

# Evaluación de cuatro clasificadores para el reconocimiento de síndrome de muerte súbita del lactante utilizando ácidos grasos de cadena corta como fuente de información

Karen E. Villagrana-Bañuelos<sup>1</sup>, Laura A. Zanella-Calzada<sup>2</sup>,  
Carlos E. Galván-Tejada<sup>1</sup>, Hamurabi Gamboa-Rosales<sup>1</sup>,  
Jorge I. Galván-Tejada<sup>1</sup>

<sup>1</sup> Universidad Autónoma de Zacatecas,  
Unidad Académica de Ingeniería Eléctrica,  
México

<sup>2</sup> LORIA (INRIA, CNRS, Université de Lorraine),  
France

{kvillagrana;ericgalvan;hamurabigr;gatejo}@uaz.edu.mx,  
laura.zanella-calzada@univ-lorraine.fr

**Resumen.** La inteligencia artificial y su aplicación en el área médica, ha venido a coadyuvar al diagnóstico de diversas enfermedades. La manipulación, procesamiento, análisis, e interpretación de grandes y complejos datos, ha permitido, identificar relaciones, que en muchas ocasiones, permanecen imperceptibles para la mente humana. El desarrollo del aprendizaje automático, permite, diseñar modelos, que ayuden a predecir enfermedades; algunas de ellas, se desconoce aún con los avances científicos, la razón que las genera, por lo que encontrar patrones, permite inferir sobre dichas enfermedades, para prevención y diagnóstico. El síndrome de muerte súbita del lactante (SMSL), es la muerte de un menor de un año de edad, sin causa aparente. Una de las posibles explicaciones al SMSL son los errores innatos del metabolismo (EIM), especialmente los defectos en la beta-oxidación de los ácidos grasos, por lo que este estudio se centra en analizar la relación de los ácidos grasos de cadena corta (AGCC) con casos identificados de muerte por SMSL contra controles, evaluando y comparando el rendimiento de los clasificadores k Nearest Neighbors, Random Forest, Naïve Bayes y Regresión Logística, para incrementar las pruebas de que existe una relación de los procesos metabólicos con este síndrome. Los modelos de Random Forest y de Naïves Bayes obtienen un desempeño igual para esta muestra de sujetos de SMSL, con un área bajo la curva de 0.75 y una exactitud de 0.8889 para ambos clasificadores.

**Palabras clave:** Síndrome de muerte súbita del lactante, ácidos grasos de cadena corta, k vecinos más cercanos, bosque aleatorio, Naïve Bayes, regresión logística.

## Evaluation of Four Classifiers for the Recognition of Sudden Infant Death Syndrome Using Short Chain Fatty Acids as a Source of Information

**Abstract.** Artificial intelligence and its application in the medical area, has come to contribute to the diagnosis of various diseases. The manipulation, processing, analysis, and interpretation of large and complex data has allowed us to identify relationships that, on many occasions, remain imperceptible to the human mind. The development of machine learning allows to design models that help to predict diseases; some of them, even with scientific advances, the reason that generates them is unknown, so finding patterns allows us to infer about these diseases, for prevention and diagnosis. Sudden infant death syndrome (SIDS) is the death of a child under one year of age, with no apparent cause. One of the possible explanations for SIDS, is the inborn errors of metabolism (IEM), especially defects in fatty acid beta-oxidation, so this study focuses on analyzing the relationship of short chain fatty acids (SCFAs) in cases identified from death by SIDS versus control cases, evaluating and comparing the performance of k Nearest Neighbors (kNN), Random Forest (RF), Naïve Bayes and Logical Regression to increase the evidence that there is a relationship of metabolic processes with this syndrome. The Random Forest and Naïves Bayes models obtained the same performance for this sample of SIDS subjects, with an area under the curve of 0.75 and an accuracy of 0.8889 for both classifiers.

**Keywords:**Sudden infant death syndrome, short chain fatty acids, k nearest neighbors, random forest, Naïve Bayes, logical regression.

