

Clasificación de trastornos hipertensivos durante el embarazo mediante Factor Analysis of Mixed Data y perceptrones multicapa

Classification of hypertensive disorders during pregnancy using mixed data factor analysis and multilayer perceptrons

QUINTANAR-GÓMEZ, Juan†*, ROBLES-CAMARILLO, Daniel, TREJO-MACOTELA, Francisco Rafael y QUINTANAR-GÓMEZ, Samuel

Universidad Politécnica de Pachuca. Carretera Pachuca-Cd. Sahagún Km. 20, Zempoala, Hidalgo, México.

ID 1^{er} Autor: Juan, Quintanar-Gómez / ORC ID: 000-0002-3298-4126, Researcher ID Thomson: W-8838-2019, CVU CONACYT ID: 923315

ID 1^{er} Coautor: Daniel, Robles-Camarillo / ORC ID: 0000-0002-7637-5904, Researcher ID Thomson: T-3496-2018, CVU CONACYT ID: 161666

ID 2^{do} Coautor: Francisco Rafael, Trejo-Macotela / ORC ID: 0000-0003-2133-3456, Researcher ID Thomson: T-3684-2018, CVU CONACYT ID: 37643

ID 3^{er} Coautor: Samuel, Quintanar-Gómez / ORC ID: 0000-0003-4220-6093, Researcher ID Thomson: V-4617-2018, CVU CONACYT ID: 232464

DOI: 10.35429/JP.2020.11.4.1.8

Recibido 19 de Marzo, 2020; Aceptado 29 Junio, 2020

Resumen

La detección temprana de trastornos hipertensivos durante el embarazo (THE) desempeña un papel importante para garantizar un embarazo seguro. Ante ello, es esencial identificar el surgimiento de THE mediante los parámetros clínicos de la paciente. Por tanto, el presente manuscrito describe el proceso de identificación y selección de las variables que presenten una correlación con los trastornos hipertensivos en el embarazo mediante Factor Analysis of Mixed Data (FAMD). Así mismo, proponemos la clasificación de THE empleando un perceptrón multicapa y las variables seleccionadas del análisis FAMD. El rendimiento del modelo propuesto presenta un accuracy de 0.84 para la clasificación de pacientes con THE. Las variables identificadas en esta investigación pueden brindar al personal médico un antecedente respecto a que parámetros pueden sugerir el desarrollo de THE. Así mismo, el modelo propuesto puede ser una herramienta complementaria para la identificación de THE, brindando al personal médico un punto de vista adicional al diagnóstico proporcionado.

FAMD, Perceptrón multicapa, Trastornos hipertensivos

Abstract

Early detection of hypertensive disorders during pregnancy (HDP) has an important part in ensuring a safe pregnancy. Therefore, it is essential to identify the appearance of HDP through the patient's clinical parameters. Therefore, this manuscript describes the process of identifying and selecting variables that correlate with hypertensive disorders in pregnancy through Factor Analysis of Mixed Data (FAMD). Furthermore, we propose the classification of HDP using a multilayer perceptron and the selected variables of FAMD analysis. The performance of the proposed model presented an accuracy of 0.84 for the classification of patients with HDP. The variables identified in this research may provide the medical staff a background about which parameters may suggest the development of HDP. Likewise, the proposed model can be a complementary tool for the identification of HDP, giving the medical staff an additional point of view to the diagnosis provided.

FAMD, Hypertensive diseases, Multilayer perceptron

Citación: QUINTANAR-GÓMEZ, Juan, ROBLES-CAMARILLO, Daniel, TREJO-MACOTELA, Francisco Rafael y QUINTANAR-GÓMEZ, Samuel. Clasificación de trastornos hipertensivos durante el embarazo mediante Factor Analysis of Mixed Data y perceptrones multicapa. Revista de Fisioterapia y Tecnología Médica. 2020. 4-11: 1-8

* Correspondencia del Autor (juanquingom@gmail.com)

† Investigador contribuyendo como primer autor.

