22nd ECORFAN® International Conference - Science, Technology and Innovation

*Booklets*

RENIECYT 1702902 SECIHTI

RENIECYT - LATINDEX - Research Gate - DULCINEA - CLASE - Sudoc - HISPANA - SHERPA UNIVERSIA - Google Scholar DOI - REDIB - Mendeley - DIALNET - ROAD - ORCID - V|LEX - EBSCO

# Title: Application of machine learning models for the classification of diseases requiring hospitalization: an approach based on NOM-035 and the Social and Solidarity Economy in Mexico

**Authors:** Lara-Escamilla, Samuel, Rueda-Gutiérrez, Allan Balam, Morales-Piña, Diana and Rodríguez-Molina, Alejandro

Instituto Tecnológico de Tlalnepantla • ABE-1839-2022 • 0000-0002-2337-0878 • 609026
Instituto Tecnológico de Tlalnepantla • OJV-4600-2025 • 0009-0002-1887-8148 • 295451
Instituto Tecnológico de Tlalnepantla • OJV-4621-2025 • 0009-0003-1585-9803 • 217846
Universidad Autónoma de la Ciudad de México • AAY-9817-2020 • 0000-0002-6901-3833 • 555081

**Editorial label MARVID:** 607-8695
**BMARVID Control Number:** 2025-01
**BMARVID Classification (2025):** 121225-0001
**RNA:** 03-2010-032610115700-14
**Pages:** 13

**SECIHTI classification:**
**Area:** Engineering
**Field:** Technological Sciences
**Discipline:** Computer Technology
**Subdiscipline:** Artificial Intelligence

**MARVID-México**
Park Pedregal Business. 3580, Anillo Perif., San Jerónimo Aculco, Álvaro Obregón, 01900 Ciudad de México, CDMX,
Phone: +52 1 55 6159 2296
Skype: MARVID-México S.C.
E-mail: contact@marvid.org
Facebook: MARVID-México S. C.
X: @Marvid_México

**www.marvid.org**

**Holdings**
Mexico, Bolivia, Spain, Ecuador, Peru, Colombia, Cameroon, El Salvador, Taiwan, Paraguay, Guatemala, Democratic Republic of Congo, Nicaragua

# Introduction

The use of artificial intelligence (AI) and machine learning has become increasingly relevant in recent years as an analytical tool in predictive medicine and occupational health (Goodfellow, Bengio, & Courville, 2016; LeCun & Hinton, 2015).

Among the most outstanding techniques for this are RNA and SVM, since they allow the identification of complex patterns and the making of accurate classifications in contexts where traditional statistical methods have limitations (Nwanosike, E. M., Conway, B. R., Merchant, H. A., & Hasan, S. S., 2022).

# Introduction

Hospital overcrowding in Mexico and many other countries is a significant problem, often attributed to the hospitalization of patients based on isolated symptoms and clinical indicators (Cisterna-García, A., Guillén-Teruel, A., Caracena, M., Pérez, E., Jiménez, F., Francisco-Verdú, F. J., & Botía, J. A., 2022). However, patient admissions cannot always be justified by this information alone. The lack of tools to assess the impact of various clinical indicators, as well as psychosocial and sociodemographic factors, leads to medical interventions that reduce the efficiency of the healthcare system.

# Introduction

Therefore, this research proposes the development of machine learning models, specifically Multilayer Perceptrons (MLP) and Support Vector Machines (SVM), to recommend hospitalization of patients when the combination of key physiological variables, the psychosocial indicators of NOM-035 and the principles of ESS, suggest that there is a high health risk.

# Metodology

A quantitative, explanatory, and predictive approach was adopted, based on a non-experimental, cross-sectional design. The objective was to develop and validate machine learning models to recommend hospitalization for patients with high health risks. The model makes recommendations based on a supervised binary classification of characteristics related to clinical and psychosocial variables, avoiding unnecessary admissions based on isolated indicators (e.g., high blood pressure or elevated glucose).

# Metodology

A dataset called Dataset. CSV was used, consisting of 500 and 1000 observations respectively, which served as the basis for modeling and classifying the hospitalization requirement status coded in binary format (1 = hospitalized, 0 = not hospitalized).

Observations include: age, sex, body mass index (BMI); measurements of systolic pressure, diastolic pressure, total cholesterol and glucose; in addition to information such as level of physical activity, tobacco use, diabetes and family history of disease.

