

Arquitectura de un módulo para la identificación de factores de riesgo para la detección de desórdenes hepáticos a partir del análisis de biomarcadores utilizando métodos de ensambles de aprendizaje automático

Luis Rodolfo Cabrera-Elías¹, José Luis Sánchez-Cervantes²,
Giner Alor-Hernández¹, Beatriz Alejandra Olivares-Zepahua¹,
Luis Ángel Reyes-Hernández¹

¹ Tecnológico Nacional de México
Instituto Tecnológico de Orizaba
México

² Consejo Nacional de Ciencia y Tecnología,
Tecnológico Nacional de México,
México

{ml6011057, jose.sc, giner.ah, luis.rh}@orizaba.tecnm.mx,
bolivares@ito-depi.edu.mx

Resumen. La salud es importante para llevar una vida estable y duradera, pero la falta de cuidado aumenta el riesgo de enfermedades y dañar órganos, especialmente el hígado. En este trabajo se presenta la arquitectura de un módulo de aprendizaje automático basado en ensambles para identificar factores de riesgo en la detección de desórdenes hepáticos a través del análisis de biomarcadores. Además, se presenta el análisis de los trabajos relacionados con el tema y se examinan algoritmos de ensamble de aprendizaje automático, como bagging, boosting, para la detección de desórdenes hepáticos. Como trabajo a futuro se identificarán los principales biomarcadores para la detección de factores de riesgo, así como se diseñará un modelo de entrenamiento basado en el algoritmo de ensamble para la identificación de factores de riesgo. Finalmente, el módulo se entrenará y se integrará con interfaces Web y un repositorio de información.

Palabras clave: Biomarcadores, daño hepático, ensambles de aprendizaje automático.

Architecture of a module for identification of risk factors for the detection of liver disorders from biomarker analysis using ensemble machine learning method

Abstract. Health is important to lead a stable and long-lasting life, but lack of care increases the risk of diseases and damaging organs, especially the liver. This work presents the architecture of a machine learning module based on ensembles

to identify risk factors in the detection of liver disorders through biomarker analysis. In addition, an analysis of related works on the topic is presented and machine learning ensemble algorithms, such as bagging and boosting, are examined for the detection of liver disorders. As future work, the main biomarkers for detecting risk factors will be identified, and a training model based on the ensemble algorithm will be designed for the identification of risk factors. Finally, the module will be trained and integrated with web interfaces and an information repository.

Keywords: Biomarkers, liver damage, machine learning ensembles.

1. Introducción

El daño hepático causado por obesidad (hígado grado), o hepatotoxicidad generada por el consumo de medicamentos, drogas o incluso remedios caseros, por mencionar algunos casos; se está convirtiendo en un problema principal en la salud en México, tal como, la enfermedad por hígado graso no alcohólico (EHGNA) que afecta a un tercio de la población mundial, siendo México uno de los países cuya población reúne varios factores de riesgo para esta enfermedad y su prevalencia podría superar el 50% [1].

Los ensambles de aprendizaje automático son algoritmos integrados por algoritmos más simples (tradicionales), que permiten incrementar los porcentajes de sensibilidad y especificidad al momento de clasificar y detectar. Por lo cual, el uso de ensambles de aprendizaje automático permitirá identificar los factores de riesgo para la detección de desórdenes hepáticos a partir del análisis de biomarcadores.

En este artículo se propone un módulo de aprendizaje automático basado en ensambles para identificar factores de riesgo en la detección de desórdenes hepáticos y tomar medidas preventivas necesarias. El resto del trabajo incluye una revisión del estado del arte en la sección 2, la arquitectura del módulo en la sección 3, resultados de los datos en la sección 4, y las conclusiones y el trabajo futuro en la sección 5.

2. Trabajos relacionados

En esta sección se exponen los trabajos más destacables, los cuales tienen relación con las técnicas de ensamble automático. Doganer et al. [2] discutieron la importancia del uso de un ensamble de aprendizaje automático para la detección temprana de carcinomas de células renales, comparando métodos de stacking, bagging y boosting.

