



Informe Técnico Final del Proyecto

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

Responsable:

Dr. Luis Carlos Padierna García¹

Participantes internos:

Dr. Carlos Villaseñor Mora¹

Dr. Arturo González Vega¹

Participante externo:

Dr. Pablo Genaro Martínez Torres²

¹Departamento de Ingenierías Química, Electrónica y Biomédica, División de Ciencias e Ingenierías, Universidad de Guanajuato Campus León

²Instituto de Física y Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo



Introducción

El contenido del presente reporte técnico parte de los objetivos, metas y entregables planteados al inicio del proyecto de investigación. Para cada objetivo, meta y entregable se describen las actividades efectuadas para alcanzarlos y se incluyen evidencias de los productos generados. Las evidencias de software y bases de conocimiento son brevemente descritas y referenciadas a un repositorio que las concentra y que estará disponible para la comunidad de la Universidad de Guanajuato.

Objetivos General

El Proyecto CIIC 232/2019 tuvo como objetivo general el *desarrollar y optimizar métodos de inteligencia computacional para resolver problemas abiertos y actuales de biomedicina*.

Objetivo cumplido. Se desarrolló un método automático de inteligencia computacional para generación de clasificadores que resuelven problemas de biomedicina. El método que denominamos Smart Evolution of Ensembled Kernels for Support Vector Machines (SEEKS) fue capaz de resolver 15 problemas de clasificación biomédica (ver Tabla 1 en objetivo específico 2) con rendimientos similares o superiores a los obtenidos con técnicas desarrolladas por expertos. Como evidencia se tiene un manuscrito enviado a revista, el cual se encuentra en archivo “**Evidencia_1_manuscrito_enviado.pdf**” https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_1_manuscrito_enviado.pdf

Objetivos específicos

1. Construir una base de conocimiento formada por artículos científicos del estado del arte, imágenes, señales, software y bases de datos sobre problemas biomédicos. La selección de los elementos que integrarán la base tendrá como requisito el que sirva de plataforma para fortalecer y extender el desarrollo de proyectos de investigación actuales y futuros del cuerpo académico.

Objetivo cumplido. La base de conocimiento consta de 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 fueron adquiridos por participantes del proyecto, el resto fue seleccionado de diversos repositorios públicos. El documento describiendo esta evidencia se encuentra en el archivo:

https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_5_base_de_conocimiento.pdf

Las imágenes, señales y bases de datos se encuentran en:

https://github.com/padiernacarlos/CIIC-232-2019/tree/master/Base_de_Conocimiento



2. Aplicar los métodos computacionales del Cuerpo Académico de Ingeniería Biomédica (**CAIB**), especialmente Máquinas de Soporte Vectorial y Metaheurísticas, a una selección de problemas biomédicos. Los problemas seleccionados serán aquellos que se encuentren en el estado del arte y que sean susceptibles de comparación directa para la evaluación de nuestros métodos.

Objetivo cumplido. Se aplicaron Máquinas de Soporte Vectorial con diversas funciones kernel y 2 metaheurísticas (Programación Genética y Algoritmos de Estimación de Distribución) a una selección de 15 problemas de clasificación biomédica. La Tabla 1 presenta una breve descripción de los problemas. La información sobre los problemas de clasificación puede ser consultada en:

https://github.com/padiernacarlos/CIIC-232-2019/tree/master/Base_de_Conocimiento

Tabla 1. 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	wdbc	194 (46-148)	33	80.09 81.22

Los datasets descritos en la Tabla 1. (y sus etiquetas cortas) 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-cirugía en pacientes de cáncer de pulmón después de cirugía torácica (*thoracic*), predicción sobre la donación de sangre (*transfusion*) y pronóstico sobre cáncer de pecho (*wdbc*). Todos los datasets están públicamente disponibles en el repositorio de la UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/index.php>).