Introducción

Los trastornos hipertensivos durante el embarazo (THE) son anomalías caracterizadas por una presión constantemente alta en las paredes sanguíneas, provocando un mayor esfuerzo al corazón durante el proceso de dilatación, generando daños a la mujer y al bebé durante el periodo gestacional. En consecuencia, la falta de tratamiento de los THE contribuye al incremento de nacimientos prematuros, restricción del crecimiento intrauterino, hipoxia, desprendimiento placentario, bajo peso al nacer y, en el peor de los casos, muerte perinatal y mortalidad materna, incrementando este último aspecto a nivel global en los últimos años (Garovic & August, 2013; Harper et al., 2016). Identificando la incidencia de dichos trastornos entre el 5-10% de los embarazos a nivel mundial (de Integración, 2015). Ante ello, la Agenda 2030 para el Desarrollo Sostenible establece en uno de sus objetivos principales la reducción en al menos 70 casos por cada 100,000 nacidos vivos los aspectos relacionados a las razones de defunción materna (Ruelas Valdés, 2017); encontrándose entre dichos aspectos los THE.

La comunidad científica ha propuesto la detección de los THE a través de modelos de clasificación a partir de un conjunto de datos recopilado de casos de estudio. Ejemplo de ello es reportado por Lan et al. (2019) estableciendo un modelo de clasificación de THE basado la fusión de Random Forest y XgBoost, presentando un accuracy de 0.8368, el modelo propuesto fue entrenado a partir de un conjunto de datos de 3000 pacientes en etapa gestacional; con 143 casos con THE durante el embarazo.

Moreira et al. (2018) propone el uso de un modelo neuro-difuso para la predicción del síndrome de hemólisis, aumento enzimas hepáticas y trombocitopenia (HELLP); uno de los THE. El modelo propuesto por Moreira et al. (2018) fue entrenado mediante un conjunto de datos de 205 pacientes en etapa gestacional; identificando en dicho conjunto 7 pacientes con síndrome HELLP. El rendimiento del modelo propuesto destaca un accuracy de 0.685 y un F1-score de 0.705.

Moreira et al. (2016) realiza una clasificación de THE comparando el rendimiento de un J48 Decision Tree y Bayesian Network, entrenando ambos modelos mediante un registro de 25 casos de hipertensión; resultando en un accuracy de 0.518 y 0.517 en el rendimiento del modelo J48 Decision Tree y el clasificador bayesiano respectivamente, proponiendo el uso de otros modelos de clasificación en trabajos futuros.

Dado que las causas que propician el surgimiento de los THE no han sido identificadas, la comunidad médica ha establecido el monitoreo los índices de presión arterial y la identificación de proteinuria como uno de los factores de detección temprana de los THE; no obstante, la identificación de proteinuria en la orina de la paciente gestante ha presentado en diferentes casos de estudio una correlación para la identificación de THE (Thangaratinam et al., 2009; Visintin et al., 2010).

Ante ello, la comunidad científica ha establecido el monitoreo de parámetros adicionales a fin de ser complementados en la identificación de THE, entre los parámetros identificados están: hipertensión crónica previa, edad, semana de gestación, multiparidad, obesidad, antecedentes previos de diabetes, enfermedades renales, entre otras (Brown et al., 2018; Fraser et al., 2010; Visintin et al., 2010). Ejemplo previo de clasificación de THE mediante diferentes atributos es reportado por Moreira et al (2018), utilizando 3 atributos para el entrenamiento del modelo anteriormente mencionado, los cuales son: índices de presión arterial, proteinuria y presencia de edema. Así mismo, Ganzevoort et al. (2006) establece una correlación entre la edad de la paciente embarazada, el peso del bebé y la multiparidad en un estudio de caso realizado a 216 pacientes a fin de identificar parámetros correlacionados con el desarrollo de THE.

Adicionalmente, se ha reportado el uso de algoritmos genéticos (Yoshida et al., 2001), Análisis de Componentes Principales (PCA) (Song et al., 2010), margen de pérdida de la clasificación del vecino más cercano (Li & Lu, 2009) para la selección de características a utilizar en modelos de clasificación.

Dichas técnicas son utilizadas en problemas del mundo real como son: reconocimiento facial (Song et al., 2010), detección de cáncer, detección de enfermedades retinales, diagnóstico de enfermedades (Remeseiro & Bolon-Canedo, 2019), entre otros.