Los algoritmos de stacking obtuvieron los mejores resultados con una precisión del 86.7%. Verma et al. [3] compararon un nuevo método de ensamble de aprendizaje automático llamado "BBS method" (Bagging, Boosting, Stacking method, en español: Método de Bagging, Boosting y Stacking) con los algoritmos clasificadores apropiados.

El "BBS method" obtuvo una mayor precisión que sus contrapartes individuales. En [4], Buyrukoglu propuso mejorar los enfoques existentes para la detección temprana del Alzheimer mediante el uso de bagging, boosting y stacking, lo que resultó en una mejora en la clasificación del 3.2% al 7.2% respecto a los métodos anteriores.

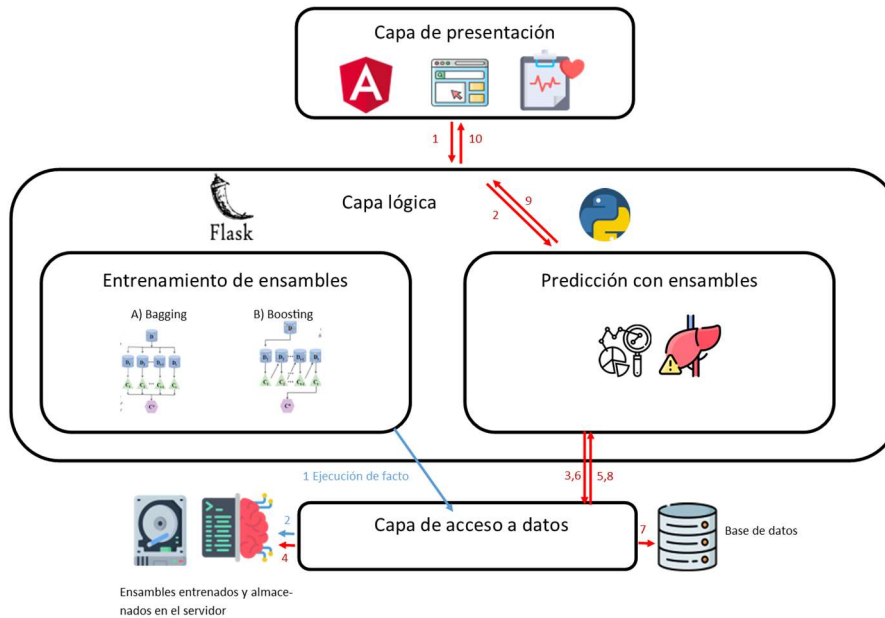


Fig. 1. Arquitectura del módulo de detección.

En [5], se utilizó un algoritmo de ensamble automático de bagging y boosting con 5 algoritmos de clasificación para predecir los síntomas de un infarto al miocardio en una etapa temprana. El algoritmo de bagging en conjunto con Random Forest obtuvo la mayor precisión con un 96.50%. Yadav et al. [6] compararon tres algoritmos basados en reglas utilizando ensambles de bagging y boosting para abordar la problemática de la diabetes mellitus.

El ensamble de bagging obtuvo una precisión del 98%. En [7], se propuso el uso de un ensamble de Bagging Weighted Voting (EBWvc) para abordar la problemática de la detección tardía del cáncer de mama. El EBWvc obtuvo una clasificación superior a las demás técnicas existentes.

Rahman F y Mahmood M [8] propusieron el uso de tres algoritmos clasificadores base, uno de ellos dirigido por bagging, para predecir enfermedades cardíacas mediante biomarcadores destacables. El modelo creado a partir del uso de Random Forest con Bagging obtuvo una precisión del 85.18%.

En [9], se mostró la importancia del uso de los registros electrónicos de salud (EHR) para obtener las variables de diagnóstico más relevantes para la predicción de la hepatitis C y la cirrosis mediante el uso de ensambles de aprendizaje automático.

Santos et al. [10] propusieron un enfoque de agrupamiento con sobre muestreo robusto para tratar con datos incompletos y predecir la supervivencia en un lapso de un año para pacientes con carcinoma hepatocelular.