### 1. Introducción

Por décadas la inteligencia artificial, ha existido, pero su relación con la medicina se vio retrasada debido a cuestiones éticas y culturales [18]. A pesar de ello el aprendizaje automático, por medio de los modelos predictivos, ha cobrado gran relevancia en las ciencias médicas, estudiando diferentes enfermedades multifactoriales [1, 10, 23], especialmente en el área pediátrica [16], donde ha permitido incursionar en la asistencia diagnóstica tal como el trabajo de Homer Warnerm, que en 1961 desarrolló un sistema de diagnóstico automatizado para diagnosticar correctamente a pacientes con cardiopatías congénitas a partir de cincuenta variables clínicas, obtuvo como resultado, que el sistema fue más certero, comparado con los diagnósticos de los cardiólogos [18], otro ejemplo más reciente es el trabajo de Halabi et al [13], quienes trabajaron con técnicas de aprendizaje automático en la identificación de la edad ósea pediátrica; y así múltiples investigaciones han consolidado la pertenencia del uso de técnicas de aprendizaje automático en el área médica [24], buscando proporcionar un instrumento de apoyo.

Existen múltiples enfermedades que en nuestro siglo, siguen siendo un reto para la comunidad científica, ya que no se ha logrado esclarecer consistentemente

el origen de las mismas. Resultan de interés, ya que la inteligencia artificial permite manipular grandes y complejas cantidades de datos, encontrando relaciones que a simple vista, parecen inexistentes. Así patologías, de las cuales aún no se ha descubierto la causa certera, podrían beneficiarse, de la inteligencia artificial, para lograr identificar patrones, que no se tenían contemplados, y así inferir con el objetivo de prevenirlas o tratarlas.

El síndrome de muerte súbita del lactante (SMSL) es definido como la muerte de un menor de un año de edad, que permanece sin explicación después de una autopsia y de una investigación exhaustiva de la escena de muerte [14]. A pesar de los múltiples esfuerzos de investigaciones en área médica, las explicaciones posibles abarcan una amplia rama de teorías sobre el origen de éste síndrome, dentro de las cuales destacan los defectos en el corazón [7, 19, 26, 27], teorías del triple riesgo [11, 15], pérdida del automatismo cardiorrespiratorio [2, 3], los errores innatos del metabolismo [17], entre otros.

Atendiendo, a la teoría patogénica que los errores innatos del metabolismo (EIM) pudieran explicar las muertes por SMSL; entenderemos por EIM a un grupo de enfermedades causadas por una mutación genética que tiene como efecto la producción de una proteína anómala que conlleva la alteración del funcionamiento fisiológico de la célula; según sea la función alterada, los efectos fisiopatológicos del acumulo de sustancias no metabolizadas dependen del grado de acumulación y de su posible toxicidad [6], de ahí la gran variedad de presentaciones clínicas, y la dificultad en el diagnóstico si no se busca de manera intencionada. Los trastornos del ciclo de beta-oxidación de ácidos grasos, forman parte de los EIM. La enzima acetil coenzima A, es generada a partir de ciclos repetidos de oxidación, necesitando de grupos de cuatro enzimas específicas para cada una de las longitudes de cadena de los ácidos de cadena muy larga, larga, media y corta. Se han estudiado los defectos de oxidación de los ácidos de cadena media, sin embargo, Opedal et al [20] destaca la necesidad de investigar otros tipos de defectos como los EIM, en especial los defectos del metabolismo de los ácidos grasos de cadena corta (AGCC) [5]. Siendo de especial interés identificar, si existe relación entre el SMSL y los niveles de ácidos grasos de cadena corta, lo cual permitiría tener marcadores para pronosticar riesgo de presentar el síndrome.

Debido a lo anterior, el objetivo de este estudio es evaluar y comparar el rendimiento de los clasificadores k Nearest Neighbors (kNN), Random Forest(RF), Naïve Bayes(NB) y Regresión Lógica(RL) en el reconocimiento de SMSL utilizando AGCC.

## **2. Materiales y métodos**

En la presente sección se describen los datos utilizados para el desarrollo de los modelos de clasificación, al igual que una breve descripción de los mismos. También, las métricas utilizadas para la validación y la comparación de los modelos son mencionadas.