# Metodology

Prior to training the machine learning models, a logical-mathematical modeling process based on operations research (OR) was applied to transform the original clinical values into six derived variables, representing composite indicators of clinical and psychosocial risk. The purpose was to construct derived variables that more coherently represented the interactions between physiological factors and risk conditions associated with hospitalization.

# Metodology

**Table 1.** Coding structures used.

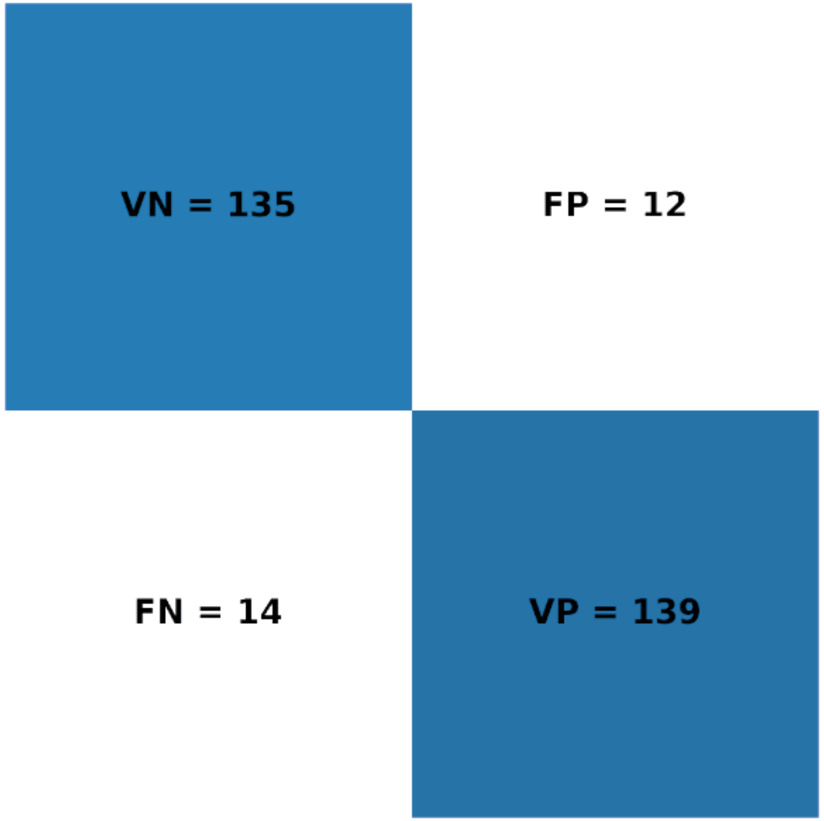
| Categoría | Ejemplos | Tipo | Codificación |
|---|---|---|---|
| Sociodemográficas | Edad, sexo, antigüedad, tipo de organización | Nominal | *One-hot* |
| Diagnóstico médico | Presión arterial, glucosa, cortisol, IMC | Continua | Normalización /Binarización |
| Psicosociales (NOM-035) | Estrés, Estrés térmico | Ordinal | 0-1-2 |

**Table 2.** Coding structures used.

| Operador | Expresión lógica | Restricciones |
|---|---|---|
| AND | $z = x \wedge y$ | $z \leq x; z \leq y; z \geq x + y - 1$ |
| OR | $z = x \vee y$ | $z \geq x; z \geq y; z \leq x + y$ |
| XOR | $z = x \bigoplus y$ | $z = x + y - 2(x \cdot y)$ |
| NAND | $\neg(x \wedge y)$ | $z = 1 - (x \cdot y)$ |
| NOR | $\neg(x \vee y)$ | $z = 1 - (x + y)$ |

# Results

**Figura 1.** Matriz de confusión del modelo MLP




•**VP (139 / 46.3%) Cases correctly classified as hospitalized.** This reflects the model's effectiveness in detecting actual cases.

•**VN (135 / 45.0%) Cases not hospitalized but correctly classified.** This indicates high specificity, avoiding false alarms.

•**FP (12 / 4.0%) Cases incorrectly classified as hospitalized.** This represents a slight and acceptable overfit in medical prevention contexts.