Basado en la premisa anterior, proponemos la selección de características para la identificación de THE mediante Factor Analysis of Mixed Data (FAMD). Así mismo, proponemos un modelo de perceptrón multicapa (MLP) con 3 capas ocultas a fin de proveer a la literatura de un modelo para el diagnóstico de casos de pacientes que se encuentren en riesgo de presentar THE. El presente manuscrito describe la metodología utilizada para interpretación, preparación y selección de características empleando FAMD en un conjunto de datos con características cualitativas y cuantitativas. Así mismo, la metodología utilizada describe el procedimiento utilizado para el entrenamiento y prueba del MLP, identificando en la etapa de evaluación los resultados obtenidos durante la investigación.

Metodología

La metodología utilizada en la investigación es Cross-Industry Standard Process for Data Mining (CRISP-DM) (Wirth, 2000), dividida en 4 etapas: interpretación de datos, preparación de datos, modelado y evaluación; las cuales son descritas a continuación:

Interpretación de los datos

El conjunto de datos fue recopilado a partir de los registros procedentes de mujeres en etapa gestacional del periodo 2018-2019 en el centro de salud "Jesús del Rosal", ubicado en la zona centro de Pachuca, Hidalgo, México. El conjunto de datos utilizado se compone de 102 características relacionadas al expediente clínico de la paciente. El conjunto de datos posee un registro de 350 pacientes en etapa gestacional, con un promedio de 3 registros por paciente. Como se observa en la Tabla 1, es posible visualizar la estadística descriptiva referente a: edad, peso, altura, semana de gestación (SG), presión sistólica (PS), presión diastólica (PD) y pulsaciones por minuto (bpm) del conjunto de datos utilizado. Cabe destacar que el comité de ética del centro clínico aprobó el estudio realizado, obteniendo un consentimiento para el manejo de los datos empleados en esta investigación.

Parámetro	Min	Max	Media	Std
Edad	14	47	24.1	5.89
Peso	29.6	95	62.13	11.69
Altura	117	197	154.79	7.10
SG	1	34	20.4	8.99
PS	60	170	100	11.77
PD	45	110	63.74	8.2
bpm	54	120	76.59	9.79

Tabla 1 Resumen estadístico del conjunto de datos utilizado (Min; valor mínimo, Max; valor máximo, Std; desviación estándar).

Fuente: *Elaboración Propia*

Preparación de datos

El conjunto de datos fue limpiado de acuerdo a los procedimientos establecidos por la literatura (Wirth, 2000), imputando valores faltantes a través de las medidas de tendencia central analizadas en las características del conjunto de datos. Los valores atípicos fueron modificados empleando la Transformada de Sujeción, como se observa en la ecuación (1); donde Q_1 y Q_3 es el primer y tercer rango intercuartílico respectivamente y f_i es un valor específico de la característica f . Las características cualitativas del conjunto de datos fueron binarizadas a fin de ser procesadas por la técnica FAMD. Así mismo, se removieron registros duplicados del conjunto de datos utilizado (McCallum, 2012; Mckinney, 2017).

$$f_i = \begin{cases} Q_1, & \text{si } f_i < Q_1 \\ Q_3, & \text{si } f_i > Q_3 \\ f_i, & \text{en caso contrario} \end{cases} \quad (1)$$

Selección de características

El conjunto de datos fue procesado, en primera instancia, empleando FAMD; análisis empleado para la extracción de características y reducción de dimensionalidad en conjuntos de datos con variables cuantitativas y cualitativas. El funcionamiento del FAMD es descrito a continuación (De Leon & Chough, 2013):

El FAMD divide el conjunto de datos en dos grupos de características. Denotamos la letra Q para representar al subconjunto de características cualitativas q y el subconjunto de datos K para representar las características cuantitativas k .

