validación de consistencia interna

Instrumentos de Validación: Para la validación clínica se diseñaron instrumentos específicos:

- Cuestionario de evaluación de concordancia clínica (α de Cronbach = 0.87)
- Escala de usabilidad del sistema (System Usability Scale - SUS)
- Rúbrica de evaluación de recomendaciones médicas

Desarrollo del sistema de inteligencia artificial

Arquitectura del Sistema: El sistema se desarrollaría siguiendo arquitectura modular que incluye:

- Módulo de preprocesamiento de datos
- Motor de machine learning con algoritmos múltiples
- Sistema de validación y testing
- Interfaz de usuario para médicos ocupacionales
- Dashboard gerencial para tomadores de decisiones

Algoritmos a Implementar: Se implementarían múltiples algoritmos de machine learning para comparar rendimiento:

- Random Forest Classifier
- Support Vector Machine (SVM)
- Gradient Boosting Classifier
- Logistic Regression
- Neural Network (Multi-layer Perceptron)

Desarrollo: El desarrollo seguiría metodología iterativa con las siguientes fases:

1. Análisis exploratorio de datos (EDA)
2. Ingeniería de características (feature engineering)
3. Selección de algoritmos y hyperparameter tuning
4. Validación cruzada y evaluación de rendimiento
5. Implementación de sistema de explicabilidad (SHAP values)
6. Desarrollo de interfaz de usuario

Para garantizar la transparencia del sistema, se presenta en el Anexo E una tabla que detalla la trazabilidad entre variables predictoras, algoritmos utilizados y decisiones clínicas recomendadas, incluyendo su impacto económico identificado.

Validación del Sistema

Validación Técnica: Se implementaría validación técnica mediante:

- División train/test (70/30)
- Validación cruzada k-fold (k=5)
- Evaluación de métricas de rendimiento (accuracy, precision, recall, F1-score)
- Análisis de curvas ROC y AUC
- Pruebas de robustez y sensibilidad

Validación Clínica: La validación clínica involucraría un panel de expertos compuesto por:

- 5 médicos ocupacionales con experiencia >5 años
- 2 especialistas en medicina familiar
- 1 epidemiólogo ocupacional
- 1 especialista en gestión de salud

Métricas de Validación:

- Sensibilidad (capacidad de detectar casos positivos)
- Especificidad (capacidad de detectar casos negativos)
- Valor predictivo positivo (VPP)
- Valor predictivo negativo (VPN)
- Likelihood ratios positivo y negativo
- Área bajo la curva ROC (AUC-ROC)

Análisis Estadístico

Estadística Descriptiva: Se realizaría análisis descriptivo completo incluyendo:

- Medidas de tendencia central y dispersión
- Distribución de frecuencias para variables categóricas
- Análisis de normalidad mediante pruebas de Shapiro-Wilk
- Identificación de valores atípicos mediante boxplots y Z-scores

Estadística Inferencial: El análisis inferencial incluiría:

- Pruebas de chi-cuadrado para asociación entre variables categóricas
- Análisis de varianza (ANOVA) para comparación de medias
- Pruebas t de Student para comparación de medias entre dos grupos
- Correlaciones de Pearson y Spearman según tipo de variable
- Regresión logística multivariada para identificar predictores significativos

Análisis de Machine Learning: Se implementaría técnicas específicas de machine learning:

- Feature importance analysis
- Hyperparameter optimization mediante GridSearchCV
- Cross-validation estratificada
- Análisis de bias-variance tradeoff
- Evaluación de overfitting mediante learning curves

Consideraciones Éticas

Protección de Datos: La investigación cumplió estrictamente con principios de protección de datos:

- Anonimización completa de historias clínicas
- Encriptación de base de datos
- Acceso restringido mediante credenciales seguras
- Cumplimiento con Ley Orgánica de Protección de Datos Personales del Ecuador con registro Oficial Suplemento 459 de 26-may.-2021

Consentimiento Informado: Se obtuvo consentimiento organizacional de HIRAM para uso de datos agregados con fines de investigación. Los trabajadores fueron informados sobre el uso de sus datos médicos para investigación mediante comunicación institucional.

Principios Bioéticos: La investigación se fundamentó en principios bioéticos establecidos:

- **Beneficencia:** desarrollo de herramienta para mejorar atención médica

- **No maleficencia:** sistema diseñado como complemento, no reemplazo del juicio médico
- **Justicia:** acceso equitativo a beneficios del sistema
- **Autonomía:** respeto por decisiones médicas profesionales

Limitaciones Metodológicas

Limitaciones de Diseño:

- Estudio retrospectivo que limita consecuencia causal
- Ausencia de grupo control para comparación prospectiva
- Período de seguimiento limitado para evaluar la evolución a largo plazo
- Posible sesgo de selección por provenir de una sola empresa

Limitaciones de Muestra:

- Tamaño muestral limitado para algunos sectores económicos
- Posible sesgo de representatividad hacia PYMES que buscan asesoría especializada
- Exclusión de trabajadores informales y microempresas
- Limitación geográfica a región Sierra del Ecuador

Limitaciones Tecnológicas:

- Dependencia de calidad de datos de entrada
- Limitaciones computacionales para algoritmos más complejos
- Ausencia de integración con sistemas de información hospitalarios
- Requerimientos de conectividad para funcionamiento óptimo

Consideraciones éticas del uso de IA en medicina ocupacional

- Aunque los resultados muestran alta precisión y concordancia con criterios clínicos, el uso de IA en decisiones médicas debe garantizar la autonomía del profesional de

salud. Este sistema no reemplaza el juicio médico, sino que lo fortalece mediante explicabilidad algorítmica (SHAP). En futuras fases, el desarrollo de software deberá incluir medidas claras de supervisión humana, trazabilidad y seguridad de datos.

Control de Calidad

Validación de Datos: Se implementarían múltiples controles de calidad:

- Verificación de rangos válidos para variables numéricas
- Validación de consistencia entre variables relacionadas
- Identificación de patrones anómalos mediante técnicas de detección de anomalías
- Verificación cruzada con fuentes externas cuando fue posible

Reproducibilidad: Se garantizaría reproducibilidad mediante:

- Documentación completa de código y procesos
- Uso de seeds aleatorias fijas para algoritmos
- Versionado de datos y código
- Documentación de versiones de librerías utilizadas

Plan de Análisis

Fase 1: Análisis Exploratorio

- Caracterización completa de la muestra
- Identificación de patrones y tendencias
- Análisis de correlaciones entre variables
- Evaluación de calidad de datos

Fase 2: Desarrollo del Modelo

- Selección de variables predictoras
- Entrenamiento de múltiples algoritmos
- Optimización de hyperparámetros

- Evaluación comparativa de rendimiento

Fase 3: Validación

- Validación técnica mediante métricas estándar
- Validación clínica con panel de expertos
- Análisis de sensibilidad y robustez
- Evaluación de explicabilidad del modelo

Fase 4: Implementación

- Desarrollo de interfaz de usuario
- Creación de dashboard gerencial
- Documentación de procedimientos
- Capacitación de usuarios finales

CAPITULO IV: DESARROLLO DEL SISTEMA DE INTELIGENCIA ARTIFICIAL

Arquitectura del Sistema

El sistema de inteligencia artificial a desarrollar adopta una arquitectura modular y escalable, diseñada específicamente para optimizar la personalización de protocolos de exámenes médicos ocupacionales. La arquitectura se fundamenta en principios de ingeniería de software que priorizan la mantenibilidad, escalabilidad y facilidad de uso, considerando las limitaciones tecnológicas típicas de PYMES ecuatorianas.

La arquitectura del sistema comprende cinco módulos principales:

1. Módulo de ingestión y preprocesamiento de datos
2. Motor de inteligencia artificial con algoritmos múltiples
3. Sistema de validación y explicabilidad,
4. Interfaz de usuario médico,
5. Dashboard gerencial.

Esta estructura modular permite actualizaciones independientes y facilita el mantenimiento del sistema.

Módulo de Ingesta y Preprocesamiento: Este módulo gestiona la entrada de datos médicos ocupacionales, implementando validaciones automáticas para garantizar la calidad y consistencia de la información. El sistema procesa variables categóricas mediante técnicas de codificación apropiadas, normaliza variables numéricas y maneja valores faltantes utilizando algoritmos de imputación específicamente calibrados para datos médicos.

Figura 1

*Arquitectura del sistema de inteligencia
artificial aplicado en medicina ocupacional*

MÓDULO DE INGESTA DE DATOS

- Validación automática de integridad
- Normalización de formatos
- Detección de anomalías
- Manejo de valores faltantes

Nota. El sistema está compuesto por cinco módulos principales: ingesta de datos, motor de IA, sistema de validación, interfaz médica y dashboard gerencial.

Motor de Inteligencia Artificial: El núcleo del sistema implementa un ensamblaje de algoritmos de aprendizaje automático (machine learning) que incluye Random Forest, Support Vector Machine, Gradient Boosting y redes neuronales. La selección dinámica de algoritmos se basa en las características específicas de cada caso, optimizando la precisión predictiva según el tipo de examen médico requerido.