En [11], se propuso la incorporación de la regresión lógica homogénea para apoyar la toma de decisiones y diagnósticos rápidos en el caso de un accidente cerebrovascular. La regresión logística homogénea obtuvo una mayor precisión que su contraparte simple. En la tabla 1 se presenta la comparación de los enfoques analizados en la sección 2. Trabajos Relacionados.

Tabla 1. Comparativa de los trabajos relacionados.

Artículo	Problema	Contribución	Tecnologías empleadas	Resultados
Doganer et al. [2]	El cáncer de células renales es normalmente asintomático, incluso en las etapas avanzadas y la posibilidad de dar con un diagnóstico temprano es baja.	Un modelo de ensamble automático de alto rendimiento para la detección temprana de carcinomas renales.	Ensamblados de boosting, bagging, stacking, tales como: IB1, IBk, Kstar, LWL, REPTree, Random Forest y SMO.	El algoritmo de stacking REPTree obtuvo la más alta precisión del 86.7%.
Hakim et al. [5]	La carencia de detección temprana de los síntomas de un infarto al miocardio.	La comparación de dos enfoques de ensamble complementados por cinco algoritmos clasificadores base.	Bagging, Boosting, Support Vector Machine, K-Nearest Neighbor, Naive Bates, Árbol de decisión, Random Forest	El ensamble de bagging en conjunción con Random Forest obtuvo la clasificación de precisión más alta del 96.50%
Anisha y Saranya [11]	La falta de detección de los síntomas de un accidente cerebrovascular y las decisiones rápidas a tomar.	Un sistema que provee ayuda en la detección de un accidente cerebrovascular y las recomendaciones a seguir en dicho momento.	Regresión logística homogénea	El sistema propuesto obtuvo errores reducidos en la predicción y una precisión del 91% frente a la regresión logística simple para el diagnóstico de accidentes cerebrovasculares
Buyrukoglu[4]	La detección tardía del Alzheimer provoca pérdidas a la memoria del paciente.	Mejorar los enfoques ya existentes mediante la implementación de ensambles de aprendizaje automático.	Boosting, Bagging, Stacking	La clasificación incrementó en un porcentaje de entre 3.2% y el 7.2% respecto a los enfoques previamente diseñados.
Rahman y Mahmood [8]	La carencia de la capacidad de predicción para enfermedades coronarias.	Un modelo que permita predecir una enfermedad cardíaca, comparando tres clasificadores base con bagging.	Random Forest con ensamble de bagging, K-Nearest, Naive Bayes	El algoritmo de Random Forest con bagging alcanzó la puntuación más alta para predecir enfermedades coronarias.

Tabla 2. Comparativa de los atributos e instancias de los conjuntos de datos.

Conjunto de datos	Repositorio	Instancias	Atributos
Cirrhosis Prediction Dataset	Kaggle	424	19
Indian Liver Patient Dataset (ILPD):	UCI Machine Learning Repository	583	10
Liver Disorders Data Set	UCI Machine Learning Repository	345	6
Non-alcohol fatty liver disease (NAFLD)	Kaggle	17549	10

Tabla 3. Métricas obtenidas en el conjunto de datos Cirrhosis Prediction Dataset

Clasificador base	Exactitud	Precisión	Exhaustividad	F1-Score	Area ROC
LR	71.43%	62.50%	45.45%	52.63%	65.41%
SVC	72.22%	62.50%	40.91%	50.70%	64.97%
KNN	68.25%	58.33%	31.82%	41.18%	59.81%
DT	73.02%	67.86%	43.18%	52.78%	66.10%
RF	69.05%	57.58%	43.18%	49.35%	63.05%

3. Arquitectura propuesta para el módulo

En esta sección se describe la propuesta y las tecnologías con las cuales se pretende desarrollar el módulo propuesto en este trabajo, además, se presenta los resultados obtenidos en el entrenamiento de los algoritmos de ensamble de aprendizaje automático seleccionados.

