Base de Conocimiento Generada dentro del Proyecto

**Optimización y Desarrollo de Métodos de Inteligencia Computacional Aplicados a la Solución de Problemas en Biomedicina**

**CIIC 232/2019**

Responsable: Dr. Luis Carlos Padierna García

Introducción

En este documento se describen los artículos, imágenes, señales y bases de datos generadas por participantes del proyecto o recopiladas de repositorios públicos indicados en artículos relacionados con el tema de inteligencia computacional aplicada a biomedicina. La base de conocimiento contiene

5X artículos del estado del arte sobre técnicas de inteligencia computacional aplicadas a la solución de problemas en biomedicina. También contiene 3 bancos de imágenes biomédicas, uno de señales ABR y 15 bases de datos con problemas de clasificación. 2 de los bancos de imágenes (*pie\_infrarrojo* y *Segmentación\_celular\_C*) fueron adquiridos por participantes del proyecto, el resto fue seleccionado de diversos repositorios públicos. Su evidencia se encuentra en el documento “Evidencia 5\_base conocimiento.pdf”

Tabla 1. Datasets generados con equipo y personal de la Universidad de Guanajuato, y datasets que abordan la misma problemática que los generados, disponibles en repositorios públicos.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| N | Dataset | Descripción | Propietario | Descripción(REF o URL) |
| 1 | Susan\_overnigth | 1250 imágenes de zona ventricular del cerebro |  |  |
| 2 | Pie\_infrarrojo | 200 imágenes en escala de grises de pies. 100 sanos y 100 con diabetes. | DCI\_UG (Dr. Carlos Villaseñor)/ IMSS T1 León | Tomadas con cámara infrarrojo por investigadores y estudiantes asociados |
| 3 | Segmentación\_celular | A) 100 originales bmp y máscaras png  B) 300 ya segmentadas |  |  |
| C) 150 5 clases png y máscaras  La carpeta "Clasificados 256x256" es el dataset con 150 imágenes. | DCI\_UG (Dr. Arturo González) / Hospital General Regional de León | Tomadas de frotis de AMO de 5 pacientes (A,B,C,D,E) con microscopio óptico a Mx1000 y cámara de celular de 12 Mpx |
| 4 | ABR | 465 señales ABR (232 pareadas) filtradas con pasa altos a 30kHz y pasabajos a 3kHz. |  | Obtenidas de 8 personas con SPL de 5 a 100. Frecuencia 1 o 4 kH |
| 5 | Benchmark datasets | 15 datasets biomédicos de uso común para prueba de algoritmos de clasificación. Los datasets están en formato LIBSVM | UCI Machine Learning Repository  Learning Repository [1]. | La descripción se adjunta en la Tabla 2. |

El dataset Susan\_overnight es una secuencia de videos del área ventricular del cerebro. La secuencia consiste en 2500 imágenes, 1250 para cada uno de dos canales A y B. El canal B resultó ser de interés para al Dra. Silvia Alejandra López Juárez, investigadora adscrita al Cuerpo Académico de Ingeniería Biomédica (CAIB), debido a que muestra el proceso de neurogénesis en una zona del cerebro.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Dataset  short label | Total cases (positive - negative) | Fts1 | Best Accuracy2 reported with RBF or *Multiple Kernels* | References |
| 1 | breast | 683 (239-444) | 10 | *98.03* | 97.31 | 97.18 | [2] [3] [4] |
| 2 | chronic | 400 (150-250) | 24 | 99.60 | [4] |
| 3 | column\_2C | 310 (100-210) | 7 | 87.00 | 86.02 | [5] [4] |
| 4 | cryotherapy | 90 (43-47) | 6 | 91.00 | [6] |
| 5 | diabetes | 768 (268-500) | 8 | *81.25* | 77.73 | 76.83 | [2] [3] [7] |
| 6 | fertility | 100 (12-88) | 9 | 88.00 | 89.19 | 88.04 | [5] [4] [8] |
| 7 | haberman | 306 (81-225) | 3 | 73.55 | 75.77 | 75.91 | [3] [9] [10] |
| 8 | heart | 270 (120-150) | 13 | *86.98*| 83.70 | 84.67 | [2] [7] [4] |
| 9 | immuno | 90 (19-71) | 7 | 88.00 | 85.46 | [6] [11] |
| 10 | liver | 345 (145-200) | 7 | 72.45 | 74.20 | 74.78 | [7] [12] [8] |
| 11 | mammo | 961(445-516) | 5 | 86.44 | [9] |
| 12 | parkinsons | 195 (48-147) | 22 | 95.30 | 95.98 | 98.88 | [13] [4] [14] |
| 13 | thoracic | 470 (70-400) | 17 | 85.30 | 85.15 | [5] [4] |
| 14 | transfusion | 748 (178-570) | 4 | 75.00 | 80.53 | [5] [9] |
| 15 | wpbc | 194 (46-148) | 33 | 80.09 | 81.22 | [7] [9] |