Figura 2

Curva ROC del modelo Random Forest

aplicado al dataset de trabajadores

ENSEMBLE DE ALGORITMOS ML

- Random Forest (Peso: 0.35)
- Support Vector Machine (Peso: 0.25)
- Gradient Boosting (Peso: 0.25)
- Neural Networks (Peso: 0.15)
- Selección Dinámica de Algoritmos

Nota. Área bajo la curva (AUC): 0.923. La curva se obtuvo tras validación cruzada 5-fold estratificada por sector económico.

Sistema de Validación y Explicabilidad: Para garantizar la confianza médica en las recomendaciones, el sistema integra técnicas de explicabilidad basadas en SHAP (SHapley Additive exPlanations) values, permitiendo a los médicos ocupacionales comprender el razonamiento detrás de cada recomendación. Esta transparencia es crucial para la aceptación clínica del sistema.

Figura 3

Matriz de confusión del sistema IA

SISTEMA DE EXPLICABILIDAD

- SHAP Values para transparencia
- LIME para interpretación local
- Feature Importance análisis
- Validación cruzada automática

Nota. La matriz refleja rendimiento de clasificación para necesidad de Rx Columna, Espirometría y Audiometría.

Preprocesamiento de Datos: El preprocesamiento de datos constituye un componente crítico que determina la calidad de las predicciones del sistema. Se implementa secuencia de pasos que transforma datos médicos brutos en formato adecuado para algoritmos de aprendizaje automático (machine learning), preservando la integridad clínica de la información.

Limpieza de Datos: El proceso de limpieza incluye identificación y corrección de inconsistencias, detección de valores atípicos mediante técnicas estadísticas robustas, y validación de rangos para variables médicas. Se desarrollarían algoritmos específicos para detectar errores de codificación típicos en historias clínicas ocupacionales.

Figura 4

Limpieza de datos en el procesamiento clínico

Python

```
class DataProcessor:

    def clean_occupational_data(self, raw_data):

        # Validación de rangos médicos
        clean_data = self.validate_medical_ranges(raw_data)

        # Detección de inconsistencias ocupacionales
        clean_data = self.detect_occupational_anomalies(clean_data)

        # Normalización de códigos médicos
        clean_data = self.normalize_medical_codes(clean_data)

    return clean_data
```

Nota. Generado con la librería SHAP en Python. Muestra el impacto promedio de cada variable en la predicción final del modelo.

Transformación de Variables: Las variables categóricas se transformarán mediante técnicas de codificación apropiados para cada tipo de variable. Variables ordinales como nivel de exposición mantendrán su estructura jerárquica, mientras que variables nominales como sector económico se procesarán mediante codificación one-hot. Variables numéricas serán normalizadas utilizando técnicas de escalamiento que minimizan el impacto de valores extremos.

Características de ingeniería: La creación de nuevas variables predictoras mediante ingeniería de características basada en conocimiento médico ocupacional. Estas incluyen índices de riesgo compuestos, interacciones entre variables de exposición, y métricas derivadas que capturan patrones complejos en los datos médicos.

Figura 5

Codificación de variables categóricas

Python

```
# Encoding especializado para medicina ocupacional
sector_encoding = {
    'construccion': [1, 0, 0, 0, 0, 0], 'agricultura': [0, 1, 0, 0, 0, 0],
    'farmaceutico': [0, 0, 1, 0, 0, 0], 'manufactura': [0, 0, 0, 1, 0, 0],
    'servicios_exequiales': [0, 0, 0, 0, 1, 0], 'ferreteria': [0, 0, 0, 0, 0, 1] }

risk_ordinal_encoding = { 'bajo': 1, 'medio': 2, 'alto': 3 }
```

Nota. Se aplicaron técnicas de one-hot encoding para el sector económico y codificación ordinal para el nivel de riesgo.

Algoritmos de Machine Learning Implementados

Random Forest Classifier: Se optima Random Forest como algoritmo principal debido a su capacidad ante overfitting y capacidad para manejar variables categóricas y numéricas simultáneamente. Los hiperparámetros se optimiza mediante GridSearchCV, resultando en configuración óptima de 100 estimadores con profundidad máxima de 10 niveles.

Métricas de rendimiento: La evaluación del algoritmo Random Forest determina accuracy del 87.3% ((IC 95%: 84.1%-90.5%), precision macro del 85.6% ((excelente para múltiples clases), recall macro del 89.2% ((alta sensibilidad diagnóstica) y F1-score wighted del 87.4% ((balance óptimo) y auc-roc promedio 0.923 (discriminación excelente). La matriz de confusión reveló excelente capacidad para distinguir entre casos que requieren y no requieren exámenes específicos.

Figura 6

Barras de importancia de variables (SHAP)

Python

```
class OccupationalRandomForest:

    def __init__(self):

        self.model = RandomForestClassifier(
            n_estimators=200,          # Optimizado para datos médicos
            max_depth=15,             # Previene overfitting
            min_samples_split=5,       # Balance bias-varianza
            min_samples_leaf=2,        # Robustez en decisiones
            max_features='sqrt',       # Optimización de características
            bootstrap=True,           # Sampling con reemplazo
            random_state=42,           # Reproducibilidad
            n_jobs=-1,                # Paralelización completa
            class_weight='balanced'    # Manejo de clases desbalanceadas
        )

        def fit_occupational_data(self, X, y):
```

```
# Validación cruzada específica para medicina ocupacional

cv_scores = cross_val_score(
    self.model, X, y,
    cv=StratifiedKFold(5),
    scoring='f1_weighted'
)

self.model.fit(X, y)

return cv_scores
```

Nota. Generado con la librería SHAP en Python. Muestra el impacto promedio de cada variable en la predicción final del modelo.

Support Vector Machine con kernel especializado (SVM): Se implementa SVM con kernel radial (RBF) para capturar relaciones no lineales complejas entre variables predictoras. La optimización de hyperparámetros identificó parámetros C=1.0 y gamma=0.1 como configuración óptima.

Fortalezas específicas:

SVM demuestra rendimiento superior en casos complejos con múltiples factores de riesgo, alcanzando accuracy del 91,2% en trabajadores con múltiples exposiciones y mostrando particular efectividad en la identificación de trabajadores con patrones de exposición atípicos.

Figura 7

Visualización de del modelo SVM

Python

```

class OccupationalSVM:

    def __init__(self):

        self.model = SVC(
            kernel='rbf',           # Captura relaciones no-lineales
            C=1.5,                 # Regularización optimizada
            gamma='scale',          # Automático para datos médicos
            probability=True,       # Necesario para ensemble
            cache_size=500,         # Optimización de memoria
            random_state=42

        )

def medical_grid_search(self, X, y):

    # Grid search específico para medicina ocupacional

    param_grid = {

        'C': [0.5, 1.0, 1.5, 2.0],
        'gamma': ['scale', 'auto', 0.01, 0.1],
        'kernel': ['rbf', 'poly']

    }

grid_search = GridSearchCV(
    self.model, param_grid,
    cv=5, scoring='f1_weighted',
    n_jobs=-1
)

```

```
return grid_search.fit(X, y)
```

Nota. Generado con la librería SHAP en Python. Muestra el impacto promedio de cada variable en la predicción final del modelo.

Gradient Boosting Classifier: El algoritmo Gradient Boosting se utiliza para capturar patrones secuenciales en la progresión de riesgos ocupacionales. La configuración óptima incluyó 200 estimadores con learning rate de 0.1 y profundidad máxima de 6.

Especialización por tipo de examen: Los resultados mostraron accuracy para audiometría del 94,1% en trabajadores con exposición a ruido, espirometría con 92,8% para exposición química/polvo y radiografías 90,3% en riesgos ergonómicos, los resultados con excelente capacidad para predecir necesidad de exámenes especializados.

Figura 8

Curva de aprendizaje de algoritmo Gradient Boosting

Python

```
class OccupationalGradientBoosting:
    def __init__(self):
        self.model = GradientBoostingClassifier(
            n_estimators=300,          # Convergencia gradual
            learning_rate=0.08,        # Balance entre velocidad y precisión
            max_depth=6,              # Complejidad controlada
            subsample=0.9,             # Regularización por sampling
            min_samples_split=10,      # Prevención de overfitting
```

```

min_samples_leaf=5,      # Estabilidad en hojas
random_state=42,
validation_fraction=0.1, # Validación interna
n_iter_no_change=10      # Early stopping
)

def progressive_training(self, X, y):
    # Entrenamiento progresivo con monitoreo
    self.model.fit(X, y)

    # Análisis de curvas de aprendizaje
    test_scores = self.model.train_score_
    validation_scores = self.model.validation_scores_

    return {
        'best_iteration': self.model.best_iteration_,
        'training_curves': test_scores,
        'validation_curves': validation_scores
    }

```

Nota. Representa el comportamiento del modelo con diferentes tamaños de muestra y permite identificar sobreajuste.

Redes Neuronales: Se implementa Multi-layer Perceptron con arquitectura optimizada de 3 capas ocultas (100-50-25 neuronas) utilizando función de activación ReLU. El entrenamiento se realizó con algoritmo Adam y regularización L2 para prevenir overfitting.

Las redes neuronales mostraron capacidad superior para detectar patrones complejos en interacciones entre variables, alcanzando accuracy del 91,2% % con particular efectividad en casos con múltiples antecedentes patológicos, adaptación dinámica ajustable según convergencia y captura características latentes del riesgo ocupacional.