3.1. Diseño del módulo

En la Fig. 1 se muestra el diagrama de la arquitectura propuesta para este módulo de detección de factores de riesgo de desórdenes hepáticos. La arquitectura presentada en la Fig. 1 está separada en tres capas de las cuales se explicarán a detalle a continuación.

3.2. Capa de presentación

En esta capa se encuentra la interfaz gráfica en la que el médico o personal sanitario interactuará con la aplicación, se designó al framework Angular [12] el cual destaca por su flexibilidad y su capacidad de trabajar con aplicaciones Web grandes. En dicha interfaz, se le proporcionará al sistema los biomarcadores de un paciente en específico para que el módulo proporcione una predicción pertinente y mostrar en la pantalla dicho resultado obtenido.

Tabla 4. Métricas obtenidas en el conjunto de datos Indian Liver Patient Dataset.

Clasificador base	Exactitud	Precisión	Exhaustividad	F1-Score	Area ROC
LR	72.00%	74.53%	93.75%	83.04%	53.26%
SVC	73.14%	73.14%	100.00%	84.49%	50.00%
KNN	68.57%	76.64%	82.03%	79.25%	56.97%
DT	73.14%	77.93%	88.28%	82.78%	60.10%
RF	73.71%	80.15%	85.16%	82.58%	63.85%

Tabla 5. Métricas obtenidas en el conjunto de datos Liver Disorders Data set.

Clasificador base	Exactitud	Precisión	Exhaustividad	F1-Score	Area ROC
LR	75.00%	74.29%	86.67%	80.00%	72.88%
SVC	72.12%	75.41%	76.67%	76.03%	71.29%
KNN	67.31%	70.97%	73.33%	72.13%	66.21%
DT	70.19%	69.86%	85.00%	76.69%	67.50%
RF	73.08%	72.22%	86.67%	78.79%	70.61%

Cabe resaltar, que al usuario del módulo se le solicitarán únicamente los biomarcadores más importantes entre los que hasta el momento se han identificado los siguientes: tiempo de protrombina (en segundos), plaquetas por ml³ y albúmina para cirrosis, y fosfatasa alcalina, aspartato aminotransferasa (AST) y alanina aminotransferasa (ALT) para desórdenes hepáticos y peso, género, altura e índice de masa corporal para hígado graso no alcohólico, los cuales se determinarán de manera definitiva en los trabajos a futuro de la presente investigación, por lo que no se le permitirá proseguir con la predicción en caso de no proporcionarlos.

3.3 Capa lógica

La capa lógica es la más importante de dicho módulo, ya que aquí se aloja el núcleo principal de este sistema, como backend se propone utilizar a Python [13] por ser un lenguaje de programación versátil y ampliamente utilizado en el campo de la inteligencia artificial en conjunción del framework Flask [14] que está diseñado para este lenguaje, además, permitirá darle funcionalidad de API REST al módulo propuesto.

Una vez introducidos los datos en la capa anterior son procesados en la sección de predicción de ensambles mostrada en la Fig. 1, donde se ejecutan una serie de modelos previamente entrenados mediante algoritmos de ensamble de aprendizaje automático.

El módulo estará integrado por dos de los mejores clasificadores base de boosting y de bagging. El sistema determinará de manera automática cual utilizar dependiendo de la carga de trabajo actual que sostenga. Consecutivamente almacenará dicha predicción

Tabla 6. Métricas obtenidas en el conjunto de datos Non-alcohol fatty liver disease.

Clasificador base	Exactitud	Precisión	Exhaustividad	F1-Score	Area ROC
LR	92.27%	56.00%	10.10%	17.11%	54.71%
SVC	92.10%	0.00%	0.00%	0.00%	50.00%
KNN	92.00%	47.71%	12.50%	19.81%	55.66%
DT	92.27%	66.67%	4.33%	8.13%	52.07%
RF	92.33%	54.29%	18.27%	27.34%	58.47%

en la base de datos y el resultado obtenido se mostrará al médico para que proceda a prescribir un tratamiento oportuno.