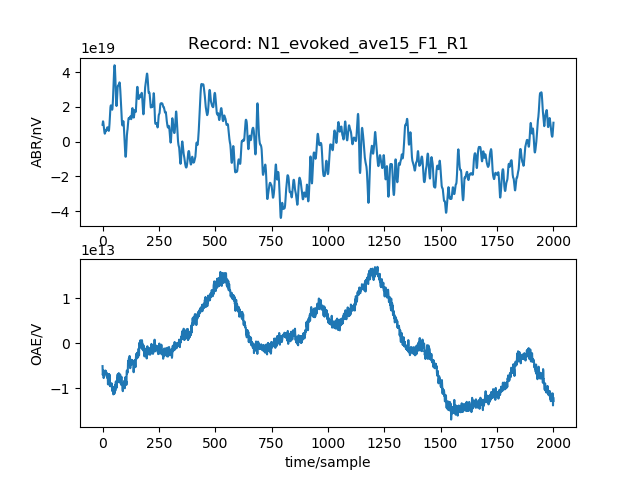


Figura 1. Ejemplo de señal ABR pareada. Las gráficas fueron generadas a partir de los datos en archivos binarios con scripts del lenguaje de programación Python.

# Referencias

|  |  |
| --- | --- |
| [1] | D. Dua y C. Graff, *UCI Machine Learning Repository [http://archive.ics.uci.edu/ml],* Irvine, CA: University of California, School of Information and Computer Science, 2019. |
| [2] | L. Dioşan, A. Rogozan y J. Pecuchet, «Improving classification performance of support vector machine by genetically optimising kernel shape and hyper-parameters,» *Applied Intelligence,* vol. 36 , nº 2, pp. 280-294, 2012. |
| [3] | A. López, X. Li y W. Yu, «Support Vector Machine Classification for Large Datasets Using Decision Tree and Fisher Linear Discriminant,» *Future Generation Computer Systems (36) 57-65,* vol. 36, pp. 57-65, 2014. |
| [4] | A. Rojas-Domínguez, L. C. Padierna, J. M. Carpio, H. J. Puga y H. Fraire, «Optimal Hyper-parameter Tuning of SVM Classifiers with Application to Medical Diagnosis,» *IEEE Access,* vol. 6, pp. 7164-7176, 2017. |
| [5] | R. Mantovani, A. Rossi, J. Vanschoren y B. d.-C. A. Bischl, «Effectiveness of Random Search in SVM hyper-parameter tuning,» de *International Joint Conference on Neural Networks (IJCNN)*, 2015. |
| [6] | A. Cüvitoglu y Z. Isik, «Evaluation Machine-Learning Approaches for Classification of Cryotherapy and Immunotherapy Datasets,» *International Journal of Machine Learning and Computing,* vol. 8, nº 4, pp. 331-335, 2018. |
| [7] | L. Sun, K.-A. Toh y Z. Lin, «A center sliding Bayesian binary classifier adopting orthogonal polynomials,» *Pattern Recognition,* vol. 48, nº 6, pp. 2013-2028, 2015. |
| [8] | Y. Xu, Z. Yang y X. Pan, «A Novel Twin Support-Vector Machine With Pinball Loss,» *IEEE Transactions on Neural Networks and Learning Systems,* vol. 28, nº 2, pp. 359-370, 2017. |
| [9] | H. l. Chen, B. Yang, S. j. Wang, G. Wang, D. y. Liu, H. z. Li y W. b. Liu, «Towards an optimal suppport vector machine classifier using a parallel particle swarm optimization strategy,» *Applied Mathematics and Computation,* vol. 239, pp. 180-197, 2014. |
| [10] | V. H. Moghaddam y J. Hamidzadeh, «New Hermite orthogonal polynomial kernel and combined kernels in Support Vector Machine classifier,» *Pattern Recognition,* vol. 60, pp. 921-935, 2016. |
| [11] | Y. F. Hernández-Julio, M. J. Prieto-Guevara, W. Nieto-Bernal, I. Meriño-Fuentes y A. Guerrero-Avendaño, «Framework for the Development of Data-Driven Mamdani-Type Fuzzy Clinical Decision Support Systems,» *Diagnostics,* vol. 9, nº 2, p. 52, 2019. |
| [12] | J. Zhao, Z. Yang y X. Yitian, «Nonparallel least square support vector machine for classification,» *Applied Intelligence,* pp. 1-10, 2016. |
| [13] | L. Shen, H. Chen, Yu, W. Kang, B. Zhang, H. Li, Y. Bo y D. Liu, «Evolving support vector machines using fruit fly optimization for medical data classification,» *Knowledge-Based Systems,* vol. 96, nº 15, pp. 61-75, March 2016. |
| [14] | M. Li, X. Lu, X. Wang, S. Lu y N. Zhong, «Biomedical classification application and parameters optimization of mixed kernel SVM based on the information entropy particle swarm optimization,» *Computer Assited Surgery,* vol. 21, nº 1, pp. 132-141, 2016. |