

UNIVERSIDAD DE SANTIAGO DE CHILE
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA INFORMÁTICA

**Algoritmo de Optimización por Enjambre de
Partículas con computación paralela de
modelos de Autorregulación Cerebral en R**

Informe N° 3
Reformulación

Nombre: Benjamin Jorquera Jorquera

Rut: 19.182.719-8

Año estimado de egreso: 2022

Teléfono: 9 93362209

E-mail: benjamin.jorquera@usach.cl

Profesor Patrocinador: José Luis Jara

Fecha: 14 de Julio de 2022

RESUMEN

El mecanismo biológico de la Autorregulación Cerebral no puede ser cuantificado de manera efectiva, ya que los métodos dinámicos de cambios bruscos de presión utilizados pueden ser perjudiciales para la salud de las personas. En la descripción del problema y análisis de la solución del presente documento se plantea que los métodos de medición actuales aún no han sido validados para pacientes enfermos, y si bien presentan buenos resultados tampoco podrían ser utilizados como muestras reales para estas señales debido a su complejidad y utilidad. En la sección de descripción de la solución propuesta se plantea evaluar la posibilidad de utilizar nuevos métodos para resolver el problema, utilizando un nuevo enfoque de solución de un algoritmo en R de optimización metaheurística y computación paralela, denominado Optimización por Enjambre de Partículas. El objetivo de esta investigación es mejorar el rendimiento y calidad de generar buenos modelos de autorregulación para prevenir posibles problemas cerebrales en las personas, en base a acotar mejor el espacio de búsqueda de estas soluciones, escoger modelos de calidad de manera automática sin necesidad de intervenir en el paciente, y reducir el tiempo de procesamiento del procedimiento, para poder diagnosticar a tiempo posibles afecciones cerebrales. Luego se explica la metodología, herramientas y ambiente de desarrollo a utilizar, y por último el plan de trabajo del proyecto.

Palabras clave: *autorregulación cerebral, optimización por enjambre de partículas, selección automática, computación paralela.*

TABLA DE CONTENIDOS

1. DESCRIPCIÓN DEL PROBLEMA.....	1
1.1. MOTIVACIÓN	1
1.2. ENUNCIADO DEL PROBLEMA.....	2
2. ANÁLISIS DE LA SOLUCIÓN	2
2.1. ESTADO DEL ARTE	2
2.2. ENFOQUES DE SOLUCIÓN	5
2.3. JUSTIFICACIÓN DEL ENFOQUE SELECCIONADO	6
3. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA	7
3.1. PROPÓSITOS DE LA SOLUCIÓN	7
3.2. CARACTERÍSTICAS DE LA SOLUCIÓN	7
3.3. ALCANCES Y LIMITACIONES DE LA SOLUCIÓN	8
3.4. EVALUACIÓN DE LA SOLUCIÓN	8
4. OBJETIVO DEL PROYECTO	9
4.1. OBJETIVO GENERAL.....	9
4.2. OBJETIVOS ESPECÍFICOS	9
5. METODOLOGÍA, HERRAMIENTAS Y AMBIENTE DE DESARROLLO	9
5.1. METODOLOGÍA PARA USAR.....	9
5.2. HERRAMIENTAS DE DESARROLLO.....	11
5.3. AMBIENTE DE DESARROLLO	11
6. PLAN DE TRABAJO.....	12
7. REFERENCIAS.....	14

1. DESCRIPCIÓN DEL PROBLEMA

1.1. MOTIVACIÓN

La Autorregulación Cerebral (AC) es un mecanismo de defensa homeostático que controla la Velocidad del Flujo Sanguíneo Cerebral (VFSC), para mantener un flujo sanguíneo cerebral relativamente constante y estable, a pesar de cualquier tipo de cambio en la Presión de Perfusión Cerebral (PPC), la cual se encarga de la irrigación sanguínea del cerebro a través de su lecho vascular (Muñoz, 2009).

Los estudios demuestran que este mecanismo se encarga de suministrar la sangre necesaria para la demanda metabólica de las actividades y situaciones críticas del cerebro. Así que un mal funcionamiento o deterioro de la AC puede provocar graves daños y consecuencias para la salud (Czosnyka et. al., 2001), como por ejemplo los accidentes cerebro vasculares (*stroke*), que son estadísticamente la primera causa de muerte en Chile (MINSAL, 2017) y la segunda a nivel mundial (WHO, 2020). Por lo tanto, es importante realizar evaluaciones de la AC que permitan identificar y prevenir estas situaciones, además de mejorar el diagnóstico, monitoreo y pronóstico de las enfermedades cerebrales en las personas (Panerai, 1998).