La sección de entrenamiento de los algoritmos de ensamble que se muestra en la Fig. 1, se ejecutará únicamente por primera vez o al mejorar su entrenamiento. Si se realiza constantemente el entrenamiento, afectará en el rendimiento del módulo propuesto.

3.4 Capa de acceso a datos

La capa de acceso a datos provee las APIs necesarias para la comunicación de los componentes externos al sistema, tales como: la base de datos (donde se almacenarán los registros para llevar un historial) y el entrenamiento entrenado con anticipación. En la Fig. 1 se indica con flechas rojas y numeraciones el orden de ejecución de este módulo para evitar confusiones y se define de la siguiente manera:

- Punto 1: Lectura de biomarcadores en la interfaz gráfica.
- Punto 2: Recepción de la lectura en la capa lógica.
- Puntos 3, 4 y 5: Uso del ensamble entrenado con anterioridad y almacenado en la capa de acceso a datos para realizar una predicción con la información recibida y la obtención del resultado.
- Punto 6 y 7: Almacenaje del resultado en la base de datos de la aplicación para generar una historia del paciente.
- Punto 8 y 9: Devolver el resultado a la Capa de presentación.
- Punto 10: Observar el resultado en pantalla.

4. Conjuntos de datos para la identificación de desórdenes hepáticos

En esta sección se describirán 4 conjuntos de datos que se utilizarán para detección de desórdenes hepáticos mediante la ejecución de este módulo.

- **Cirrhosis prediction dataset:** [15] contiene información sobre pacientes con cirrosis hepática crónica y no crónica. Incluye 27 variables de entrada que incluyen edad, sexo, síntomas, signos de laboratorio, hallazgos histopatológicos y

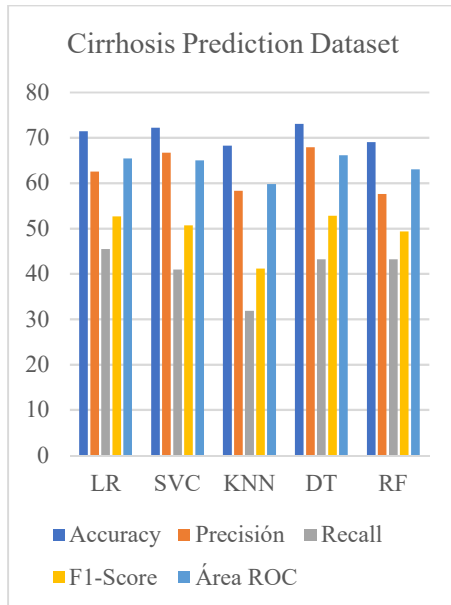


Fig. 2. Gráfica de las métricas obtenidas en el conjunto de datos cirrhosis prediction dataset.

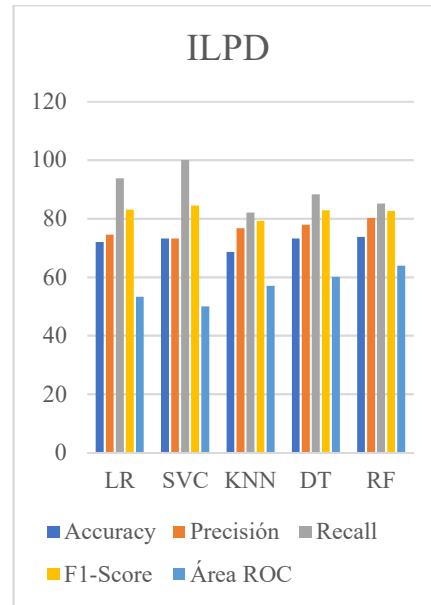


Fig. 3. Gráfica de las métricas obtenidas en el conjunto de datos indian liver patient dataset.

diagnósticos, así como una variable de salida que indica si el paciente tiene cirrosis crónica o no. El objetivo del dataset es predecir si un paciente tiene cirrosis crónica o no en función de las variables de entrada. El conjunto de datos contiene 386 casos, 186 de los cuales son casos positivos de cirrosis crónica. El dataset se encuentra en formato CSV y está disponible para descargar en Kaggle.