El FAMD se basa en la reducción de dimensionalidad mediante la combinación del análisis PCA en el subconjunto Q y el Análisis Múltiple de Correspondencia (MCA) en el subconjunto K . En primera instancia, es utilizado el análisis PCA; como se observa en la ecuación (2), buscando una función que maximice el coeficiente de correlación r entre las variables z y k ; siendo la variable z la dimensión a construir en respuesta al análisis PCA, convirtiendo los datos en vectores propios o eigenvectores respecto a la varianza de las características a fin de encontrar los componentes que representen el resultado de PCA.

$$\sum_k \max(r^2(z, k)) \quad (2)$$

Posteriormente, el análisis MCA es efectuado al subconjunto Q ; como se identifica en la ecuación (3), se busca una función que maximice la correlación entre las variables z y q , encontrando los componentes que proporcionan la mayor varianza entre las variables cualitativas.

$$\sum_q \max(\eta^2(z, q)) \quad (3)$$

Finalmente, el análisis FAMD busca la función que presenta mayor relación entre los subconjuntos K y Q , mediante la ecuación (4):

$$\sum_k \max(r^2(z, k)) + \sum_q \max(\eta^2(z, q)) \quad (4)$$

El análisis FAMD fue ejecutado mediante el lenguaje de programación Python, empleando la librería Prince. El análisis FAMD fue implementado a fin de reducir el conjunto de datos a 4 dimensiones, aprovechando el 92% de la varianza explicada del conjunto de datos (Gráfico 1).

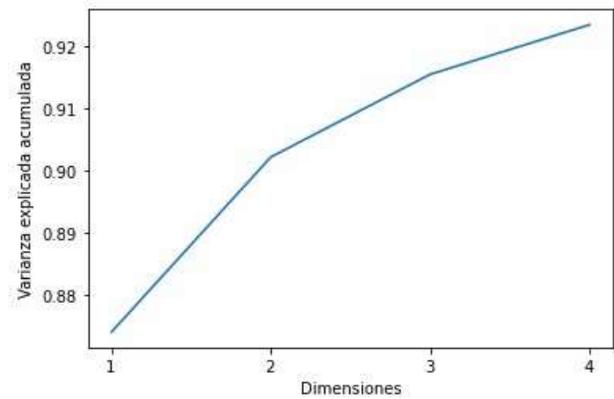


Gráfico 1 Porcentaje de varianza explicada por el número de dimensiones utilizado en el análisis FAMD

Fuente: *Elaboración Propia*

El resultado del análisis FAMD proporciona una matriz con el número de dimensiones menor al proporcionado por el conjunto de datos original. No obstante, establecemos el análisis FAMD en este manuscrito para la selección de características que proporcionan una mayor correlación en las dimensiones de la matriz resultante.

Modelado

A partir de las variables seleccionadas del análisis FAMD fue entrenado un MLP (Yun et al., 2018). Los perceptrones producen una salida basada en varias entradas al formar una combinación lineal utilizando pesos de entrada; como se observa en la ecuación (5), donde w_i denota al vector de pesos, x_i al vector de entradas, b el coeficiente bias y g la función de activación (Gupta, 1999):

$$y = g\left(\sum_{i=1}^n w_i x_i + b\right) \quad (5)$$

Un MLP tiene una estructura integrada por más de un perceptrón, compuesto por una capa de entrada encargada de la recepción de datos, una capa de salida que toma una decisión o predicción sobre los datos de entrada y un número arbitrario de capas ocultas, colocadas entre las capas descritas anteriormente. El aprendizaje del MLP se efectuado mediante backpropagation; al modificar los pesos de las conexiones en capa época; dependiendo del error producido en la salida respecto al resultado esperado, como se observa en la ecuación (6); donde d es el valor objetivo, j el nodo de salida, n el dato de entrenamiento y y el valor producido por el perceptron.

$$e_j^2(n) = d_j(n) - y_j(n) \quad (6)$$

Los pesos de los nodos son ajustados mediante las correcciones que minimicen el error en la salida, mediante la ecuación (7).

$$\varepsilon(n) = \frac{1}{2} \sum_j e_j^2(n) \quad (7)$$

El cambio de pesos en el perceptrón es realizado mediante el algoritmo de gradiente descendiente, como se observa en la ecuación (8), donde η es la tasa de aprendizaje y y_i es la salida de la neurona anterior.

$$\Delta w_{ji}(n) = -\eta \frac{\partial \varepsilon(n)}{\partial v_j(n)} y_i(n) \quad (8)$$

El MLP propuesto presenta una estructura de 3 capas internas con función de activación sigmoide en la segunda capa interna y en la capa de salida, mientras que la función de activación Rectified Linear Unit (ReLU) fue utilizada en la primer y tercer capa interna (Tabla 2).