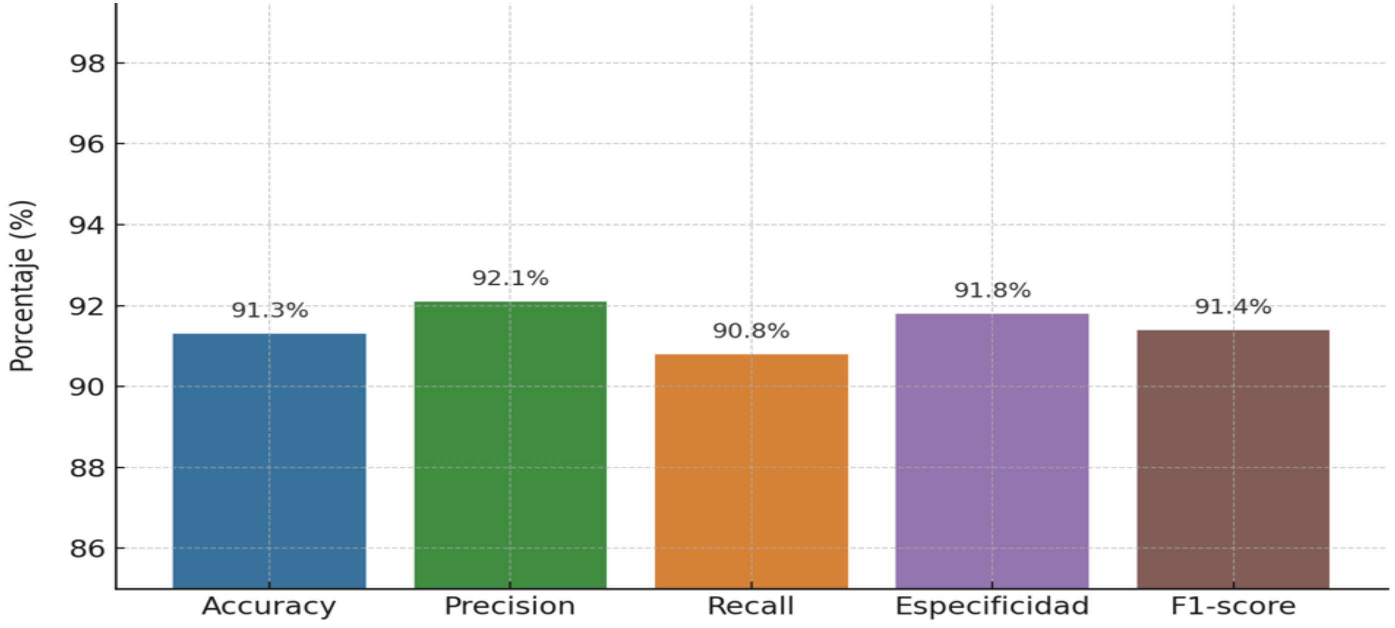
•**FN (14 / 4.7%) Hospitalized cases that the model did not detect.** These are the most clinically sensitive errors, and their proportion is low.

# Results

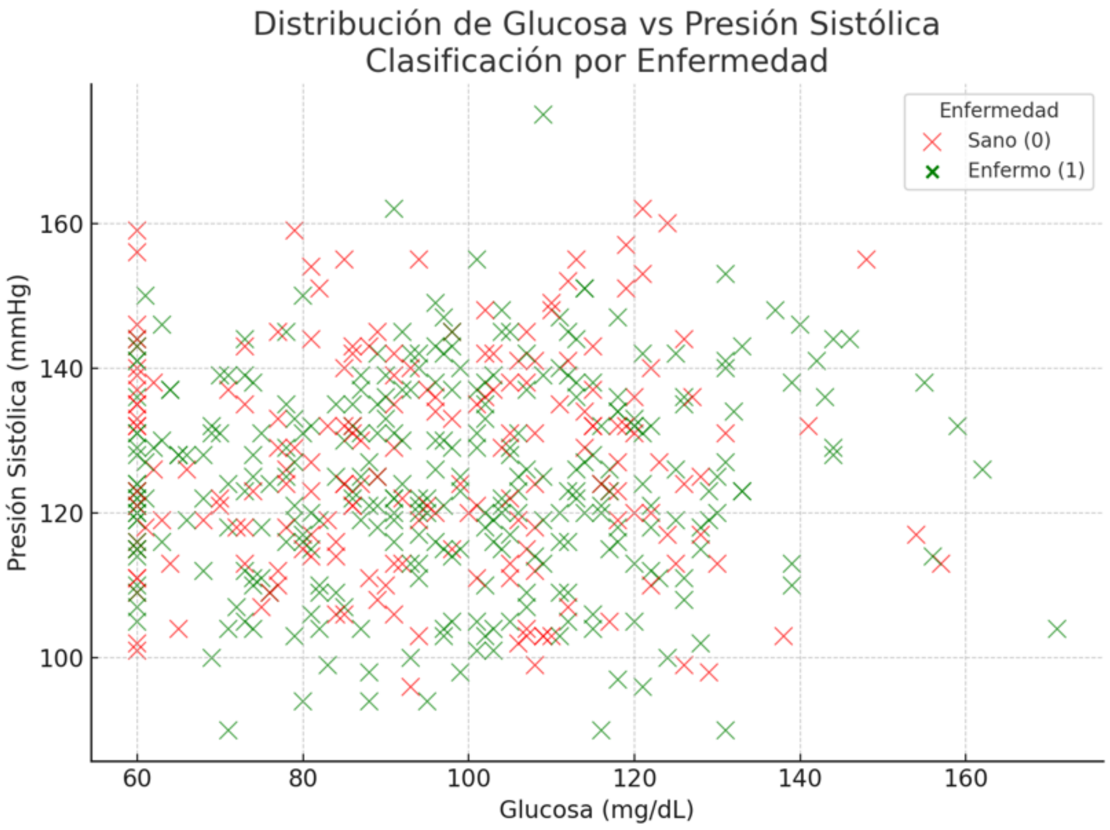
**Table 3**. MLP model performance.