- **Indian liver patient dataset (ILPD):** ILPD [16] es un conjunto de datos médicos que contiene información sobre pacientes indios con y sin enfermedades hepáticas. El conjunto de datos consta de 583 instancias con 10 atributos, incluyendo información demográfica del paciente, resultados de pruebas de sangre y diagnósticos de enfermedades hepáticas. El objetivo del conjunto de datos es predecir si un paciente tiene una enfermedad hepática o no, lo que lo convierte en un conjunto de datos útil para tareas de clasificación binaria en el campo de la salud.
- **Liver disorders data set:** [17] este conjunto de datos proveniente del repositorio UCI Machine Learning contiene los resultados de pruebas de laboratorio de pacientes con trastornos hepáticos, como hepatitis viral, cirrosis, hemocromatosis y otros. El conjunto de datos consta de 345 instancias y 6 atributos, incluyendo el volumen corpuscular medio, la fosfatasa alcalina, la alanina aminotransferasa y el aspartato aminotransferasa. El objetivo del conjunto de datos es predecir si un paciente tiene o no un trastorno hepático basado en los valores de los atributos.
- **Non-alcohol fatty liver disease (NAFLD):** NAFLD [18] alojado en Kaggle contiene información sobre pacientes diagnosticados con hígado graso no alcohólico en un hospital universitario de Corea del Sur. Incluye datos demográficos y clínicos de

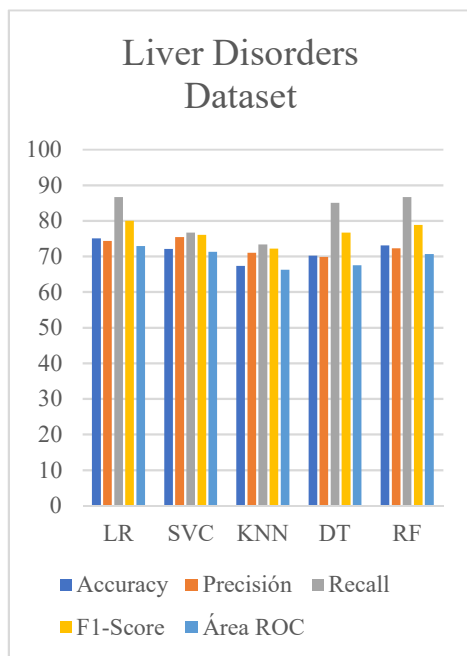


Fig. 4. Gráfica de las métricas obtenidas en el conjunto de datos liver disorders dataset.

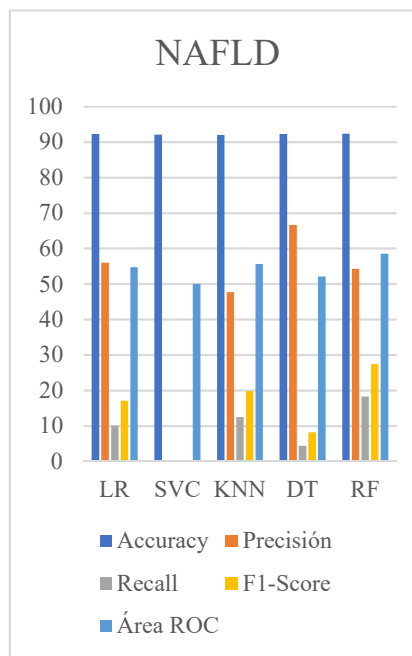


Fig. 5. Gráfica de las métricas obtenidas en el conjunto de datos non-alcohol fatty liver disease.

17,549 pacientes, como edad, género e índice de masa corporal. El objetivo del conjunto de datos es ayudar a predecir la progresión de la enfermedad y desarrollar un modelo de diagnóstico temprano.

La Tabla 2 muestra una comparación breve entre los cuatro conjuntos de datos mencionados anteriormente, haciendo hincapié en el número de instancias y los atributos.