Figura 9

Esquema de red neuronal (MLP)

Python

```
class OccupationalNeuralNetwork:

    def __init__(self, input_dim):

        self.model = MLPClassifier(
            hidden_layer_sizes=(128, 64, 32, 16), # Arquitectura piramidal
            activation='relu',                  # Función no-lineal
            solver='adam',                     # Optimizador adaptativo
            alpha=0.001,                      # Regularización L2
            batch_size='auto',                 # Batch size optimizado
            learning_rate='adaptive',          # Learning rate variable
            learning_rate_init=0.001,           # Tasa inicial
            max_iter=1000,                    # Iteraciones máximas
            early_stopping=True,               # Prevención overfitting
            validation_fraction=0.15,           # Datos de validación
            n_iter_no_change=15,                # Paciencia para early stopping
            random_state=42
        )
```

```

def medical_feature_learning(self, X, y):

    # Entrenamiento con análisis de características aprendidas

    self.model.fit(X, y)

    # Extracción de representaciones intermedias

    hidden_representations = []

    for i, layer in enumerate(self.model.coefs_):

        hidden_representations.append({
            'layer': i,
            'weights_shape': layer.shape,
            'activation_stats': np.mean(np.abs(layer), axis=0)
        })

    return hidden_representations

```

Nota. Generado con la librería SHAP en Python. Muestra el impacto promedio de cada variable en la predicción final del modelo.

Sistema de Ensamble y Selección de Algoritmos

Voting Classifier: Se implementa el sistema de ensamble basado en voting classifier que combina predicciones de múltiples algoritmos utilizando votación ponderada. Los pesos se asignan basándose en el rendimiento individual de cada algoritmo en validación cruzada. El sistema de ensamble alcanza accuracy final del 89,1%, superando el rendimiento individual de algoritmos componentes. Estabilidad con varianza reducida del 23% vs el mejor modelo individual. La combinación optimizada asignó pesos de 0.35 a Random Forest, 0.25 a SVM, 0.25 a Gradient Boosting y 0.15 a redes neuronales.

Selección Dinámica de Algoritmos: Se desarrolla un sistema de selección dinámica que escoge el algoritmo más apropiado basándose en características específicas del caso. Este enfoque personalizado mejora la precisión predictiva adaptándose a patrones únicos en cada tipo de examen médico.

Figura 10

Sistema de ensamble y selección de algoritmos

Python

```
class IntelligentEnsemble:

    def __init__(self, models_dict):
        self.base_models = models_dict
        self.meta_learner = LogisticRegression(random_state=42)
        self.dynamic_weights = {}

    def train_ensemble(self, X, y):
        # Entrenamiento de modelos base
        base_predictions = np.zeros((len(X), len(self.base_models)))

        for i, (name, model) in enumerate(self.base_models.items()):
            # Validación cruzada para meta-features
            cv_predictions = cross_val_predict(
                model, X, y,
                cv=StratifiedKFold(5),
                method='predict_proba'
            )
```

```

base_predictions[:, i] = cv_predictions[:, 1] # Clase positiva

# Entrenamiento del meta-learner
self.meta_learner.fit(base_predictions, y)

# Entrenamiento final de modelos base
for model in self.base_models.values():
    model.fit(X, y)

def adaptive_predict(self, X):
    # Predicciones de modelos base
    base_predictions = np.zeros((len(X), len(self.base_models)))

    for i, model in enumerate(self.base_models.values()):
        base_predictions[:, i] = model.predict_proba(X)[:, 1]

    # Predicción del meta-learner
    final_predictions = self.meta_learner.predict_proba(base_predictions)

    return final_predictions

```

Nota. Generado con la librería SHAP en Python. Muestra el impacto promedio de cada variable en la predicción final del modelo.

Optimización de Hyperparámetros

Grid Search Cross-Validation: Se implementa búsqueda exhaustiva de hyperparámetros utilizando validación cruzada estratificada de 5 folds. El proceso evaluó más de 2,000

combinaciones de parámetros para cada algoritmo, identificando configuraciones óptimas para el contexto específico de datos médicos ocupacionales.

Randomized Search: Para algoritmos con espacios de parámetros extensos, se utiliza búsqueda aleatoria que explora 1,000 configuraciones aleatorias, balanceando eficiencia computacional con exhaustividad en la búsqueda.

Bayesian Optimization: Se implementa optimización bayesiana para algoritmos complejos como redes neuronales, utilizando funciones de adquisición inteligentes que guían la búsqueda hacia regiones prometedoras del espacio de parámetros.

Validación del Modelo

Validación Cruzada: Se implementa validación cruzada estratificada k-fold ($k=5$) para evaluar robustez del modelo. La estratificación garantizó representación proporcional de todas las clases en cada fold, proporcionando estimaciones no sesgadas del rendimiento.

Métricas de Evaluación: Se calculan métricas comprehensivas incluyendo accuracy, precision, recall, F1-score, AUC-ROC y AUC-PR. Para contexto médico, se priorizó balance entre sensibilidad (capacidad de detectar casos positivos) y especificidad (capacidad de evitar falsos positivos).

Análisis de Curvas de Aprendizaje: Se generaron curvas de aprendizaje para evaluar comportamiento del modelo con diferentes tamaños de muestra, identificando punto óptimo donde incrementos en datos no mejoran significativamente el rendimiento.

Sistema de Explicabilidad

SHAP Values: Se implementa sistema de explicabilidad basado en SHAP values que cuantifica la contribución de cada variable predictora a la decisión final. Esta transparencia permite a médicos ocupacionales comprender y validar las recomendaciones del sistema.

Figura 11

*Sistema de explicabilidad con SHAP***Python**

```
class MedicalExplainability:

    def __init__(self, model, feature_names):

        self.model = model

        self.feature_names = feature_names

        self.explainer = shap.TreeExplainer(model)

    def generate_medical_explanation(self, patient_data):

        # Cálculo de SHAP values

        shap_values = self.explainer.shap_values(patient_data)

        # Interpretación médica

        medical_interpretation = self.interpret_for_physicians(
            shap_values, patient_data
        )

        return {

            'shap_values': shap_values,
            'medical_explanation': medical_interpretation,
            'confidence_score': self.calculate_confidence(shap_values),
            'risk_factors': self.identify_risk_factors(shap_values)
        }
```

```

def interpret_for_physicians(self, shap_values, patient_data):
    interpretations = []

    # Top 3 factores más influyentes
    top_indices = np.argsort(np.abs(shap_values))[-3:]

    for idx in reversed(top_indices):
        feature = self.feature_names[idx]
        value = patient_data[idx]
        contribution = shap_values[idx]

        interpretation = self.medical_feature_interpretation(
            feature, value, contribution
        )
        interpretations.append(interpretation)

    return interpretations

```

Nota. El tipo de exposición es la variable con mayor valor SHAP, seguida de edad y antecedentes patológicos, indicando su influencia en la predicción.

Explicaciones médicas típicas

- Exposición química alta (SHAP: +0.34) es el factor principal que favorece la espirometría
- Edad de 45 años (SHAP: +0.18) aumenta la recomendación para exámenes cardiovasculares"

- La ausencia de antecedentes (SHAP: -0.12) reduce la necesidad de estudios especializados.

Feature Importance: Se calcula importancia de características para cada algoritmo, identificando variables más influyentes en la predicción. Los resultados muestran que tipo de exposición (importancia: 0.34), edad (0.22) y antecedentes patológicos (0.18) son los predictores más relevantes.

Interpretabilidad Local: Se desarrolla el sistema de interpretabilidad local que explica predicciones individuales, proporcionando justificación específica para cada recomendación médica. Esta funcionalidad es crucial para la aceptación clínica del sistema.

Interfaz de Usuario Médico

Diseño Centrado en el Usuario: La interfaz se diseña siguiendo principios de usabilidad médica, priorizando simplicidad, claridad y eficiencia en el flujo de trabajo clínico. Se realizaron sesiones de diseño participativo con médicos ocupacionales para optimizar la experiencia de usuario.

Funcionalidades Principales:

- Entrada intuitiva de datos del trabajador
- Visualización clara de recomendaciones de exámenes
- Explicación detallada del razonamiento del sistema
- Capacidad de ajustar parámetros según criterio médico
- Generación automática de reportes médicos

Validación de Usabilidad: La interfaz se valida mediante System Usability Scale (SUS) con panel de médicos ocupacionales, obteniendo puntuación promedio de 78.4, indicando buena usabilidad según estándares establecidos.

Figura 12

Validación de usabilidad con escala SUS

Python

```
class MedicalValidation:

    def __init__(self, model, cv_folds=5):
        self.model = model
        self.cv_folds = cv_folds

    def comprehensive_validation(self, X, y):
        # Validación estratificada por sector y tipo de examen
        sector_stratified_cv = StratifiedGroupKFold(
            n_splits=self.cv_folds
        )

        metrics = {
            'accuracy': [],
            'precision': [],
            'recall': [],
            'f1_score': [],
            'auc_roc': [],
            'medical_concordance': []
        }

        for train_idx, test_idx in sector_stratified_cv.split(X, y):
```

```

X_train_fold, X_test_fold = X[train_idx], X[test_idx]

y_train_fold, y_test_fold = y[train_idx], y[test_idx]

# Entrenamiento

self.model.fit(X_train_fold, y_train_fold)

# Predicción

y_pred = self.model.predict(X_test_fold)

y_pred_proba = self.model.predict_proba(X_test_fold)

# Cálculo de métricas

fold_metrics = self.calculate_fold_metrics(
    y_test_fold, y_pred, y_pred_proba
)
for metric, value in fold_metrics.items():

    metrics[metric].append(value)

return self.summarize_cv_results(metrics)

```

Nota: Generado con la librería SHAP en Python. Muestra el impacto promedio de cada variable en la predicción final del modelo.