La técnica k-fold cross-validation fue utilizada para dividir el conjunto de datos en un subconjunto de datos de entrenamiento y prueba, empleando el 80% del conjunto de datos para el entrenamiento y el 20% restante para la etapa de prueba (Mckinney, 2017). El modelo fue entrenado en el lenguaje de programación Python empleando 1500 épocas. Las especificaciones del equipo de cómputo utilizado son: procesador AMD Ryzen 1700, disco sólido de 1 Tb y capacidad de 16 Gb de memoria RAM. Se empleó la función de pérdida binary-crossentropy y la métrica accuracy fue utilizada para medir el rendimiento del modelo.

Capa del perceptrón	Número de neuronas	Función de activación
Capa de entrada	10	Lineal
Capa interna 1	10	ReLU
Capa interna 2	20	Sigmoide
Capa interna 3	10	ReLU
Capa de Salida	1	Sigmoide

Tabla 2 Estructura del MLP propuesto
Fuente: Elaboración Propia

Resultados y Discusión

El análisis FAMD propuesto para la selección de características a utilizar en el entrenamiento del MLP refleja la identificación de 9 variables que presentaron mayor correlación en la creación de las dimensiones del análisis FAMD (Tabla 3). En este caso la presión arterial, edad, peso (Leung et al., 2017), frecuencia cardíaca (Romundstad et al., 2010) y respiratoria de la paciente son parámetros establecidos por la literatura para la detección de THE. No obstante, el análisis FAMD refleja una correlación de la temperatura corporal; parámetro que ha presentado una correlación en embarazos prematuros y bajo peso al nacer (Rashid et al., 2017), la presencia de la vacuna contra la difteria y tétanos (DT) (Pineda, 2009) y la presencia de antecedentes previos de embarazo en la creación de las dimensiones del análisis FAMD. Por tanto, las variables mencionadas anteriormente fueron utilizadas en el MLP para la clasificación de THE.

Parámetro seleccionado	Dimensión			
	1	2	3	4
Peso	0.17	-0.24	0.00	-0.28
PD	0.04	-0.05	0.79	-0.10
PS	0.11	-0.06	0.07	-0.14
Frecuencia cardíaca	-0.19	-0.07	0.16	-0.05
Frecuencia respiratoria	-0.15	-0.11	0.03	0.07
Temperatura	-0.01	0.11	-0.027	0.05
Vacuna DT	-0.01	-0.47	-0.12	-0.20
Edad	0.064	-0.26	0.19	-0.48
Embarazo anterior	0.16	-0.43	0.23	-0.41

Tabla 3 Correlación de las variables seleccionadas en las dimensiones resultantes del análisis FAMD
Fuente: Elaboración Propia

Evaluación

La evaluación del modelo propuesto emplea curvas de precisión-recall (PR), evaluando la precisión contra el recall del modelo (Gráfico 2). Dicho aspecto es reflejado en la Tabla 4, identificando un accuracy de 0.84 durante la etapa de prueba del modelo.

	Precisión	Recall	F1-score	Support
0	0.89	0.92	0.91	178
1	0.42	0.33	0.37	30
Accuracy			0.84	208
Macro avg.	0.65	0.63	0.64	208
Weighted avg.	0.82	0.84	0.83	208

Tabla 4 Rendimiento del modelo propuesto
Fuente: Elaboración Propia

El modelo MLP propuesto, al ser comparado con la información procedente de la literatura, presenta un accuracy similar al reportado por Lan et al. (2019); como se observa en la Tabla 5, reportando un accuracy de 0.8368 por medio de la combinación de modelos de Random Forest y XgBoost. No obstante, el accuracy obtenido presenta una mejora respecto a los modelos reportados por la literatura.

