UNIVERSIDAD DE GUANAJUATO

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 artículos del estado del arte sobre técnicas de inteligencia computacional aplicadas a la solución de problemas en biomedicina. También contiene tres bancos de imágenes biomédicas, uno de señales ABR y 15 bases de datos con problemas de clasificación. Dos 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 señalados por los artículos más relevantes del estado del arte. La Tabla 1 describe brevemente los datasets de la base de conocimiento.

Descripción de la Base de Conocimiento

El dataset *Susan_overnight* es una secuencia de video 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 la Dra. Silvia Alejandra López Juárez, investigadora miembro del Cuerpo Académico de Ingeniería Biomédica (**CAIB**), debido a que muestra el proceso de neurogénesis en una zona del cerebro. Actualmente se trabaja con este banco de imágenes segmentando regiones de interés para rastreo de nuevas neuronas observadas en las imágenes. La Figura 1 muestra una de las 2500 imágenes.

DEPARTAMENTO DE INGENIERÍAS QUÍMICA, ELECTRÓNICA Y BIOMÉDICA
DIVISIÓN DE CIENCIAS E INGENIERÍAS
Loma del Bosque #103, Col. Lomas del Campestre. León, Gto. C.P. 37150
Tel. +52 (477) 7 88 51 00 ext. 8522
lc.padierna@ugto.mx
www.dci.ugto.mx/depingquim

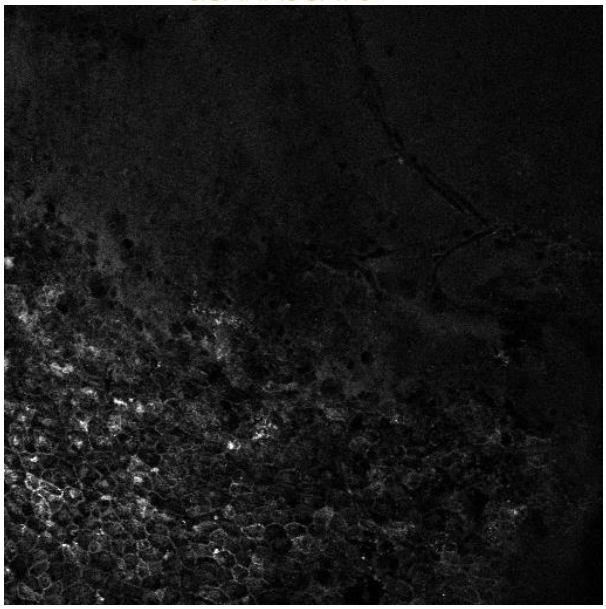

Figura 1. Ejemplo de imagen de zona ventricular del cerebro (Canal A). Los puntos más brillantes son núcleos de neuronas.

Tabla 1. Imágenes, señales y bases de datos obtenidas por los participantes del proyecto

N	Dataset	Descripción	Propietario
1	Susan_overnigth	1250 imágenes de zona ventricular del cerebro (Canal B) que representan una secuencia de video del proceso de Neurogénesis.	Repositorio público
2	Pie_infrarrojo	200 imágenes en escala de grises de 100 de pies sanos y 100 de pies con diabetes. Tomadas con cámara infrarrojo por investigadores y estudiantes asociados	DCI_UG (Dr. Carlos Villaseñor) / IMSS T1 León
3	Segmentación_celular	A) 100 imágenes de células blancas segmentadas y sus respectivas máscaras señalando la región deseada a segmentar B) 300 imágenes de células blancas segmentadas y sus respectivas máscaras señalando la región deseada a segmentar	Repositorio público
		C) 150 imágenes de células blancas segmentadas y sus respectivas máscaras señalando la región deseada a segmentar. Tomadas de frotis de Aspirado de Médula Ósea de 5 pacientes (A,B,C,D,E) con microscopio óptico a Mx1000 y cámara de celular de 12 Mpx	DCI_UG (Dr. Arturo González) / Hospital General Regional de León
4	ABR	465 señales ABR (232 pareadas) filtradas con pasa altos a 30kHz y pasa bajos a 3kHz. Obtenidas de 8 personas con audición normal, usando SPL de 5 a 100. Frecuencia 1 o 4 kHz	Repositorio público
5	Benchmark datasets	15 datasets biomédicos de uso común para prueba de algoritmos de clasificación (ver tabla 2). Los datasets están en formato LIBSVM.	UCI Machine Learning Repository Learning Repository [1].