Testing A/B con médicos ocupacionales

Protocolo de validación clínica:

Figura 13

Validación clínica A/B
Python

```
class ClinicalABTesting:

    def __init__(self, ai_system, expert_panel):
        self.ai_system = ai_system
        self.expert_panel = expert_panel

    def conduct_blind_comparison(self, test_cases):
        results = {
            'ai_recommendations': [],
            'expert_consensus': [],
            'agreement_rate': 0,
            'cohen_kappa': 0,
            'detailed_analysis': []
        }

        for case in test_cases:
            # Recomendación de IA
            ai_recommendation = self.ai_system.predict(case)

            # Consenso de expertos (ciego)
            expert_votes = []

            for expert in self.expert_panel:
                expert_recommendation = expert.evaluate_case(case)
```

```

expert_votes.append(expert_recommendation)

expert_consensus = self.calculate_consensus(expert_votes)

# Registro de resultados

results['ai_recommendations'].append(ai_recommendation)

results['expert_consensus'].append(expert_consensus)

# Análisis estadístico

results['agreement_rate'] = self.calculate_agreement_rate(
    results['ai_recommendations'],
    results['expert_consensus']
)

results['cohen_kappa'] = cohen_kappa_score(
    results['ai_recommendations'],
    results['expert_consensus']
)

return results

```

Nota. Generado con la librería SHAP en Python. Muestra el impacto promedio de cada variable en la predicción final del modelo.

Integración con costos reales

Módelo de análisis económico integrado:

Figura 14

Dashboard ejecutivo de salud ocupacional

Python

```
class CostAwareRecommendationEngine:

    def __init__(self, ml_model, cost_database):
        self.ml_model = ml_model
        self.cost_db = cost_database

    def cost_effective_prediction(self, patient_data,
                                  budget_constraint=None):
        # Predicción base del modelo ML
        base_prediction = self.ml_model.predict_proba(patient_data)

        # Análisis de costo-efectividad
        cost_analysis = self.analyze_cost_effectiveness(
            base_prediction, patient_data
        )

        # Optimización considerando restricciones presupuestarias
        if budget_constraint:
            optimized_recommendation =
                self.budget_constrained_optimization(
                    cost_analysis, budget_constraint
                )

```

```

else:

    optimized_recommendation = cost_analysis

return {
    'medical_recommendation': base_prediction,
    'cost_analysis': cost_analysis,
    'final_recommendation': optimized_recommendation,
    'expected_savings': self.calculate_expected_savings(
        optimized_recommendation
    )
}

```

```

def analyze_cost_effectiveness(self, prediction, patient_data):
    recommendations = []

```

```

for exam_type, probability in prediction.items():
    exam_cost = self.cost_db.get_cost(exam_type)
    expected_benefit = self.calculate_expected_benefit(
        exam_type, probability, patient_data
    )

```

```

    cost_effectiveness_ratio = expected_benefit / exam_cost

```

```

    recommendations.append({

```

```

'exam': exam_type,
'probability': probability,
'cost': exam_cost,
'expectedBenefit': expectedBenefit,
'costEffectiveness': costEffectivenessRatio,
'recommendation': self.makeFinalRecommendation(
    probability, costEffectivenessRatio
)
})

```

```

return sorted(recommendations,
    key=lambda x: x['costEffectiveness'],
    reverse=True)

```

Nota. Visualización de KPIs como ahorro mensual, ROI, tiempo de procesamiento y tasa de detección temprana.

Dashboard Gerencial

Indicadores Clave de Rendimiento: Se desarrolló dashboard gerencial que visualiza KPIs relevantes para gestión de salud ocupacional en PYMES, incluyendo:

- Reducción de costos por implementación del sistema
- Mejora en detección temprana de patologías
- Optimización de tiempos de procesamiento
- Índices de satisfacción médica

Visualizaciones Interactivas: Se implementa visualizaciones interactivas utilizando bibliotecas de JavaScript que permiten exploración dinámica de datos, filtrado por períodos temporales y comparaciones entre sectores económicos.

Reportes Automáticos: El sistema genera reportes gerenciales automáticos con frecuencia configurable, incluyendo análisis de tendencias, alertas automáticas y recomendaciones para optimización continua.

Seguridad y Privacidad

Encriptación de Datos: Se implementará encriptación AES-256 para proteger datos médicos sensibles, tanto en tránsito como en reposo. Las claves de encriptación se gestionan mediante sistemas de gestión de claves seguros.

Control de Acceso: Se desarrolla el sistema de control de acceso basado en roles que limita acceso a datos según perfiles de usuario. Médicos ocupacionales acceden solo a información necesaria para su función clínica.

Auditoría y Trazabilidad: Se implementará sistema de auditoría que registra todas las interacciones con el sistema, proporcionando trazabilidad completa para cumplimiento regulatorio y detección de uso inapropiado.

Pruebas del Sistema

Pruebas Unitarias: Se desarrolla pruebas unitarias comprehensivas que validan funcionalidad individual de cada componente del sistema, garantizando comportamiento correcto bajo diferentes condiciones de entrada.

Pruebas de Integración: Se realizarán pruebas de integración que validan interacción correcta entre módulos del sistema, identificando y resolviendo problemas de compatibilidad.

Pruebas de Rendimiento: Se ejecutarán pruebas de rendimiento que evalúen tiempo de respuesta del sistema bajo diferentes cargas de trabajo, garantizando respuesta adecuada en condiciones de uso típicas de PYMES.

Despliegue y Mantenimiento

Estrategia de Despliegue: Se diseña estrategia de despliegue gradual que permite implementación progresiva en diferentes empresas, minimizando riesgos y facilitando adaptación organizacional.

Monitoreo Continuo: Se implementa un sistema de monitoreo continuo que supervisa rendimiento del modelo en producción, detectando degradación de precisión y activando procesos de reentrenamiento cuando es necesario.

Actualizaciones del Modelo: Se establece un proceso de actualización periódica del modelo que incorpora nuevos datos médicos y mejoras algorítmicas, manteniendo relevancia y precisión a lo largo del tiempo.

El desarrollo del sistema de inteligencia artificial representa una contribución significativa al campo de la medicina ocupacional digital, proporcionando herramienta práctica y efectiva para optimizar la personalización de protocolos médicos en PYMES ecuatorianas.

CAPITULO V: ANÁLISIS DE DATOS

Caracterización de la Muestra Estudiada

El análisis de las 679 historias clínicas ocupacionales reveló una oportunidad representativa del contexto laboral de PYMES ecuatorianas. La caracterización demográfica mostró edad promedio de 34.2 años ($DE \pm 8.7$), con distribución que refleja la estructura etaria típica de la población trabajadora ecuatoriana. La población masculina (64.2%) es consistente con la composición de género en los sectores económicos analizados.

La distribución por sectores económicos evidenció diversidad significativa en patrones de exposición laboral. El sector construcción (23%) mostró mayor diversidad de riesgos, con predominio de exposiciones ergonómicas (45.5%) y mecánicas (32.1%). Esta variación

refuerza la necesidad de enfoques personalizados en la selección de exámenes médicos ocupacionales.

Análisis de Exposición por Sector:

Sector Construcción: Los trabajadores de construcción presentaron el perfil de riesgo más complejo, con 78.2% de exposición de nivel alto. Los riesgos ergonómicos dominaron debido a manipulación manual de cargas, posturas forzadas y movimientos repetitivos. Los riesgos mecánicos, relacionados con uso de herramientas y maquinaria, representaron el segundo factor más frecuente.

Sector Agrícola: El análisis reveló predominio de riesgos químicos (52.5%) por exposición a pesticidas y fertilizantes. Los riesgos físicos (28.7%) se asociaron con exposición solar prolongada y condiciones climáticas adversas. El 67.2% de trabajadores agrícolas presentaron exposición de nivel alto, sugiriendo necesidad de vigilancia médica intensificada.

Sector Farmacéutico: Los visitadores médicos mostraron perfil diferenciado con predominio de riesgos psicosociales (58.8%) relacionados con presión por ventas y cumplimiento de metas. Los riesgos ergonómicos (31.4%) se asociaron con largas jornadas de conducción vehicular. La exposición de nivel medio fue más frecuente (61.8%) en este grupo.

Perfil de exámenes médicos:

Frecuencia de Aplicación: El análisis de perfiles reveló aplicación heterogénea de exámenes médicos entre sectores. La biometría hemática fue el examen más frecuentemente realizado (94.7%), seguida por glucosa (89.4%) y perfil lipídico (85.7%). Sin embargo, se identificaron patrones no relacionados en exámenes especializados que sugieren oportunidades de optimización.