5. Análisis comparativo de los conjuntos de datos mediante los algoritmos de ensambles con enfoque de Bagging

En esta sección se presentan la comparativa realizada a los conjuntos de datos mediante el uso de diferentes clasificadores base utilizando el enfoque de embolsado (bagging, en inglés) mediante el uso de Python y las bibliotecas de scikit-learn [19], las cuales están enfocadas en aprendizaje automático. Los algoritmos utilizados como clasificadores base para este análisis fueron los siguientes: LogisticRegression [20], SVM (Support Vector Machine) [21], KNN(K-Nearest-Neighbor) [22], DecisionTree [23] y RandomForest [24].

De los algoritmos se evaluaron y compararon métricas tales como: exactitud (accuracy), precisión (precision), exhaustividad (recall), F1-Score y el área ROC y se aplicaron a los 4 conjuntos de datos listados en la Tabla 2. Algunos conjuntos de datos

tienen mejores resultados para las predicciones que otros y esto es dependiente de los atributos de los cuales se encuentren conformados y mediante este análisis comparativo es posible determinar cuál clasificador base es más eficiente con los diferentes conjuntos que se le proporcionen.

6. Conclusiones y trabajo a futuro

La detección temprana de factores de riesgo para la detección de desórdenes hepáticos es esencial para prevenir daños graves y potencialmente mortales en el hígado. El desarrollo de este módulo tiene como objetivo ayudar a los médicos a identificar y tratar el problema antes de que se convierta en algo más grave y difícil de tratar y tomar las medidas pertinentes para dar un tratamiento que permita a los afectados por dichos desordenes mejorar su calidad de vida.

Como trabajo a futuro se considerarán los biomarcadores más importantes para la detección de los factores de riesgo para la detección de desórdenes hepáticos, tal cual, el análisis comparativo mediante el enfoque de algoritmos de ensamble de boosting, así como como los algoritmos de ensamble de aprendizaje automático a utilizar para este módulo, se tiene previsto utilizar las bibliotecas orientadas para Python enfocadas en aprendizaje automático conocidas como scikit-learn como parte del núcleo para realizar las predicciones pertinentes y proporcionar un resultado fidedigno, así como evaluar y comparar, el conjunto de datos y tipos de ensambles más precisos para esta problemática.

Para la validación del módulo propuesto en el futuro se hará uso de un caso de estudio como prueba de concepto que permita describir los resultados y conclusiones obtenidas. Finalmente se realizará una comparativa de los resultados obtenidos de los clasificadores mediante este módulo respecto a los trabajos mencionados anteriormente en la sección 2, estado del arte.

Agradecimientos. Los autores dan las gracias al Tecnológico Nacional de México por respaldar este trabajo. A sí mismo como el patrocinio por parte del Consejo Nacional de Ciencia y Tecnología (CONACYT).

Referencias

1. Bernal-Reyes, R., Castro-Narro, G., Malé-Velázquez, R., Carmona-Sánchez, R., González-Huezo, M., García-Juárez, I., Chávez-Tapia, N., Aguilar-Salinas, C., Aiza-Haddad, I., Ballesteros-Amozurrutia, M., Bosques-Padilla, F., Castillo-Barradas, M., Chávez-Barrera, J., Cisneros-Garza, L., Flores-Calderón, J., García-Compeán, D., Gutiérrez-Grobe, Y., Tijera, M. H., Kershenobich-Stalnikowitz, D., Guevara-Cetina, L. L., et. al.: Consenso mexicano de la enfermedad por hígado graso no alcohólico. *Revista de Gastroenterología de México*, vol. 84, no. 1, pp. 69–99 (2019) doi: 10.1016/j.rgmx.2018.11.007
2. Doğaner, A., Çolak, C., Küçükdurmaz, F., Ölmez, C.: Prediction of renal cell carcinoma based on ensemble learning methods. *Middle Black Sea Journal of Health Science*, vol. 7, no. 1, pp. 104–114 (2021) doi: 10.19127/mbsjohs.889492