## 2.1. Descripción de la base de datos

Para el desarrollo de este trabajo fue utilizada la base de datos, 'SCFA profile in babies dying from SIDS', la cual se encuentra públicamente disponible en <https://www.metabolomicsworkbench.org> proporcionada por el *NIH Common Fund's National Metabolomics Data Repository*, con el número de proyecto asignado PR000512.

La información contenida en esta base de datos se refiere a 18 sujetos, cinco controles y 13 pacientes diagnosticados con SMSL. Esta información incluye los valores obtenidos para las características presentadas en la Tabla 1, los cuales fueron calculados post-mortem a través de la extracción fría medida con ionización de electrones - cromatografía de gases espectrometría de masas sin derivación.

**Tabla 1.** Características contenidas en la base de datos de perfil de AGCC, en lactantes fallecidos por SMSL.

Características
1 Horas de intervalo post-mortem
2 Semanas de edad gestacional
3 Semanas de edad post-natal
4 Ácido isovalérico
5 Ácido octanoico
6 Ácido propiónico
7 Ácido isobutírico
8 Ácido butírico
9 Ácido hexanoico
10 Ácido valérico
11 Ácido acético

## 2.2. Modelos de clasificación

En esta sección se describen brevemente los modelos de clasificación que han sido utilizados para la identificación de lactantes que presentaron síndrome de muerte súbita. Estos clasificadores fueron seleccionados debido a sus diferentes características, así como por ser utilizados en el campo de la biomédica para diferentes enfermedades, ya sea para dar pronósticos o diagnóstico de las mismas [9, 24, 28].

### Random Forest

Random forest (RF) o bosques aleatorios, se refiere a un conjunto de clasificadores que aprenden de manera aleatoria múltiples árboles de decisión. Este método consiste en dos etapas principales, entrenamiento y prueba. En la etapa de entrenamiento se construyen los múltiples árboles de decisión, mientras que

en la etapa de prueba se lleva a cabo la clasificación o la predicción de la variable de salida con base en las variables de entrada.

La distribución que se obtiene de los árboles de decisión de cada bosque es inicialmente promediada para después llevar a cabo la clasificación. Los predictores de cada muestra son combinados usando la media para el caso donde las variables son continuas y el voto mayoritario para el caso donde las variables son categóricas.

Una característica de esta técnica es que no presenta problemas significativos de sobreajuste, y no se ve afectada por el ruido o los valores atípicos que pueden estar presentes en los datos. Además, puede reducir los errores de generalización, permitiendo lograr un valor alto de precisión [4].

### **Regresión Logística**

Los modelos de regresión logística (RL) presentan como requisito que la variable de salida sea categórica, y calculan generalmente la probabilidad, en un rango de 0 a 1, de que una clase pertenezca a alguna de las categorías del conjunto de datos.

La probabilidad estimada que se obtiene como salida es usada para inferir qué tan confiable puede ser el valor que se predijo en comparación con el valor real, dado un valor de entrada.

Existen diferentes tipos de RL, tales como regresión logística binaria, multinomial y ordinal, las cuales varían dependiendo del número de categorías que presente la variable de salida [8].

### ***K*-Nearest Neighbor**

*K*-NN o *K*-vecinos cercanos, es una técnica regularmente utilizada cuando se tiene poco o nulo conocimiento sobre la distribución de los datos.

*K*-NN consiste en minimizar la distancia Euclidiana de la muestras de entrada con respecto a las características, buscando de esa manera que cada muestra se categorice con sus vecino más cercanos.

Por lo tanto, esta técnica se basa en una regla de clasificación en donde se asigna a cada muestra la etiqueta de la categoría mayoritaria de sus *K* muestras más cercanas. Regularmente *K* es asignado con un número impar, para evitar de esta manera lazos [21].

### **Naïve Bayes**

Naïve Bayes (NB) se refiere a una técnica de clasificación basada en probabilidad, en donde se calcula el conjunto de probabilidades a través de contar la frecuencia y las combinaciones de valores en un conjunto de datos.