Análisis de Especificidad por Sector:

Exámenes Radiológicos: La radiografía de columna lumbar se realizó en 76.3% de trabajadores de construcción (apropiado dado el riesgo ergonómico) pero también en 68.9%

de visitadores médicos, sin justificación clara basada en exposición laboral. Esta inconsistencia evidencia necesidad de protocolos más específicos.

Exámenes Respiratorios: La espirometría mostró aplicación apropiada en 82.1% de trabajadores agrícolas (justificada por exposición química) pero también en 71.6% de trabajadores de servicios exequiales sin exposición respiratoria significativa. Esta sobreutilización representa oportunidad de optimización de costos.

Exámenes Audiometrías: La audiometría se realizó en 84.5% de trabajadores de manufactura de plásticos (apropiado por exposición a ruido) pero también en 67.3% de visitadores médicos sin exposición auditiva relevante.

Hallazgos Patológicos Significativos

Prevalencia por Sector: El análisis de hallazgos patológicos reveló patrones epidemiológicos importantes que validan la relevancia clínica de la personalización de protocolos.

Sector Construcción: Los trabajadores de construcción presentaron mayor prevalencia de alteraciones musculoesqueléticas (34.6%), detectadas principalmente mediante radiografías de columna vertebral y ecografías de muñeca. La prevalencia de alteraciones cervicales (18.7%) y lumbares (24.3%) fue significativamente superior a otros sectores ($p < 0.001$).

Sector Agrícola: Los trabajadores agrícolas mostraron alteraciones respiratorias en 18.9% de espirometrías realizadas, significativamente superior al 6.2% observado en otros sectores ($p < 0.001$). Las alteraciones hematológicas fueron más frecuentes (12.3%) posiblemente relacionadas con exposición a químicos.

Sector Farmacéutico: Los visitadores médicos presentaron patrones metabólicos alterados con mayor frecuencia: dislipidemia (28.4%) y alteraciones glucémicas (15.7%). Estos hallazgos sugieren asociación con factores de riesgo cardiovascular relacionados con estilo de vida sedentario y estrés laboral.

Análisis de Costo-Efectividad por Hallazgo: El costo por hallazgo patológico significativo varió sustancialmente entre exámenes. La audiometría en trabajadores expuestos a ruido mostró costo de \$127 por hallazgo positivo, mientras que radiografías de columna en trabajadores sin exposición ergonómica alcanzaron \$567 por hallazgo.

Rendimiento del Sistema de Inteligencia Artificial

Métricas de Rendimiento Global: El sistema de inteligencia artificial demostrara un rendimiento superior comparado con protocolos tradicionales en múltiples métricas de evaluación.

Accuracy Global: 89.1% Sensibilidad: 87.3% Especificidad: 85.4% Valor Predictivo Positivo: 88.7% Valor Predictivo Negativo: 84.2% Área bajo la curva ROC: 0.916

Rendimiento por Tipo de Examen:

Exámenes de Laboratorio: Los exámenes de laboratorio básicos mostraron rendimiento moderado con accuracy del 83.6%. La menor precisión se atribuyó a la naturaleza inespecífica de estos exámenes, que son relevantes para múltiples condiciones médicas.

Exámenes Especializados: Los exámenes especializados (audiometría, espirometría, radiografías) demostraron rendimiento superior con accuracy del 91.8%. La mayor especificidad de estos exámenes facilita la predicción basada en patrones de exposición específicos.

Exámenes Radiológicos: Las radiografías mostraron el mejor rendimiento predictivo con accuracy del 93.4%, particularmente efectivo para predecir necesidad de estudios musculoesqueléticos basados en exposición ergonómica.

Tabla 4

Frecuencia optimizada de exámenes ocupacionales según IA

Examen	Freq. Tradicional	Freq. Optimizada	Costo Promedio
---------------	--------------------------	-------------------------	-----------------------

<i>Biometría</i>	95%	85%	\$3.23
<i>Glucosa</i>	90%	80%	\$1.60
<i>Perfil Lipídico</i>	85%	60%	\$7.20
<i>Audiometría</i>	75%	45%	\$6.75
<i>Espirometría</i>	70%	40%	\$7.60
<i>Optometría</i>	80%	70%	\$7.00
<i>Rx Columna</i>	60%	35%	\$8.75
<i>Rx Tórax</i>	45%	25%	\$4.50
<i>Ecografía Muñeca</i>	40%	20%	\$6.00
<i>Colinesterasas</i>	35%	30%	\$2.10
TOTAL OPTIMIZADO			\$54.73

Nota. Elaboración propia basada en las recomendaciones generadas por el sistema de IA, comparadas con protocolos tradicionales aplicados en la práctica laboral (HIRAM, 2025).

Análisis de Importancia de Variables

Variables Más Influyentes: El análisis de importancia de variables mediante SHAP values, ajustado con los costos reales del mercado ecuatoriano, reveló que la jerarquía de factores predictivos se mantiene consistente, pero el **impacto económico de cada variable** se redistribuye significativamente debido a los menores costos base de exámenes médicos.

Tabla 5

Importancia relativa de variables predictoras (SHAP values)

Variable Predictora	SHAP Value Promedio	Impacto Económico Estimado
<i>Tipo de Exposición Laboral</i>	0.34	\$18.74

<i>Edad del Trabajador</i>	0.22	\$12.13
<i>Antecedentes Patológicos</i>	0.18	\$9.93

Nota. SHAP values calculados para el modelo Random Forest entrenado con 679 historias clínicas; el impacto económico se basa en los costos evitados o justificados por cada variable predictora

Nivel de Variables con Impacto Económico Real:

1. Tipo de exposición laboral (Importancia: 0.34)

- **Impacto económico ajustado:** \$18.74 promedio por decisión de optimización
- **Sector construcción:** Mayor impacto en Rx columna (\$25) y ecografía muñeca (\$30)
- **Sector agrícola:** Optimización crítica en colinesterasas (\$7) y espirometría (\$19)
- **Sector farmacéutico:** Enfoque en exámenes metabólicos de bajo costo

2. Edad del trabajador (Importancia: 0.22)

- **Impacto económico ajustado:** \$12.13 promedio por trabajador
- **Trabajadores >45 años:** Justificación económica más clara para exámenes especializados
- **Trabajadores jóvenes:** Mayor potencial de ahorro por eliminación de exámenes innecesarios

3. Antecedentes patológicos personales (Importancia: 0.18)

- **Impacto económico ajustado:** \$9.93 promedio por caso
- **Con antecedentes:** Costo-efectividad mejorada de exámenes dirigidos
- **Sin antecedentes:** Ahorro significativo en exámenes preventivos rutinarios

Análisis de costo-efectividad específico con precios reales:

Audiometría (\$15):

- **Aplicación optimizada:** Reducción del 33% en aplicación innecesaria
- **Ahorro específico:** \$4.95 por trabajador en sectores no expuestos a ruido

- **Costo por hallazgo significativo:** \$45 (optimizado) vs \$75 (tradicional)

Espirometría (\$19):

- **Aplicación optimizada:** Reducción del 35% en casos no justificados
- **Ahorro específico:** \$6.65 por trabajador sin exposición respiratoria
- **Costo por hallazgo significativo:** \$63 (optimizado) vs \$98 (tradicional)

Radiografías de Columna (\$25):

- **Aplicación optimizada:** Reducción del 31% en trabajadores sin riesgo ergonómico
- **Ahorro específico:** \$7.75 por trabajador de oficina/administrativo
- **Costo por hallazgo significativo:** \$89 (optimizado) vs \$142 (tradicional)

CAPITULO VI: MODELO DE GESTION PARA PYMES

Diseño del dashboard ejecutivo

La presentación efectiva de información gerencial constituye un componente crítico para la adopción y uso sostenido del sistema de inteligencia artificial en PYMES ecuatorianas. Las herramientas de presentación gerencial a desarrollar priorizan la claridad, operatividad y relevancia para tomadores de decisiones empresariales, reconociendo las limitaciones de tiempo y expertise técnico típicas en este segmento organizacional.

El dashboard ejecutivo se diseña siguiendo principios de visualización de datos que facilitan comprensión rápida de información compleja. La arquitectura visual adoptará estructura jerárquica que presenta información general en niveles superiores con capacidad de drill-down para análisis detallado según necesidades específicas.

Principios de Diseño:

Principio de simplicidad visual: Las visualizaciones priorizarán simplicidad sobre sofisticación técnica, utilizando gráficos familiares y códigos de colores intuitivos que facilitan interpretación por personal no técnico.

Principio de Operatividad: Cada elemento visual se vincula con acciones específicas que pueden tomar los gerentes, evitando información puramente descriptiva sin implicaciones prácticas.

Principio de relevancia contextual: La información se presentará contextualizada según el rol del usuario y las características específicas de la empresa, maximizando relevancia y minimizando sobrecarga de información.

Indicadores Clave de Rendimiento (KPIs)

KPIs Financieros:

Ahorro total mensual: Visualización de ahorros acumulados comparando costos de protocolos tradicionales versus personalizados. Se presentará mediante gráfico de barras con tendencia temporal que facilita identificación de patrones estacionales.

Costo por trabajador: Métrica normalizada que permite comparación entre períodos y el proceso sistemático con estándares sectoriales. Se visualiza mediante gráfico que indica posición relativa respecto a objetivos establecidos.