3. Diseñar y realizar experimentos para comparación de nuestros métodos contra los disponibles en la literatura. **Objetivo cumplido.** El diseño experimental y comparación contra métodos de la literatura se describe en la sección 4 del manuscrito “**Evidencia_1_manuscrito_enviado.pdf**”
https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_1_manuscrito_enviado.pdf
4. Evaluar los resultados experimentales obtenidos en el objetivo anterior y buscar puntos de optimización de las técnicas aplicadas. **Objetivo cumplido.** Los resultados experimentales de los métodos desarrollados, comparados contra métodos de la literatura se describen en la sección Results and Discussion del manuscrito “**Evidencia_1_manuscrito_enviado.pdf**”
https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_1_manuscrito_enviado.pdf
5. Desarrollar nuevos modelos computacionales que compitan o mejoren los disponibles en el estado del arte. **Objetivo cumplido.** Se desarrollaron métodos computacionales para realizar evolución de kernels de máquinas de soporte vectorial. La descripción de estos modelos se encuentra en la sección III. Smart Evolution of Kernels for SVMs dentro del manuscrito “**Evidencia_1_manuscrito_enviado.pdf**”
https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_1_manuscrito_enviado.pdf
6. Reportar la revisión de la literatura, los diseños y los resultados de la evaluación comparativa en foros de divulgación y difusión, académicos y científicos. **Objetivo cumplido.** Se dictó una plática científica, una plática de divulgación y se presentó un póster en un evento académico local y otro internacional. Las constancias de participación en eventos académicos se encuentran en el documento “**Evidencia_2_constancias_presentaciones.pdf**” y pueden consultarse en:
https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_2_constancias_presentaciones.pdf

Metas Científicas

- 1.1 Un manuscrito sometido a revisión en una revista JCR describiendo los hallazgos principales de la investigación. **Meta cumplida.** El manuscrito enviado a revista se encuentra en el documento denominado “**Evidencia_1_manuscrito_enviado.pdf**” accesible en:
https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_1_manuscrito_enviado.pdf
- 1.2 Un documento describiendo la revisión del estado del arte que contenga al menos entre 50 y 100 artículos científicos sobre técnicas de inteligencia computacional aplicadas a la solución de problemas biomédicos en los últimos 5 años. **Meta cumplida.** Se realizó una revisión a 55 artículos del estado del arte. El documento probatorio se encuentra en el archivo “**Evidencia_Meta_1-2_Revisión_estado_del_arte.pdf**” disponible en:

https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_Meta_1-2_Revisión_estado_del_arte.pdf



- 1.3 Un reporte técnico describiendo la construcción de una base de conocimiento en el que se describa el tipo y cantidad de imágenes, bioseñales, bases de datos, artículos y software con que cuenta el CAIB. **Meta cumplida.** El reporte técnico se encuentra en el documento denominado “Evidencia_5_base_de_conocimiento.pdf” y está disponible en:

https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_5_base_de_conocimiento.pdf

- 1.4 Un prototipo en software con los modelos computacionales desarrollados, el cual será transferible y accesible a la comunidad estudiantil y el profesorado de la Universidad de Guanajuato. **Meta cumplida.** La interfaz gráfica del prototipo de software se describe en el manual denominado “Evidencia 4_manual_interfaz_gráfica.pdf” y está disponible en:

https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_4_manual_interfaz_gráfica.pdf

- 1.5 Una presentación en evento académico o conferencia científica dando a conocer los resultados parciales del proyecto. **Meta cumplida.** Las constancias de participación en eventos académico se encuentran en el documento “Evidencia_2_constancias presentaciones.pdf” y está disponible en:

https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_2_constancias_presentaciones.pdf

Metas de Formación de Recursos Humanos

En este proyecto se fomentará la participación de estudiantes de licenciatura distribuidos de la siguiente forma:

1. Involucramiento de 3 estudiantes de licenciatura en tareas de desarrollo de software y análisis de textos científicos. Los estudiantes cursan las últimas materias de su ingeniería, 2 de ellos serán evaluados en la materia de Desarrollo Experimental y 1 más en la materia de Laboratorio Avanzado. **Meta cumplida.** Como evidencia del involucramiento de los 3 estudiantes, se presentan sus evaluaciones en las materias mencionadas y una constancia describiendo el proyecto en el que participaron. La evidencia está en el archivo “Evidencia_MetaRH_1_involucramiento.pdf” disponible en:

https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_MetaRH_1_involucramiento.pdf