3. Verma, A., Mehta, S.: A comparative study of ensemble learning methods for classification in bioinformatics. In: 7th International Conference on Cloud Computing, Data Science and Engineering Confluence, pp. 155–158 (2017) doi: 10.1109/confluence.2017.7943141
4. Buyrukoglu, S.: Improvement of machine learning models' performances based on ensemble learning for the detection of Alzheimer disease. In: 6th International Conference on Computer Science and Engineering pp. 102–106 (2021) doi: 10.1109/ubmk52708.2021.9558994
5. Hakim, M. A., Jahan, N., Zerín, Z. A., Farha, A. B.: Performance evaluation and comparison of ensemble based bagging and boosting machine learning methods for auto mated early prediction of myocardial infarction. In: 12th International Conference on Computing Communication and Networking Technologies, pp. 1–6 (2021) doi: 10.1109/ICCCNT51525.2021.9580063
6. Yadav, D. C., Pal, S.: An experimental study of diversity of diabetes disease features by bagging and boosting ensemble method with rule-based machine learning classifier algorithms. SN Computer Science, vol. 2, no. 1 (2021) doi: 10.1007/s42979-020-00446-y
7. Ponnaganti, N. D., Anitha, R.: A novel ensemble bagging classification method for breast cancer classification using machine learning techniques. Traitement du Signal, vol. 39, no. 1, pp. 229–237 (2022) doi: 10.18280/ts.390123
8. Rahman, F., Mahmood, M. A.: A dynamic approach to identify the most significant biomarkers for heart disease risk prediction utilizing machine learning techniques. Bangabandhu and Digital Bangladesh, pp. 12–22 (2022) doi: 10.1007/978-3-031-17181-9_2
9. Chicco, D., Jurman, G.: An ensemble learning approach for enhanced classification of patients with hepatitis and cirrhosis. IEEE Access, vol. 9, pp. 24485–24498 (2021) doi: 10.1109/access.2021.3057196
10. Santos, M. S., Abreu, P. H., García-Laencina, P. J., Simao, A., Carvalho, A.: A new cluster-based oversampling method for improving survival prediction of hepatocellular carcinoma patients. Journal of Biomedical Informatics, vol. 58, pp. 49–59 (2015) doi: 10.1016/j.jbi.2015.09.012
11. Anisha, C., Saranya, K.: Early diagnosis of stroke disorder using homogenous logistic regression ensemble classifier. International Journal of Nonlinear Analysis and Applications, vol. 12, pp. 1649–1654 (2021) doi: 10.22075/ijnaa.2021.5851
12. Angular: What is angular? (2023) angular.io/guide/what-is-angular
13. Python: What is Python? Executive Summary (2022) www.python.org/doc/essays/blurb
14. Python and REST APIs: Interacting with web services – real python (2022) realpython.com/api-integration-in-python/#flask
15. Cirrhosis prediction dataset. Kaggle (2023) www.kaggle.com/datasets/fedesoriano/cirrhosis-prediction-dataset
16. UCI machine learning repository: ILPD (indian liver patient dataset) data set (2022) archive.ics.uci.edu/ml/datasets/ILPD+%28Indian+Liver+Patient+Dataset%29
17. UCI machine learning repository: Liver disorders data set (2023) archive.ics.uci.edu/ml/datasets/liver+disorders
18. Non-alcohol fatty liver disease. Kaggle (2023) www.kaggle.com/datasets/jamescorden/nonalcohol-fatty-liver-disease?select=naflid.csv

19. scikit-learn: Machine learning in Python. scikit-learn 1.2.2 documentation (2023) scikit-learn.org/stable/
20. sklearn.linear_model: LogisticRegression. scikit-learn 1.2.2 documentation, scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
21. 1.4. Support vector machines. Scikit-learn 1.2.2 documentation (2023) scikit-learn.org/stable/modules/svm.html
22. sklearn.neighbors: KNeighborsClassifier. Scikit-learn 1.2.2 documentation (2023) scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html
23. sklearn.tree.: DecisionTreeClassifier. Scikit-learn 1.2.2 documentation (2023) scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html
24. sklearn.ensemble: RandomForestClassifier. Scikit-learn 1.2.2 documentation (2023) scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html