Esta técnica asume que todas las variables son independientes considerando el valor de la variable dependiente o de salida. Esta suposición de independencia condicional es raramente aceptada en aplicaciones reales, por lo que se le categoriza como 'ingenua', pero este algoritmo tiende a aprender rápidamente en problemas de clasificación controlados.

Por lo tanto, el teorema Bayes es usado para determinar la probabilidad condicional, en donde se calcula la probabilidad de la ocurrencia de un evento A, de un evento B, de un evento A cuando un evento B ocurre, y de un evento B cuando un evento A ocurre. Con base en esto, se obtiene la probabilidad de cada muestra de corresponder a alguna de las dos categorías posibles [25].

### 2.3. Validación

Con el fin de evaluar el comportamiento de los modelos desarrollados se llevó a cabo una etapa de validación, con base en tres métricas principales, la curva ROC (acrónimo de receiver operating characteristic), al área bajo la curva (AUC, por sus siglas en inglés) y la exactitud.

Para el cálculo de estas métricas es necesario conocer el número de verdaderos positivos (VP), verdaderos negativos (VN), falsos positivos (FP) y falsos negativos (FN). Con base en esto, es posible obtener la sensibilidad y la especificidad. La sensibilidad calcula con la Ecuación 1, la cual permite conocer la probabilidad de que un caso de positivo SMSL sea clasificado correctamente:

$$\text{Sensibilidad} = \frac{VP}{VP + FN}. \quad (1)$$

La especificidad se calcula con la Ecuación 2, la cual permite conocer la probabilidad de que un caso negativo de SMSL sea clasificado correctamente:

$$\text{Especificidad} = \frac{VN}{VN + FP}. \quad (2)$$

A través de estas dos métricas es posible obtener la curva ROC, la cual permite visualizar el comportamiento de los clasificadores, en conjunto con el AUC. El AUC representa la probabilidad que una muestra positiva aleatoria sea identificada de manera correcta [12].

Por otro lado, la medición de la exactitud se refiere a la tasa del número de predicciones correctas con respecto al total de predicciones hechas, y se puede calcular con la Ecuación 3:

$$\text{Exactitud} = \frac{\text{predicciones correctas}}{\text{total de predicciones}}. \quad (3)$$

## 3. Experimentos y resultados

En esta sección se describen los experimentos y los resultados obtenidos a través de la metodología seguida.

Para este trabajo fue utilizado un conjunto de datos contenido por once características que describen los valores presentados por pacientes con SMSL positivo y SMSL negativo. Estos datos fueron modelados a través de cuatro clasificadores: RF, RL,  $K$ -NN y NB, con el fin de identificar automáticamente la clase a la que pertenecía cada paciente. Para el entrenamiento de los modelos se

tomó el 70 % de los datos y el 30 % restante se utilizó para prueba. Una vez que los modelos fueron desarrollados, se sometieron a una evaluación a través de una validación estadística, para poder conocer el grado de generalización alcanzado.

En la Tabla 2 se presentan los resultados obtenidos de la validación de los modelos, en donde se pueden observar los valores de exactitud y AUC calculados para cada clasificador.

**Tabla 2.** Resultados de la validación de los modelos obtenidos.

<b>Clasificador</b>	<b>Exactitud</b>	<b>AUC</b>
RF	0.8889	0.7500
RL	0.6667	0.6071
K-NN	0.7777	0.6786
NB	0.8889	0.7500

Por otro lado, en la Figura 1 se muestran las curvas ROC obtenidas para cada modelo de clasificación, donde se puede observar además su comportamiento en términos de sensibilidad y especificidad, permitiendo conocer también el desempeño de cada clasificador para reconocer los verdaderos positivos y verdaderos negativos, siendo dos métricas comúnmente utilizadas para evaluar pruebas médicas.