El dataset *Pie_infrarrojo* consta de 200 imágenes obtenidas de 24 personas con diabetes y 21 personas sin diabetes. Las imágenes se obtuvieron con ayuda de médicos del Instituto Mexicano del Seguro Social (IMMS T1, León, Guanajuato). La Figura 2 muestra ejemplos de estas imágenes.

Figura 2. Izquierda: ejemplo de pies de una persona sin diabetes. Derecha: ejemplo de pies de una persona con diabetes.

El banco de imágenes *Segmentación_celular* consta de 3 datasets. Dos de ellos, etiquetados como A) y B) fueron obtenidos de un repositorio público creado por investigadores que utilizan técnicas de inteligencia computacional para conteo de células blancas. El dataset restante "C" fue generado por investigadores y alumnos asociados al proyecto mediante microscopía a partir de aspirados de muestras de médula ósea proporcionados por el Hospital General de León, Guanajuato. La Figura 3 muestra ejemplos de las imágenes del dataset C).

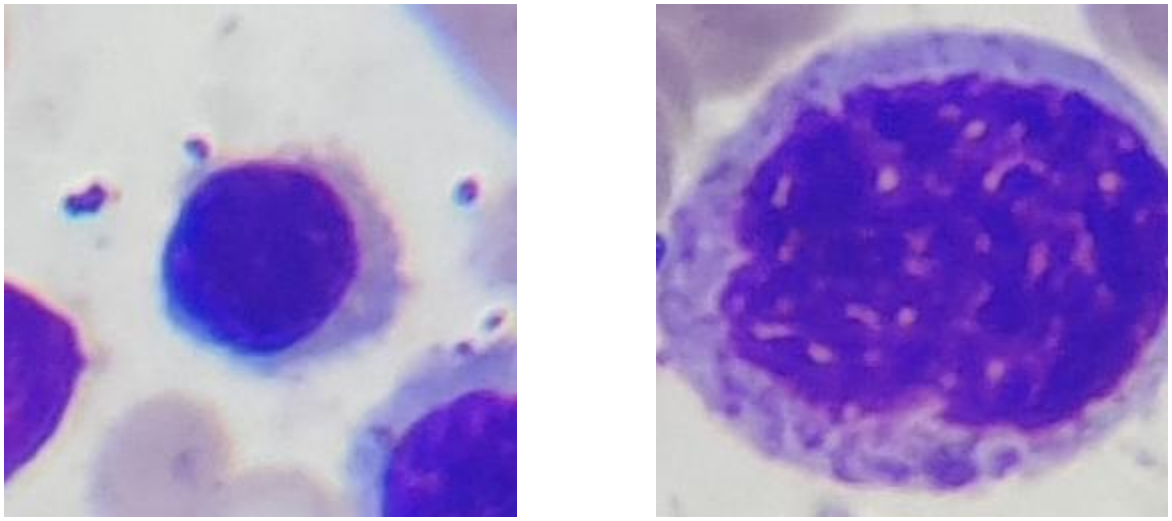

Figura 3. Izquierda: ejemplo de célula blanca (Linfocito). Derecha: ejemplo de célula blanca (Basófilo)

DEPARTAMENTO DE INGENIERÍAS QUÍMICA, ELECTRÓNICA Y BIOMÉDICA
DIVISIÓN DE CIENCIAS E INGENIERÍAS
Loma del Bosque #103, Col. Lomas del Campestre. León, Gto. C.P. 37150
Tel. +52 (477) 7 88 51 00 ext. 8522
lc.padierna@ugto.mx
www.dci.ugto.mx/depingquim