| Métrica | Valor (%) |
|---|---|
| *Accuracy* | 91.3 |
| *Precision* | 92.0 |
| *Recall* | 90.9 |
| *F1-score* | 91.4 |

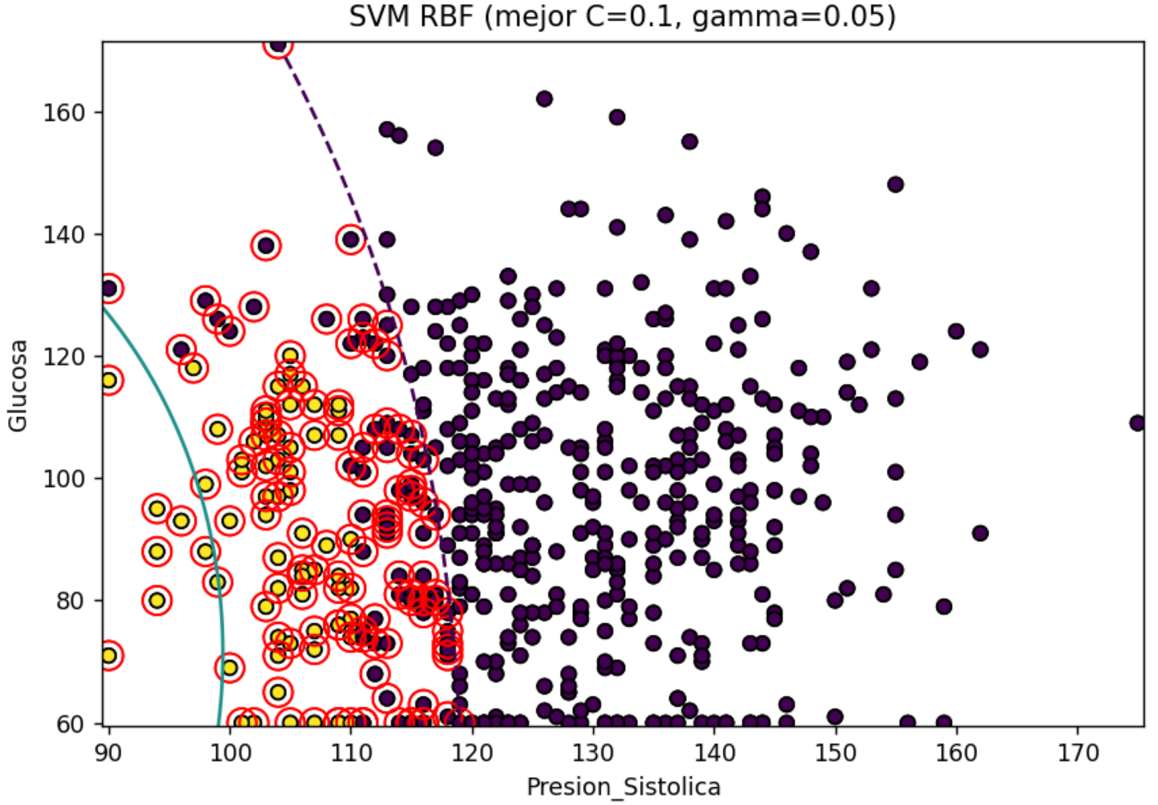
**Figure 2.** MLP performance illustrated by a bar chart.