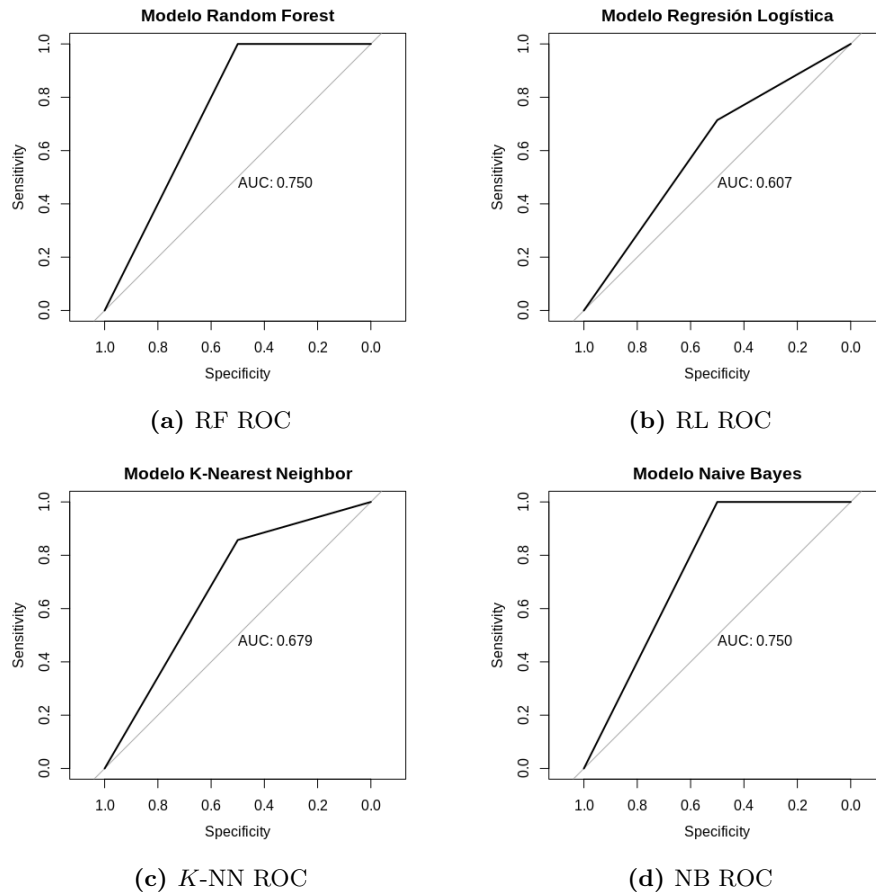
Toda la experimentación fue desarrollada en R [22], un entorno de software libre para gráficos y computación estadística.

#### 4. Discusión y conclusiones

Se realizó un estudio comparativo sobre distintos clasificadores reportados en la literatura para la clasificación de diferentes problemas en el área biomédica, tomando en cuenta el perfil de AGCC para clasificar los casos de muerte por SMSL e incrementar las pruebas de que existe una relación de los procesos metabólicos con este síndrome.

Los clasificadores que se tomaron en cuenta fueron: Regresión Logística (RL), K-Nearest Neighbor (kNN), Random Forest (RF) y Naïve Bayes (NB). En la Tabla 2, se muestran los diferentes clasificadores, con su valor respectivo de exactitud y de área bajo la curva. De esta comparación se puede concluir los siguientes puntos:

- Independientemente del modelo de clasificación seleccionado, todos tiene la capacidad de tener un área bajo la curva mayor a 0.5, lo cual es un factor a tomar en cuenta en un tamizaje de AGCC.
- El desempeño del modelo de regresión logística es menor que el resto en ambas métricas, sin embargo es un modelo de bajo costo computacional que permite evaluar el comportamiento de los algoritmos de clasificación para este problema, así como ser un algoritmo que de manera clásica se utiliza en el área de medicina.



**Fig. 1.** Curvas ROC obtenidas para los distintos modelos de clasificación.

- Los modelos de Random Forest y de Naïves Bayes obtienen un desempeño igual para esta muestra de sujetos de SMSL, con un área bajo la curva de 0.75 y una exactitud de 0.8889 para ambos clasificadores. Ambos con una sensibilidad de 1, pudiendo así clasificar correctamente a todos los positivos.