2. Formación en investigación de 2 estudiantes de licenciatura bajo la modalidad de Servicio Social Universitario o Verano de Investigación. **Meta cumplida.** Se formó como investigadores a 2 estudiantes de licenciatura (Walter Adame González y Martha Susana García Delgado). Ambos se encuentran realizando tesis de licenciatura y tienen planes de cursar estudios de posgrado. La constancia de su formación se encuentra en el documento “Evidencia_Meta_RH 2y3_formación_investigadores.pdf” disponible en:

https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_MetaRH_2y3_formación_investigadores.pdf



3. Formación en investigación de 1 estudiante de licenciatura bajo la modalidad de Servicio Social Profesional con aspiraciones a tesis. **Meta cumplida.** La constancia de su formación se encuentra en el documento “**Evidencia_Meta_RH 2y3_formación_investigadores.pdf**” disponible en:

https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_MetaRH_2y3_formación_investigadores.pdf

4. Es muy importante hacer énfasis en que las actividades asignadas a los estudiantes involucrados están pensadas para fomentar las vocaciones científicas, buscando que algunos de ellos se interesen en continuar con estudios de posgrado. **Meta cumplida.** Se fomentó la vocación científica y como resultado el participante Edwin Armando Bedolla Montiel ingresó a estudiar la Maestría en Ciencias Aplicadas. La evidencia es su formato de pago de inscripción al 1er semestre de maestría “**Evidencia_MetaRH_4_ingreso_maestría.pdf**” disponible en:

https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_MetaRH_4_ingreso_maestría.pdf

Productos Entregables

1. Envío de artículo científico a revista JCR. Integración de los experimentos, resultados y conclusiones principales del proyecto. Este producto fue derivado del cumplimiento de la meta científica 1.1 y su evidencia se encuentra en el archivo “**Evidencia_1_manuscrito_enviado.pdf**” disponible en:

https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_1_manuscrito_enviado.pdf

2. Constancia de presentación en evento académico o científico. Presentación de los resultados y avances intermedios del proyecto. Este producto fue derivado del cumplimiento de la meta científica 1.5 y su evidencia se encuentran en el documento “**Evidencia_2_constancias_presentaciones.pdf**” y está disponible en:

https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_2_constancias_presentaciones.pdf

3. Reporte Técnico. Descripción de los experimentos y actividades realizadas por los estudiantes involucrados. Este producto fue derivado del cumplimiento de la meta científica 1.3 y de las metas 1,2 y 3 de formación de recursos humanos. Su evidencia se encuentra en el documento “**Evidencia_3_reporte_técnico_alumnos.pdf**”, disponible en:

https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_3_reporte_técnico_alumnos.pdf



4. Prototipo de Software. Aplicación con interfaz gráfica que permita la ejecución y prueba de los métodos implementados a lo largo del proyecto. El prototipo será transferible y accesible a la comunidad UG. Este producto fue derivado del cumplimiento de la meta científica 1.4. La interfaz gráfica del prototipo de software se describe en el manual denominado “Evidencia 4_manual_interfaz_gráfica.pdf” y está disponible en:

https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_4_manual_interfaz_gráfica.pdf

El prototipo se encuentra disponible en:

[https://github.com/padiernacarlos/CIIC-232-2019/tree/master/Interfaz_Gráfica\(Prototipo_Software\)](https://github.com/padiernacarlos/CIIC-232-2019/tree/master/Interfaz_Gráfica(Prototipo_Software))

5. Base de Conocimiento. Descripción y copia de información médica en forma de imágenes, bases de datos y archivos de bioseñales. Este producto fue derivado del cumplimiento del objetivo específico 1. El documento describiendo esta la base de conocimiento se encuentra en el archivo:

https://github.com/padiernacarlos/CIIC-232-2019/blob/master/Evidencia_5_base_de_conocimiento.pdf

Las imágenes, señales y bases de datos y artículos se encuentran en:

https://github.com/padiernacarlos/CIIC-232-2019/tree/master/Base_de_Conocimiento