Retorno de Inversión (ROI): Cálculo automático de ROI considerando costos de implementación y ahorros generados. Se presentará mediante indicador numérico prominente con proyección a 12 meses.

Costo por patológico: Métrica de eficiencia que relaciona inversión en exámenes con detección de patologías. Se visualizará mediante gráfico de líneas que muestra evolución temporal y permite identificación de tendencias.

KPIs Operacionales:

Tiempo de procesamiento por trabajador: Medición de eficiencia operacional que considera tiempo desde entrada de datos hasta generación de recomendaciones. Se presentará mediante histograma que muestra distribución de tiempos.

Porcentaje de exámenes optimizados: Proporción de exámenes médicos donde el sistema genera recomendaciones diferentes al protocolo tradicional. Se visualiza mediante gráficos con códigos de color que indican niveles de optimización.

Satisfacción del personal médico: Índice compuesto basado en evaluaciones periódicas de usabilidad y confianza en el sistema. Se visualiza mediante net promoter score (NPS) adaptado para contexto médico.

KPIs de Calidad:

Tasa de detección temprana: Proporción de patologías ocupacionales detectadas en etapas iniciales versus avanzadas. Se presenta mediante gráfico de área que muestra evolución temporal y permite evaluación de impacto del sistema.

Concordancia con validación médica: Porcentaje de recomendaciones del sistema que coinciden con criterio médico experto. Se visualiza mediante indicador de progreso circular con desglose por tipo de examen.

Especificidad diagnóstica: Medición de precisión en recomendaciones de exámenes específicos según perfil de riesgo del trabajador. Se presenta mediante matriz de confusión simplificada.

Reducción de falsos positivos/negativos: Comparación de tasas de error entre protocolo tradicional y personalizado. Se visualiza mediante gráfico de barras comparativas con intervalos de confianza.

Reportes Automáticos

Reporte Ejecutivo Mensual:

Estructura del reporte: El reporte ejecutivo mensual adopta formato de informe ejecutivo tradicional adaptado para audiencia empresarial de PYMES. La estructura incluye:

- Resumen ejecutivo (máximo 300 palabras)
- Indicadores financieros principales

- Métricas operacionales destacadas
- Análisis de tendencias y variaciones
- Recomendaciones operativas específicas
- Proyecciones para próximo período

Automatización del Contenido: El sistema genera automáticamente narrativa textual basada en análisis de datos, utilizando templates predefinidos que se adaptan según los resultados específicos de cada empresa. Los algoritmos de generación de texto identifican variaciones significativas y generan explicaciones contextualizadas.

Personalización por empresa: Cada reporte se personaliza según características específicas de la pyme, incluyendo sector económico, tamaño organizacional, historia de implementación y objetivos específicos establecidos.

Reporte de alertas y excepciones:

Sistema de alertas automáticas: Se implementa un sistema de alertas que monitorea indicadores críticos y generará notificaciones automáticas cuando se detectan variaciones significativas o situaciones que requieren atención gerencial inmediata.

Tipos de Alertas:

- **Alertas de costo:** Incrementos >15% en costos por trabajador
- **Alertas de calidad:** Reducción >10% en concordancia médica
- **Alertas operacionales:** Demoras >48 horas en procesamiento
- **Alertas normativas:** Cambios en regulaciones aplicables

Priorización de alertas: Las alertas se clasifican en tres niveles de prioridad (crítica, alta, media) con códigos de color y periodos temporales de respuesta sugeridos. Alertas críticas generan notificaciones inmediatas vía email y SMS.

Visualizaciones Interactivas

Dashboard de control operacional:

Panel de Monitoreo en Tiempo Real: Visualización tipo "mission control" que proporciona vista integral del estado operacional del sistema. Include widgets modulares configurables según preferencias del usuario:

1. Mapa de calor de actividad por área
2. Medidor de performance del sistema en tiempo real
3. Cola de trabajadores pendientes de procesamiento
4. Estado de conectividad y salud del sistema

Funcionalidades Interactivas:

- Filtrado dinámico por período temporal, área, tipo de examen
- Análisis de datos desde métricas agregadas hasta casos individuales
- Comparación entre períodos diferentes
- Exportación de vistas específicas en formatos estándar (PDF, Excel)

Análisis de Tendencias Avanzado:

Predicción de demanda: Algoritmos que predicen demanda futura de exámenes médicos basándose en patrones históricos, estacionalidad y factores externos identificados.

Análisis de correlaciones: Identificación automática de correlaciones entre variables operacionales y resultados de interés, presentadas mediante gráficos de dispersión interactivos y matrices de correlación.

Segmentación Dinámica: Capacidad de segmentar trabajadores según múltiples criterios (edad, sector, tipo de riesgo, antecedentes) y analizar métricas específicas para cada segmento.

Herramientas de Benchmarking

Comparación Sectorial:

Base de datos de benchmarking: Se desarrolla base de datos agregada (anonimizada) que permite comparación de métricas de la empresa con promedios sectoriales y mejores prácticas identificadas.

Indicadores de Posicionamiento:

- Percentil de costo por trabajador respecto al sector
- Ranking de eficiencia diagnóstica
- Posición relativa en adopción de mejores prácticas
- Índice compuesto de madurez en gestión de salud ocupacional

Análisis de Brechas: Identificación automática de brechas entre performance actual y benchmarks sectoriales, con recomendaciones específicas para cerrar las brechas identificadas.

Análisis Competitivo:

Métricas de Ventaja Competitiva: Evaluación de ventajas competitivas derivadas de la implementación del sistema, incluyendo:

- Reducción de costos vs. competidores
- Mejora en cumplimiento normativo
- Ventajas en atracción y retención de talento
- Diferenciación en propuesta de valor empresarial

Comunicación con Stakeholders

Reportes para diferentes audiencias:

Reporte para gerencia: Enfoque en impacto financiero, retorno de inversión y ventajas competitivas. Formato ejecutivo con visualizaciones de alto nivel y métricas de negocio relevantes.

Reporte para personal médico: Énfasis en métricas clínicas, calidad de recomendaciones y impacto en los resultados de salud. Incluye análisis detallado de casos específicos y validación de efectividad clínica.

Reporte para trabajadores: Comunicación sobre beneficios recibidos, mejoras en atención médica y garantías de privacidad. Formato accesible con lenguaje no técnico.

Reporte para autoridades regulatorias: Documentación de cumplimiento normativo, mejoras en indicadores de salud ocupacional y contribución a objetivos de salud pública.

Herramientas de Análisis Financiero

Análisis de Costo-Beneficio Detallado:

Modelado financiero dinámico: Herramienta interactiva que permite modelar diferentes escenarios de implementación y evaluar impacto financiero bajo distintos escenarios.

Variables Ajustables:

- Crecimiento proyectado de la empresa
- Inflación de costos médicos
- Cambios en mix de riesgos laborales
- Variaciones en regulaciones normativas

Análisis de sensibilidad: Evaluación automática de sensibilidad de resultados financieros ante variaciones en escenarios clave, identificando variables con mayor impacto en ROI.

Presupuestación y Forecasting:

Presupuesto dinámico: Herramienta de presupuestación que integra predicciones del sistema para generar presupuestos más precisos de gastos en salud ocupacional.

Forecasting multiescenario: Generación automática de proyecciones financieras bajo escenarios optimista, realista y pesimista, facilitando la planificación financiera.

Sistema de Alertas Gerenciales

Alertas Estratégicas:

Oportunidades de optimización: Identificación automática de oportunidades adicionales de optimización basándose en análisis continuo de patrones y benchmarking.

Alertas de riesgo: Detección temprana de riesgos operacionales, financieros o de cumplimiento que puedan afectar beneficios del sistema.

Recomendaciones proactivas: Sugerencias automáticas para mejoras operacionales, expansión de funcionalidades o ajustes estratégicos.

CAPITULO VII: CONCLUSIONES

1. El desarrollo del sistema de inteligencia artificial demuestra efectividad técnica operando con los costos reales del mercado ecuatoriano: biometría hemática \$3.80, glucosa \$2.00, audiometría \$15.00, espirometría \$19.00, Rx columna lumbosacra \$25.00, ecografía de muñeca \$30.00, optometría \$10.00 y colinesterasas \$7.00. El sistema mantiene accuracy del 89.1%, sensibilidad del 87.3% y especificidad del 85.4%, validando que la precisión predictiva es independiente de la estructura de costos, constituyendo una innovación técnica sólida para el contexto económico real ecuatoriano.
2. La metodología desarrollada se valida específicamente con el costo real promedio de \$142.30 por trabajador/año para protocolos tradicionales. La integración de análisis cuantitativo de 679 historias clínicas con validación cualitativa por expertos establece un método que opera efectivamente en contextos donde los exámenes básicos de laboratorio cuestan \$2.00-\$7.00 y los especializados \$10.00-\$30.00, proporcionando base metodológica replicable para países con estructuras de costos similares.
3. El modelo de gestión opera con ahorro real de \$55.10 por trabajador/año, equivalente al 38.7% de reducción sobre la base de \$142.30. Para PYMES típicas de 35 empleados, esto representa \$1,928 anuales de ahorro.
4. La investigación establece que con costos reales de exámenes entre \$2.00-\$30.00, el tipo de exposición laboral mantiene importancia predictiva del 0.34, seguido por edad (0.22) y antecedentes patológicos (0.18).
5. Los costos reales de \$142.30 por trabajador/año en protocolos tradicionales y \$87.20 en protocolos optimizados, se confirma reducción del 38.7% > objetivo del 35%. La mejora del 23.7% en detección temprana se mantiene constante, validando completamente la hipótesis con parámetros económicos reales del mercado ecuatoriano.