Modelo	Accuracy
MLP	0.84
Random Forest y XgBoost, Lan et al. (2019)	0.8368
Fuzzy model, Moreira et al. (2018)	0.685
J48 Decision Tree, Moreira et al. (2016)	0.518
Bayesian Network, Moreira et al. (2016)	0.517

Tabla 5 Correlación de las variables seleccionadas en las dimensiones resultantes del análisis FAMD

Fuente: Elaboración Propia

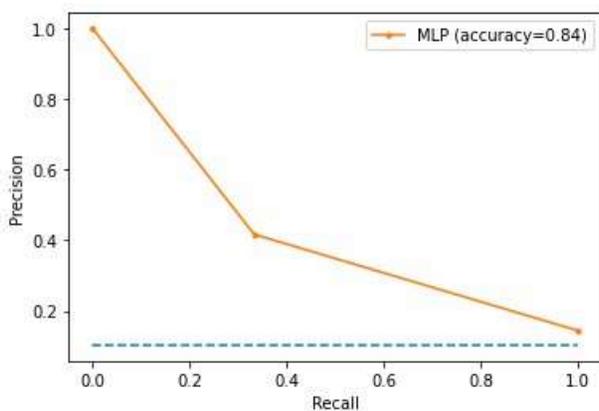


Gráfico 2 Curva PR para clasificador THE.

Fuente: Elaboración Propia

Agradecimientos

Los autores agradecen al Consejo Nacional de Ciencia y Tecnología de México (CONACYT) por financiar la investigación de este trabajo. Un agradecimiento especial al personal docente de la Universidad Politécnica de Pachuca por los conocimientos y habilidades proporcionados y al Centro de Salud "Jesús del Rosal" por la cooperación para el desarrollo de la presente investigación.

Conclusiones

La detección de trastornos hipertensivos durante el embarazo es un aspecto esencial para la reducción de Razones de Mortalidad Materna, ante ello se han propuesto diferentes modelos para la clasificación de THE a partir de diferentes parámetros clínicos recopilados. Basado en lo anterior, un modelo de perceptrón multicapa fue propuesto para la clasificación de trastornos hipertensivos empleando las características que presentaron mayor correlación durante el análisis FAMD. En este caso fueron seleccionadas variables que presentan en la literatura una correlación en el desarrollo de los trastornos THE; no obstante, se identificaron variables que proporcionan correlación con aspectos relacionados al peso del bebé y embarazos prematuros.

El rendimiento del modelo propuesto presenta un accuracy del 0.84 para la predicción de THE. El modelo propuesto puede ser integrado en un dispositivo o servicio web a fin de proveer un segundo punto de vista en el diagnóstico de pacientes en etapa gestacional. Sin embargo, en trabajos futuros puede realizarse un análisis comparativo del rendimiento de modelos a fin de identificar una mejora en el rendimiento propuesto. Así mismo, en trabajos futuros es menester la identificación y selección de características mediante análisis de un conjunto de datos mixto a fin de identificar características adicionales que proporcionen mejoras en la detección de THE.

Referencias

Brown, M. A., Magee, L. A., Kenny, L. C., Karumanchi, S. A., McCarthy, F. P., Saito, S., Hall, D. R., Warren, C. E., Adoyi, G., & Ishaku, S. (2018). Hypertensive disorders of pregnancy: ISSHP classification, diagnosis, and management recommendations for international practice. *Hypertension*, 72(1), 24–43.

de Integración, S. (2015). Informe sobre la salud de los mexicanos 2015. *Diagnostico General de La Salud Poblacional. 1ª Edición. Secretaria de Salud: México DF.*

De Leon, A. R., & Chough, K. C. (2013). *Analysis of mixed data: methods & applications*. CRC Press.