Es importante mencionar que la base de datos utilizada para realizar este análisis es limitada a 18 sujetos, y es un conjunto desbalanceado ya que 5 son sujetos de control y 13 casos diagnosticados con SMSL, lo cual dificulta la generalización del comportamiento de los clasificadores, así como el hecho de que RF y NB tengan el mejor desempeño, ya que debido a la naturaleza de ambos clasificadores pueden estar sobre ajustados a la clase de casos y por ello tener una sensibilidad de 1.



## 5. Trabajo futuro

Debido a que el SMSL es un proceso multifactorial el cual incluye procesos metabólicos de los ácidos grasos, se propone como trabajo a futuro, realizar pruebas utilizando técnicas de reducción de características para identificar los ácidos grasos de cadena corta que tienen mayor influencia en este proceso, así como la revisión del desempeño de algoritmos de mayor costo computacional. De manera específica quedando los siguientes puntos:

- Selección de características utilizando métodos stepwise.
- Análisis univariado de los ácidos grasos de cadena corta para observar su comportamiento relacionado con el SMSL.
- Implementación de algoritmo genético para observar la interacción de diferentes combinaciones de ácidos grasos así como la importancia de los mismos.
- Implementar técnicas complejas de aprendizaje por computadora, con la limitante del número de sujetos, que es pequeña.

## Referencias

1. Alcalá-Rmz, V., Zanella-Calzada, L.A., Galván-Tejada, C.E., García-Hernández, A., Cruz, M., Valladares-Salgado, A., Galván-Tejada, J.I., Gamboa-Rosales, H.: Identification of diabetic patients through clinical and para-clinical features in Mexico: an approach using deep neural networks. *International journal of environmental research and public health* 16(3), 381 (2019)
2. Bright, F., Vink, R., Byard, R., Duncan, J., Krous, H., Paterson, D.: Abnormalities in substance p neurokinin-1 receptor binding in key brainstem nuclei in sudden infant death syndrome related to prematurity and sex. *PLoS ONE* 12(9), 1–14 (2017)
3. Buchanan, G.: Impaired CO<sub>2</sub>-induced arousal in SIDS and SUDAP. *Trends in Neurosciences* 42(4), 242–250 (2019)
4. Byeon, H.: Is the random forest algorithm suitable for predicting Parkinson's disease with mild cognitive impairment out of Parkinson's disease with normal cognition? *International Journal of Environmental Research and Public Health* 17(7), 2594 (2020)
5. Chávez-Ocaña, S., Bravata-Alcantara, J., Sierra-Martínez, M.: Errores innatos del metabolismo, una mirada a un tópico poco valorado. *Rev Hosp Jua Mex* 85(3), 159–167 (2018), [ ]
6. Chávez-Ocaña, S., Bravata-Alcántara, J., Sierra-Martínez, M.: Errores innatos del metabolismo, una mirada a un tópico poco valorado. *Rev Hosp Jua Mex* 85(3), 159–167 (2018)
7. Courts, C., Madea, B.: Genetics of the sudden infant death syndrome. *Forensic Science International* 203(1), 25–33 (2010)
8. Dreiseitl, S., Ohno-Machado, L.: Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics* 35(5-6), 352–359 (2002)
9. Galván-Tejada, C.E., Zanella-Calzada, L.A., Galván-Tejada, J.I., Celaya-Padilla, J.M., Gamboa-Rosales, H., Garza-Veloz, I., Martínez-Fierro, M.L.: Multivariate feature selection of image descriptors data for breast cancer with computer-assisted diagnosis. *Diagnostics* 7(1), 9 (2017)