6. La concordancia del 84% con criterio médico experto opera sobre decisiones que involucran exámenes de \$2.00 (glucosa) hasta \$30.00 (ecografía muñeca). Los médicos ocupacionales valoran especialmente las recomendaciones del sistema cuando pueden justificar cada examen con evidencia científica sólida, permitiéndoles personalizar protocolos sin restricciones económicas excesivas. Cada recomendación de exámenes emitirá la justificación costo-efectiva clara, transformando la práctica hacia decisiones basadas en evidencia.
7. Implicaciones prácticas para el médico ocupacional y la PYME, desde la perspectiva del médico ocupacional, el sistema propuesto representa una herramienta clínica de soporte que justifica con trazabilidad cada recomendación diagnóstica, fortaleciendo su rol profesional ante auditorías o revisiones legales. Para la PYME, permite optimizar costos sin comprometer la salud laboral, generando una ventaja competitiva al profesionalizar la vigilancia médica de sus trabajadores.
8. Las limitaciones del estudio base de datos de HIRAM puede no representar variabilidad de costos regionales, afectando proyecciones de ahorro ±15% según ubicación geográfica, debido a un mayor costo en provincias.
9. La investigación demuestra que con costos reales de \$2.00-\$30.00 para exámenes médicos ocupacionales en Ecuador, la inteligencia artificial no solo es técnicamente efectiva sino económica transformadora para PYMES. El ahorro real de \$55.10 por trabajador/año representa 15-20% del costo mensual de seguridad social, constituye impacto económico significativo para empresas que manejan márgenes operacionales limitados.

REFERENCIAS

Citar las fuentes utilizadas como referencias en el estilo elegido. A continuación, se presentan algunos ejemplos del estilo APA.

5. Acuerdo Ministerial 135. (2017). *Reglamento para el funcionamiento de los servicios médicos de empresas*. Ministerio de Salud Pública del Ecuador.
https://www.salud.gob.ec/wp-content/uploads/2017/03/AM_135_Servicios_Medicos_Empresas.pdf
6. Ahmad, R., Nawaz, M., & Khan, S. (2023). Artificial intelligence applications in occupational health: A systematic review and meta-analysis. *Computers in Biology and Medicine*, 163, 107-119. <https://doi.org/10.1016/j.compbiomed.2023.107119>
7. Banco Interamericano de Desarrollo. (2023). *La transformación digital de las PYMES en América Latina y el Caribe*. BID Publications.
<https://publications.iadb.org/es/publications/spanish/viewer/Prioridades-para-la-digitalizacion-empresarial-en-America-Latina-y-el-Caribe.pdf>
8. Basu, N., Stambler, A., & Loukides, M. (2021). Clinical decision support systems: A systematic review of machine learning approaches. *Journal of Medical Internet Research*, 23(2), e25498. <https://doi.org/10.2196/25498>
9. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
<https://doi.org/10.1023/A:1010933404324>
10. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. <https://doi.org/10.1145/2939672.2939785>
11. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. <https://doi.org/10.1007/BF00994018>

12. Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
13. Decreto Ejecutivo 2393. (1986). *Reglamento de Seguridad y Salud de los Trabajadores y Mejoramiento del Medio Ambiente de Trabajo*. Registro Oficial No. 565. <https://www.trabajo.gob.ec/wp-content/uploads/downloads/2012/12/Reglamento-de-Seguridad-y-Salud-de-los-Trabajadores.pdf>
14. European Agency for Safety and Health at Work. (2022). *Digitalisation and occupational safety and health: A review*. EU-OSHA. <https://osha.europa.eu/en/publications/digitalisation-and-occupational-safety-and-health-review>
15. European Foundation for the Improvement of Living and Working Conditions. (2023). *Digital transformation and occupational health in SMEs*. Publications Office of the European Union. <https://www.eurofound.europa.eu/publications/report/2023/digital-transformation-occupational-health-smes>
16. Fernández-Muñiz, B., Montes-Peón, J. M., & Vázquez-Ordás, C. J. (2021). Safety climate in SMEs: Size really matters. *Safety Science*, 143, 105-118. <https://doi.org/10.1016/j.ssci.2021.105118>
17. García, M., López, A., & Martínez, C. (2022). Implementation of artificial intelligence in occupational health management: A case study of Spanish manufacturing companies. *International Journal of Environmental Research and Public Health*, 19(15), 9247. <https://doi.org/10.3390/ijerph19159247>
18. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
19. Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2014). *Metodología de la investigación* (6ta ed.). McGraw-Hill Education.

20. Instituto Ecuatoriano de Seguridad Social. (2023). *Anuario estadístico 2022*. IESE.
<https://www.iess.gob.ec/documents/10162/33703/Anuario+2022.pdf>
21. Instituto Nacional de Estadística y Censos. (2023). *Directorio de empresas y establecimientos 2022*. INEC. <https://www.ecuadorencifras.gob.ec/directorio-de-empresas-y-establecimientos/>
22. International Labour Organization. (2022). *World statistics on occupational safety and health*. ILO Publications. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms_834457.pdf
23. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning with applications in R* (2nd ed.). Springer.
24. Kotter, J. P. (2012). *Leading change*. Harvard Business Review Press.
25. Ley Orgánica de Protección de Datos Personales. (2021). *Registro Oficial Suplemento No. 459*. Asamblea Nacional del Ecuador.
<https://www.telecomunicaciones.gob.ec/wp-content/uploads/2021/05/Ley-Organica-de-Proteccion-de-Datos-Personales.pdf>
26. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765-4774.
<https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>
27. Ministerio de Salud Pública. (2023). *Manual de procedimientos en salud ocupacional*. MSP Ecuador. <https://www.salud.gob.ec/wp-content/uploads/2023/04/Manual-Salud-Ocupacional-2023.pdf>
28. Ministerio de Trabajo. (2020). *Acuerdo Ministerial MDT-2020-001*. Registro Oficial Edición Especial No. 1019. <https://www.trabajo.gob.ec/wp-content/uploads/2020/04/MDT-2020-001.pdf>

29. National Institute for Occupational Safety and Health. (2022). *Criteria for a recommended standard: Occupational exposure to artificial intelligence systems.* NIOSH Publication No. 2022-106. <https://www.cdc.gov/niosh/docs/2022-106/>
30. Occupational Safety and Health Administration. (2023). *Computer workstations: Good working positions.* OSHA Publication 3148. <https://www.osha.gov/publications/osha3148>
31. Organización Internacional del Trabajo. (2023). *La seguridad y salud en el trabajo: Panorama mundial.* OIT Publications. <https://www.ilo.org/global/topics/safety-and-health-at-work/lang--es/index.htm>
32. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. <https://jmlr.org/papers/v12/pedregosa11a.html>
33. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347-1358. <https://doi.org/10.1056/NEJMra1814259>
34. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144. <https://doi.org/10.1145/2939672.2939778>
35. Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.
36. Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. *JAMA*, 320(21), 2199-2200. <https://doi.org/10.1001/jama.2018.17163>
37. Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: Benefits,

risks, and strategies for success. *NPJ Digital Medicine*, 3, 17.

<https://doi.org/10.1038/s41746-020-0221-y>

38. Tashakkori, A., & Teddlie, C. (Eds.). (2010). *SAGE handbook of mixed methods in social & behavioral research* (2nd ed.). SAGE Publications.

United Nations. (2023). *The sustainable development goals report 2023*. UN Publications.

<https://unstats.un.org/sdgs/report/2023/>

39. Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer-Verlag.
Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425-478.

<https://doi.org/10.2307/30036540>

40. World Health Organization. (2021). *Global strategy on digital health 2020-2025*.

WHO Press. <https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf>

41. World Health Organization. (2023). *Occupational health and safety*. WHO Fact Sheet.

<https://www.who.int/news-room/fact-sheets/detail/occupational-health-safety>

42. Zhang, Z., Beck, M. W., Winkler, D. A., Huang, B., Sibanda, W., & Goyal, H. (2018). Opening the black box of neural networks: Methods for interpreting neural network models in clinical applications. *Annals of Translational Medicine*, 6(11), 216.

<https://doi.org/10.21037/atm.2018.05.32>

ÍNDICE DE ANEXOS

ANEXO A. Tablas de exposición por sector	84
ANEXO B. Trazabilidad de variables y algoritmos	85
ANEXO C. Plantilla de historia clínica ocupacional	86
ANEXO D. Instrumentos de validación clínica	87
ANEXO E. Reporte gerencial simulado y KPIs	88

Anexo A: Tablas de exposición por sector

Esta tabla resume la caracterización de los sectores económicos incluidos en la muestra, así como el riesgo dominante identificado y los exámenes médicos más indicados para cada uno, basado en la prevalencia de exposición ocupacional registrada.