- Fraser, A., Tilling, K., Macdonald-Wallis, C., Sattar, N., Brion, M.-J., Benfield, L., Ness, A., Deanfield, J., Hingorani, A., & Nelson, S. M. (2010). Association of maternal weight gain in pregnancy with offspring obesity and metabolic and vascular traits in childhood. *Circulation*, *121*(23), 2557.
- Ganzevoort, W., Rep, A., de Vries, J. I. P., Bonsel, G. J., & Wolf, H. (2006). Prediction of maternal complications and adverse infant outcome at admission for temporizing management of early-onset severe hypertensive disorders of pregnancy. *American Journal of Obstetrics and Gynecology*, *195*(2), 495–503.
- Garovic, V. D., & August, P. (2013). Preeclampsia and the future risk of hypertension: the pregnant evidence. *Current Hypertension Reports*, *15*(2), 114–121.
- Gupta, M. M. (1999). *Soft computing and intelligent systems: theory and applications*. Elsevier.
- Harper, L. M., Biggio, J. R., Anderson, S., & Tita, A. T. N. (2016). Gestational age of delivery in pregnancies complicated by chronic hypertension. *Obstetrics and Gynecology*, *127*(6), 1101.
- Lan, X., Wu, W., Peng, D., Xu, T., Wang, J., Wang, G., & Hou, F. (2019). Classification of hypertension in pregnancy based on random forest and Xgboost fusion model. *The Third International Conference on Biological Information and Biomedical Engineering*, 1–5.
- Leung, A. A., Daskalopoulou, S. S., Dasgupta, K., McBrien, K., Butalia, S., Zarnke, K. B., Nerenberg, K., Harris, K. C., Nakhla, M., & Cloutier, L. (2017). Hypertension Canada's 2017 guidelines for diagnosis, risk assessment, prevention, and treatment of hypertension in adults. *Canadian Journal of Cardiology*, *33*(5), 557–576.
- Li, Y., & Lu, B.-L. (2009). Feature selection based on loss-margin of nearest neighbor classification. *Pattern Recognition*, *42*, 1914–1921.
<https://doi.org/10.1016/j.patcog.2008.10.011>
- McCallum, Q. E. (2012). *Bad data handbook: cleaning up the data so you can get back to work*. “O'Reilly Media, Inc.”
- Mckinney, W. (2017). *Python for Data Analysis - 2nd edition*.
<https://doi.org/10.1145/1985441.1985476>
- Moreira, Mário W L, Rodrigues, J. J. P. C., Kumar, N., Al-Muhtadi, J., & Korotaev, V. (2018). Nature-inspired algorithm for training multilayer perceptron networks in e-health environments for high-risk pregnancy care. *Journal of Medical Systems*, *42*(3), 51.
- Moreira, Mario W L, Rodrigues, J. J. P. C., Oliveira, A. M. B., Saleem, K., & Neto, A. (2016). Performance evaluation of predictive classifiers for pregnancy care. *2016 IEEE Global Communications Conference (GLOBECOM)*, 1–6.
- Pineda, V. (2009). Vacunación en la embarazada. *Anales de Pediatría Continuada*, *7*(3), 173–176. [https://doi.org/10.1016/S1696-2818\(09\)71123-0](https://doi.org/10.1016/S1696-2818(09)71123-0)
- Rashid, H., Kagami, M., Ferdous, F., Ma, E., Terao, T., Hayashi, T., & Wagatsuma, Y. (2017). Temperature during pregnancy influences the fetal growth and birth size. *Tropical Medicine and Health*, *45*(1), 1.
- Remeseiro, B., & Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. *Computers in Biology and Medicine*, *112*(February), 103375. <https://doi.org/10.1016/j.combiomed.2019.103375>
- Romundstad, P. R., Magnussen, E. B., Smith, G. D., & Vatten, L. J. (2010). Hypertension in pregnancy and later cardiovascular risk: common antecedents? *Circulation*, *122*(6), 579–584.
- Ruelas Valdés, D. M. (2017). *Objetivo 3 de Desarrollo Sostenible: Salud y Bienestar, garantizar una vida sana y promover el bienestar para todos en todas las edades*.
- Song, F., Guo, Z., & Mei, D. (2010). *Feature selection using principal component analysis*. <https://doi.org/10.1109/ICSEM.2010.14>
- Thangaratinam, S., Coomarasamy, A., O'Mahony, F., Sharp, S., Zamora, J., Khan, K. S., & Ismail, K. M. K. (2009). Estimation of proteinuria as a predictor of complications of pre-eclampsia: a systematic review. *BMC Medicine*, *7*(1), 1–9.

Visintin, C., Mugglestone, M. A., Almerie, M. Q., Nherera, L. M., James, D., & Walkinshaw, S. (2010). Management of hypertensive disorders during pregnancy: summary of NICE guidance. *Bmj*, *341*, c2207.

Wirth, R. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, *24959*, 29–39. <https://doi.org/10.1.1.198.5133>

Yoshida, H., Leardi, R., Funatsu, K., & Varmuza, K. (2001). Feature selection by genetic algorithms for mass spectral classifiers. *Analytica Chimica Acta*, *446*(1–2), 483–492.

Yun, K., Huyen, A., & Lu, T. (2018). Deep neural networks for pattern recognition. *Advances in Pattern Recognition Research*, 49–79.