10. Galván Tejada, J.I., Galván Tejada, C.E., López Monteagudo, F.E., Alonso González, O., Moreno Báez, A., Celaya Padilla, J.M., Zanella Calzada, L.A., et al.: Image registration measures and chronic osteoarthritis knee pain prediction: Data from the osteoarthritis initiative. *Revista mexicana de ingeniería biomédica* (2019)
11. Guntheroth, W., Spiers, P.: The triple risk hypotheses in sudden infant death syndrome. *Pediatrics* 110(5), 1–6 (2002)
12. Hajian-Tilaki, K.: Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation. *Caspian journal of internal medicine* 4(2), 627 (2013)
13. Halabi, S.S., Prevedello, L.M., Kalpathy-Cramer, J., Mamonov, A.B., Bilbily, A., Cicero, M., Pan, I., Araújo Pereira, L., Teixeira Sousa, R.: The rsna pediatric bone age machine learning challenge. *Radiology* 290(2), 498–503 (2018)
14. Haynes, R.: Biomarkers of sudden infant death syndrome (sids) risk and sids death. In: Duncan, J., Byard, R. (eds.) *SIDS Sudden Infant and Early Childhood Death: The Past, the Present and the Future*. Adelaide (AU), University of Adelaide Press (2018)
15. Kinney, H., Thach, B.: The sudden infant death syndrome. *N Engl J Med* 361(8), 795–805 (2009)
16. Kokol, P.; Završnik, J., Blažun Vošner, H.: Artificial intelligence and pediatrics: A synthetic mini review. *Pediatr Dimensions* 2(4), 1–5 (2017)
17. Labayru Echeverría, M.: Libro blanco de la muerte súbita infantil, chap. Errores innatos del metabolismo. *Patología neurológica y muerte súbita del lactante*, pp. 90–102. No. 2003 in NA, Ediciones Ergon, Madrid, España. Grupo de Trabajo de Muerte Súbita Infantil-AEP, 2 edn. (2013)
18. Lugo-Reyes, S., Maldonado-Colín, G., C., M.: Inteligencia artificial para asistir el diagnóstico clínico en medicina. *Revista Alergia México* 61, 110–120 (2014)
19. Nikolaos, S., Ioakeimidis, T., Meditskou, S., Iakovidou-Kritsi, Z.: Sudden infant death syndrome due to long qt syndrome: a brief review of the genetic substrate and prevalence. *Journal of Biological Research-Thessaloniki* 24(6), 1–6 (2017)
20. Opdal, S., Rognum, T.: The sudden infant death syndrome gene: does it exist? *Pediatrics* 114(4), e506–12 (2004)
21. Peterson, L.E.: K-nearest neighbor. *Scholarpedia* 4(2), 1883 (2009)
22. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2020), <https://www.R-project.org/>
23. Rodríguez-Ruiz, J.G., Galván-Tejada, C.E., Zanella-Calzada, L.A., Celaya-Padilla, J.M., Galván-Tejada, J.I., Gamboa-Rosales, H., Luna-García, H., Magallanes-Quintanar, R., Soto-Murillo, M.A.: Comparison of night, day and 24 h motor activity data for the classification of depressive episodes. *Diagnostics* 10(3), 162 (2020)
24. Romero-Rosales, B.L., Tamez-Pena, J.G., Nicolini, H., Moreno-Treviño, M.G., Trevino, V.: Improving predictive models for alzheimer’s disease using gwas data by incorporating misclassified samples modeling. *Plos one* 15(4), e0232103 (2020)
25. Saritas, M.M., Yasar, A.: Performance analysis of ann and naive bayes classification algorithm for data classification. *International Journal of Intelligent Systems and Applications in Engineering* 7(2), 88–91 (2019)
26. Sarquella-Brugada, G., Campuzano, O., Cesar, S.: Sudden infant death syndrome caused by cardiac arrhythmias: only a matter of genes encoding ion channels? *Int J Legal Med* 130, 415–420 (2016)
27. Sweeting, J., Semsarian, C.: Cardiac abnormalities and sudden infant death syndrome. *Paediatr Respir Rev* 15(4), 301–6 (2014)

28. Tamez-Pena, J., Orozco, J., Sosa, P., Valdes, A., Nezhadmoghadam, F.: Ensemble of svm, random-forest and the bswims method to predict and describe structural associations with fluid intelligence scores from t1-weighted mri. In: Challenge in Adolescent Brain Cognitive Development Neurocognitive Prediction. pp. 47–56. Springer (2019)