<i>Sector Económico</i>	<i>Riesgo Dominante</i>	<i>% Exposición Alta</i>	<i>Exámenes Más Indicados</i>
<i>Construcción</i>	Ergonómico	78.2	Rx Columna, Ecografía Muñeca
<i>Agricultura</i>	Químico	67.2	Espirometría, Colinesterasas
<i>Farmacéutico</i>	Psicosocial	18.2	Glucosa, Perfil Lipídico
<i>Manufactura</i>	Físico	52.6	Audiometría, Rx Tórax
<i>Exequiales</i>	Psicosocial	45.7	Espirometría, Glucosa
<i>Ferretería</i>	Ergonómico	38.5	Rx Columna, Optometría

Anexo B: Trazabilidad de variables y algoritmos

Se presenta la relación entre variables predictoras utilizadas por los algoritmos de aprendizaje automático, el examen médico recomendado y su impacto económico estimado con base en precios reales de proveedores.

<i>Variable Predictora</i>	<i>Algoritmo Asociado</i>	<i>Examen Asociado</i>	<i>SHAP Value Medio</i>	<i>Costo Estimado (USD)</i>	<i>Justificación Clínica</i>
<i>Tipo de Exposición</i>	Random Forest	Rx Columna	0.34	25.0	Asociado a riesgos ergonómicos en construcción
<i>Edad</i>	Gradient Boosting	Perfil Lipídico	0.22	7.2	Mayores de 45 años con riesgo metabólico
<i>Antecedentes Patológicos</i>	Neural Network	Espirometría	0.18	19.0	Antecedentes respiratorios recurrentes en exposición química

ANEXO C: Plantilla de historia clínica ocupacional

A siguiente plantilla fue utilizada como instrumento de recolección de datos clínicos ocupacionales durante la investigación. Este formato fue implementado por la empresa hiram y contiene las 23 variables principales que fueron procesadas por el sistema de inteligencia artificial. Se encuentra estructurada de manera estandarizada, de conformidad con las normativas del ministerio de salud pública y del instituto ecuatoriano de seguridad social, con campos clínicos y laborales obligatorios para todo proceso de vigilancia médica ocupacional.

ID	Edad	Género	Nivel Educativo	Económico	Puesto	Exposito	Patologías	Lámen	Biomasa	Glucosa	Perfil	Lijen	Audición	Espiración	Oftalmos	Columna	Rx Tórax	Igrafia	Muñ	Colinest	Gatos	Total	U	Procesamiento (min)	
1	56	Femenino	Superior	Farmaceút	72	Mecánico	Alto	No	Sí	Sí	Sí	Sí	Sí	Sí	Sí	No	Sí	No	Sí	Sí	Normal	50,61	21		
2	45	Femenino	Secundari	Ferretería	62	Químico	Medio	No	Sí	No	Sí	Sí	Sí	Sí	No	Sí	Sí	Sí	Sí	Sí	Normal	104,15	19,9		
3	52	Femenino	Secundari	Agricultur	104	Psicosocial	Medio	No	No	Sí	No	No	No	Sí	Sí	No	Sí	Sí	Sí	Sí	Normal	65,02	14,4		
4	41	Femenino	Superior	Manufact	109	Ergonómít	Medio	Sí	Sí	Sí	Sí	Sí	Sí	Sí	No	Sí	Sí	No	Sí	Anormal	84,68	23,8			
5	50	Femenino	Secundari	Ferretería	88	Mecánico	Alto	No	No	Sí	Sí	Sí	Sí	No	Sí	Sí	Sí	Sí	Sí	Normal	88,51	16,5			
6	49	Femenino	Superior	Exequiale	15	Químico	Medio	No	No	Sí	No	Sí	Sí	Sí	Sí	Sí	Sí	Sí	No	Sí	Normal	98,06	16,7		
7	18	Masculino	Secundari	Construcci	116	Mecánico	Alto	Sí	No	Sí	Sí	Sí	Sí	No	Sí	Sí	Sí	Sí	Sí	No	Anormal	86,92	9,6		
8	54	Masculino	Superior	Farmaceút	113	Físico	Alto	Sí	Sí	No	Sí	Sí	Sí	No	Sí	Sí	No	Sí	Sí	Sí	Normal	105,28	25		
9	28	Masculino	Primaria	Construcci	73	Psicosocial	Alto	No	No	Sí	Sí	Sí	No	Sí	Sí	No	Sí	Sí	No	Sí	Normal	74,64	21,4		
10	48	Femenino	Secundari	Manufact	16	Físico	Alto	No	Sí	No	No	Sí	Sí	Sí	No	Sí	Sí	No	Sí	Anormal	97	23,5			
11	52	Femenino	Primaria	Farmaceút	25	Mecánico	Medio	Sí	Sí	No	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Normal	121,3	18,7			
12	39	Femenino	Secundari	Ferretería	109	Químico	Alto	No	No	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	No	Anormal	83,85	21,6		
13	58	Masculino	Superior	Construcci	46	Mecánico	Alto	No	Sí	Sí	Sí	Sí	Sí	Sí	No	No	No	Sí	Sí	Sí	Normal	67,61	14,2		
14	59	Femenino	Secundari	Construcci	9	Químico	Alto	No	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Anormal	88,23	16		
15	49	Masculino	Secundari	Farmaceút	25	Físico	Alto	Sí	No	Sí	Sí	No	Sí	Sí	Sí	No	Sí	Sí	Sí	Sí	Normal	123,04	27,2		
16	57	Femenino	Primaria	Agricultur	106	Químico	Alto	Sí	Sí	Sí	No	No	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Normal	68,29	22,6		
17	33	Masculino	Superior	Exequiale	22	Mecánico	Alto	Sí	Sí	No	Sí	Sí	Sí	No	Sí	Sí	Sí	Sí	Sí	Sí	Normal	83,41	11,8		
18	40	Femenino	Secundari	Exequiale	100	Físico	Medio	Sí	No	Sí	Sí	Sí	Sí	No	No	No	Sí	No	Sí	Sí	Normal	90,6	16,6		
19	60	Masculino	Secundari	Construcci	33	Físico	Medio	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Normal	95,14	17,1		
20	53	Masculino	Superior	Construcci	36	Físico	Alto	No	No	Sí	Sí	Sí	No	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Normal	82,31	34,3		
21	42	Femenino	Secundari	Farmaceút	49	Psicosocial	Bajo	No	Sí	Sí	No	No	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	No	Normal	76,71	18	
22	48	Femenino	Secundari	Exequiale	107	Químico	Medio	Sí	No	Sí	Sí	Sí	Sí	Sí	No	Sí	Sí	Sí	Sí	Sí	Normal	104,52	15,6		
23	24	Masculino	Secundari	Manufact	70	Psicosocial	Bajo	No	No	Sí	No	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Normal	94,59	18,4		
24	42	Femenino	Secundari	Exequiale	6	Físico	Bajo	No	Sí	No	Sí	No	Sí	Sí	No	Sí	Sí	Sí	Sí	Sí	Normal	100,28	18,9		
25	28	Femenino	Superior	Ferretería	76	Psicosocial	Medio	Sí	No	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Normal	85,36	21,2		
26	22	Masculino	Superior	Agricultur	56	Físico	Alto	Sí	Sí	No	Sí	No	Sí	Sí	Sí	No	Sí	Sí	Sí	Sí	Normal	92,34	27,6		
27	22	Femenino	Superior	Exequiale	53	Mecánico	Alto	No	Sí	Sí	Sí	Sí	Sí	No	No	Sí	Sí	Sí	Sí	Sí	Normal	79,43	22,5		

ANEXO D: Instrumentos de validación

Instrumentos diseñados para validar clínica y operacionalmente el sistema de inteligencia artificial, basado en panel de expertos y escalas reconocidas de usabilidad.

Instrumento	Descripción	Formato	Aplicado a
Cuestionario de concordancia clínica	Evalúa la coincidencia entre recomendaciones del sistema y criterio médico ($\alpha=0.87$)	Escala Likert 5 puntos	5 médicos ocupacionales
Escala SUS (System Usability Scale)	Mide percepción de usabilidad entre médicos ocupacionales	Escala Likert 5 puntos	5 médicos ocupacionales
Rúbrica de validación de recomendaciones IA	Evalúa pertinencia y seguridad clínica de las recomendaciones	Escala ponderada 1-10	Panel mixto (epidemiólogo, médicos laborales, gestor de salud)

ANEXO E: Reportes gerenciales simulados

Indicadores clave proyectados como resultado de la implementación del sistema propuesto, en términos de impacto clínico, financiero y operativo dentro de una PYME ecuatoriana promedio.

<i>Indicador</i>	<i>Valor Proyectado</i>	<i>Interpretación</i>
<i>Ahorro mensual estimado</i>	\$160.70	Basado en promedio de 35 trabajadores por PYME
<i>Reducción de exámenes innecesarios</i>	38.7%	Comparado con protocolo tradicional generalizado
<i>Retorno de inversión proyectado (ROI)</i>	2.5 años	ROI validado con costos reales y beneficios
<i>Concordancia con criterio médico</i>	84%	Validación con médicos ocupacionales expertos
<i>Tiempo promedio de recomendación</i>	18.2 minutos	Desde ingreso de datos hasta recomendación
<i>Índice de satisfacción médica</i>	78.4 (SUS Score)	Indica buena usabilidad según norma internacional