# Results




Distribution of sick patients in the space formed by the variables of glucose and systolic pressure.




Decision boundary of the SVM model with radial core (RBF) considering the variables of glucose and systolic pressure.

# Results

**Table 4.** Aspects to consider regarding the use of MLP and SVM models.

| Aspecto | MLP | SVM |
|---|---|---|
| Naturaleza del modelo | Aprendizaje adaptable, basado en pesos sinápticos | Modelo estructurado, basado en vectores de soporte |
| Tipo de relación detectada | No lineal, compleja, dependiente del ajuste de pesos | Lineal o no lineal, controlada por el núcleo |
| Ventajas | Generalización progresiva, tolerante a ruido, interpretable en capas | Buena precisión con menos datos, evita sobreajuste, eficiente en alta dimensión |
| Limitaciones | Requiere ajuste fino de parámetros (número de neuronas, tasa de aprendizaje) | Sensible a la selección de C y γ. Puede ser costoso en grandes conjuntos de datos. |
| Aplicación sugerida | Escenarios con alta variabilidad en los registros clínicos | Escenarios con datos más estructurados o linealmente separables |

# Conclusions

The MLP and SVM models achieved metrics exceeding 90% in terms of accuracy, precision, and sensitivity, validating their predictive potential to assist in medical decision-making and contribute to better resource allocation in hospitals.

The proposed method incorporates the indicators of NOM-035-STPS-2018 to extend the analysis beyond clinical parameters, considering the psychosocial and organizational dimensions of occupational risk.

From the perspective of the EES (Educational Health Strategy), the results confirm the feasibility of developing ethical, reproducible, and socially oriented artificial intelligence or machine learning systems useful for the healthcare system.

Finally, it is proposed to expand the database and incorporate new psychological, environmental, and occupational variables to strengthen the predictive capacity and interpretability of the model. These actions will allow progress toward a comprehensive, sustainable healthcare system that links AI with human well-being.

# References

**References**

**Basics**
Cisterna-García, A., Guillén-Teruel, A., Caracena, M., Pérez, E., Jiménez, F., Francisco-Verdú, F. J., ... & Botía, J. A. (2022). A predictive model for hospitalization and survival to COVID-19 in a retrospective population-based study. Scientific Reports, 12(1), 18126.
**Support**
Cooper, C. L., Quick, J. C., & Schabracq, M. J. (2015). International Handbook of Work and Health Psychology (3rd ed.). Wiley-Blackwell.
LeCun, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.

**Background.**
Secretaría del Trabajo y Previsión Social (STPS). (2018). NOM-035-STPS-2018: Factores de riesgo psicosocial en el trabajo. Identificación, análisis y prevención. Diario Oficial de la Federación.

**Support**
Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
World Health Organization. (2022). Workplace Health Promotion and Well-being. WHO Publications.
Zhang, J., Liu, S., & Wang, Y. (2008). Gene association study with SVM, MLP and cross-validation for the diagnosis of diseases. Progress in Natural Science, 18(6), 741-750

**Differences.**
Musaa, F. E., & Elnoor, E. M. (2025). A Comparative Study of Machine Learning Algorithms for Lung Cancer Diagnosis: Evaluating Performance Using ROC Curve, Accuracy, Recall, and F1 Score Metrics. General Letters in Mathematics (GLM), 15(2).
Razaque, A., Ben Haj Frej, M., Almi'ani, M., Alotaibi, M., & Alotaibi, B. (2021). Improved support vector machine enabled radial basis function and linear variants for remote sensing image classification. Sensors, 21(13), 4431

**Discussions**
Nwanosike, E. M., Conway, B. R., Merchant, H. A., & Hasan, S. S. (2022). Potential applications and performance of machine learning techniques and algorithms in clinical practice: a systematic review. International journal of medical informatics, 159, 104679.
Secretaría del Trabajo y Previsión Social (STPS). (2018). NOM-035-STPS-2018: Factores de riesgo psicosocial en el trabajo. Identificación, análisis y prevención. Diario Oficial de la Federación.
World Health Organization. (2022). Workplace Health Promotion and Well-being. WHO Publications.
Zhang, J., Liu, S., & Wang, Y. (2008). Gene association study with SVM, MLP and cross-validation for the diagnosis of diseases. Progress in Natural Science, 18(6), 741-750

**MARVID®**