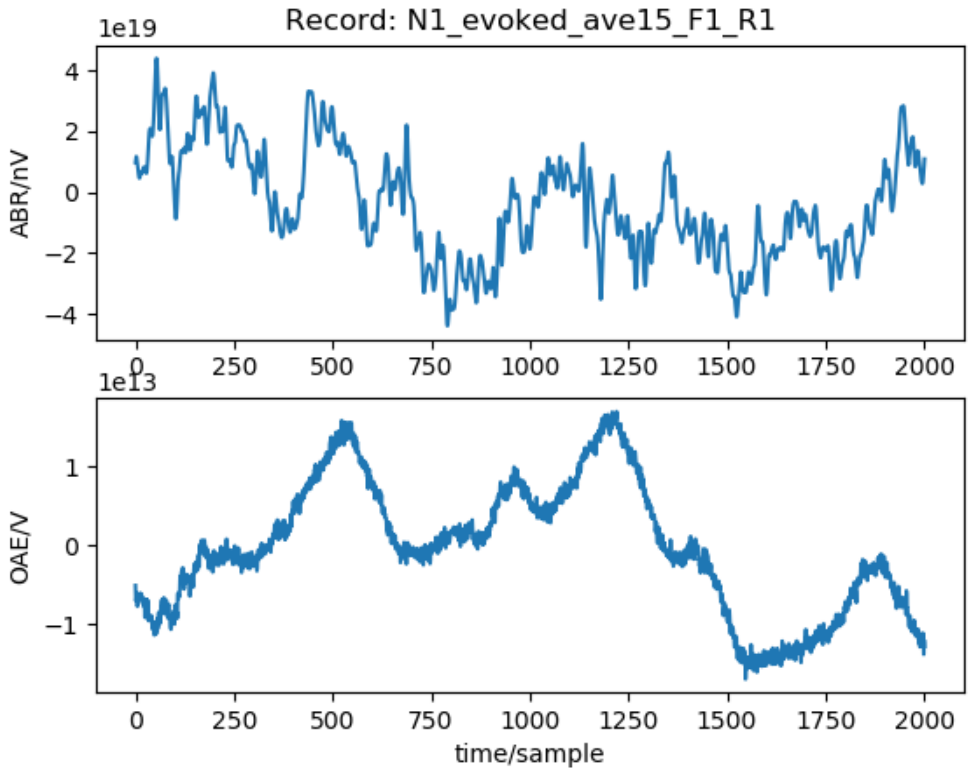

Figura 4. Ejemplo de señal ABR pareada. Las gráficas fueron generadas a partir de los datos en archivos binarios con scripts del lenguaje de programación Python.

El dataset *ABR* contiene 465 señales procesadas a partir archivos binarios. Estas señales representan la Respuesta Auditiva del Tallo Cerebral (Auditory Brainsteam Responses) de 8 personas con audición normal. El objetivo es determinar la presencia o ausencia de las ondas características de una señal ABR.

Los 15 datasets benchmark se describen en la Tabla 2, estos son: predicción de cáncer de pecho (breast), enfermedad crónica del riñón (chronic), normalidad ortopédica de columna vertebral (column_2C), resultados de tratamiento de verrugas empleando crioterapia (cryotherapy), diagnóstico de diabetes mellitus tipo 2 (diabetes), identificación de concentración alterada de esperma (fertility), predicción de sobrevivencia después de cirugía de cáncer de pecho (haberman), determinación de enfermedades en el corazón (heart), resultados de tratamiento para verrugas empleando inmunoterapia (immuno), alteraciones en el hígado causadas por alcohol (liver), discriminación de masas mamográficas malignas o benignas (mammo), predicción de Parkinson con base en mediciones de voz (parkinsons), esperanza de vida post-cirujía en pacientes de cácer de pulmón después de cirugía torácica (thoracic), predicción sobre la donación de sangre (transfusion) y pronóstico sobre cáncer de pecho (wpbc). Todos los datasets están públicamente disponibles en el repositorio de la UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/index.php).

Tabla 2. Listado de los 15 problemas biomédicos de clasificación empleados

	Dataset Etiqueta corta	Casos Totales (positivos - negativos)	Dim	Mejor exactitud en clasificación lograda con RBF o kernels múltiples
1	breast	683 (239-444)	10	*98.03* | 97.31 | 97.18
2	chronic	400 (150-250)	24	99.60
3	column_2C	310 (100-210)	7	87.00 | 86.02
4	cryotherapy	90 (43-47)	6	91.00
5	diabetes	768 (268-500)	8	*81.25* | 77.73 | 76.83
6	fertility	100 (12-88)	9	88.00 | 89.19 | 88.04
7	haberman	306 (81-225)	3	73.55 | 75.77 | 75.91
8	heart	270 (120-150)	13	*86.98* | 83.70 | 84.67
9	immuno	90 (19-71)	7	88.00 | 85.46
10	liver	345 (145-200)	7	72.45 | 74.20 | 74.78
11	mammo	961(445-516)	5	86.44
12	parkinsons	195 (48-147)	22	95.30 | 95.98 | 98.88
13	thoracic	470 (70-400)	17	85.30 | 85.15
14	transfusion	748 (178-570)	4	75.00 | 80.53
15	wpbc	194 (46-148)	33	80.09 | 81.22

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.