Sin embargo, no existe una forma estándar de medir la AC de una persona, pero sí existe una gran variedad de métodos y criterios de medición. En la actualidad estos métodos dependen del estudio realizado, presentando diferentes enfoques y características. Los estudios más recientes realizados por el Departamento de Ingeniería Informática en la Universidad de Santiago de Chile corresponden a modelar la AC a través de Máquinas de Vectores de Soporte (*Support Vector Machine*, SVM) (Miranda, 2016) utilizando técnicas no invasivas, y luego la utilización de la Optimización Multi Objetivo (OMO) para determinar la información más relevante para seleccionar modelos de AC, acotando su espacio de búsqueda (Vallejos, 2017). Los resultados obtenidos por estos estudios indican que la utilización de las SVM como método de aprendizaje de los modelos aumentan la complejidad de estos, influyendo directamente en el tiempo de ejecución, el cual depende también del tamaño de las muestras utilizadas. Por otra parte,

la OMO sobre los modelos entrenados por las SVM obtuvo buenos resultados, pero, debido a sus limitaciones no mejoraron la calidad de los modelos ya existentes.

Basándose en estos antecedentes, con esta investigación se espera utilizar un método de optimización alternativo a los existentes para que realice una búsqueda y selección más avanzada de los modelos de AC, comparando los resultados obtenidos de utilizar optimizaciones metaheurísticas sobre las muestras entrenadas y preprocesadas, y por otro lado sobre las muestras originales, utilizando el enfoque de las Variaciones Espontáneas de Presión (VEP) de los pacientes. Con el fin de reducir los tiempos de ejecución y procesamiento de los datos de las muestras, encontrando soluciones óptimas a las ya existentes de los modelos que más se aproximen a las señales biológicas de la AC. Además, aportar información valiosa sobre estos métodos de optimización para el problema planteado.

1.2. ENUNCIADO DEL PROBLEMA

¿Cómo contribuir al diagnóstico oportuno de posibles afecciones cerebrales en las personas de manera no invasiva, utilizando un método de optimización metaheurística para reducir el tiempo de procesamiento en la búsqueda y selección de modelos de calidad de Autorregulación Cerebral, comparado con los resultados de los métodos actuales?

2. ANÁLISIS DE LA SOLUCIÓN

2.1. ESTADO DEL ARTE

Se tiene constancia de que los primeros estudios de la AC comienzan con representaciones de modelos estáticos al final de la década de los 50 (Lassen, 1959), que no contaban con la tecnología necesaria para considerar los pequeños cambios en las señales del Flujo Sanguíneo Cerebral (FSC).

Luego surgen investigaciones sobre modelos dinámicos del fenómeno de la autorregulación, los cuales se dividen en dos áreas de estudio: los modelos generados a partir de VEP y los modelos que inducen cambios bruscos en la presión sanguínea. Algunas técnicas de estos últimos son la Valsalva (intento de exhalar aire con la boca y la nariz cerradas) y los manguitos (dispositivo que interrumpe la circulación sanguínea de manera temporal en los muslos).

Sin embargo, estos métodos de cambios espontáneos de presión son considerados invasivos y podrían provocar alteraciones a la salud de algunos pacientes enfermos. Debido a esto se empezaron a realizar estudios con modelos que utilizaban el enfoque VEP con régimen de respiración para estudiar el fenómeno de la AC, donde los pacientes se encontraban en estado de reposo. Estos estudios han demostrado que los modelos no lineales presentan un buen rendimiento para este método, pero algunos de estos no podían ser considerados a priori debido a la poca cantidad de sujetos de prueba utilizados, y también por que los resultados no fueron comparados con otros métodos (Miranda, 2016).

La evolución de las metodologías de medición de la AC avanza con estudios basados en nuevas técnicas comparativas, para evidenciar las diferencias entre los métodos de medición estáticos y dinámicos presentados con anterioridad. Así, se pudo descubrir la existencia de una respuesta continua, la cual seguía un patrón único para todos los individuos en un modelo de 10 niveles para clasificar su capacidad autorregulatoria (Tiecks et. al. 1995).

También se han realizado diversos estudios a partir de los métodos dinámicos con modelos de estructuras lineales y no lineales, donde estos últimos han presentado buenos resultados. Un enfoque sobre métodos dinámicos aceptado y utilizado en el Departamento de Ingeniería Informática de la Universidad Santiago de Chile son las SVM, que utilizan aprendizaje supervisado para entregar modelos representativos de la AC mediante una regresión que permite estimar la VFSC. Los estudios de Chacón et. al. (2011) aplican metodologías basadas en regresión multivariada con aplicaciones en

clasificadores, permitiendo mediciones del paciente en estado de reposo con cambios espontáneos de presión.

Actualmente el estudio de la AC consiste en medir la presión sanguínea arterial media (PAM) y la VFSC que entra por las arterias cerebrales, y así modelar su comportamiento utilizando distintas técnicas de aprendizaje.

A partir del aprendizaje se pueden filtrar los modelos para escoger los que mejor representen el fenómeno de la AC. Los filtros, en primera instancia, constan de tres reglas visuales, donde posteriormente se agregarán cuatro nuevas reglas con un algoritmo basado en los puntajes de las características de la señal. Este nuevo enfoque de filtro realiza una búsqueda exhaustiva (*grid*) en base a los filtros normales, y luego el filtro con las nuevas reglas, donde el investigador le otorga valor a ciertos comportamientos de la señal, obteniendo modelos de posibles señales autorregulatorias humanas (Miranda, 2016).

Los modelos seleccionados por este nuevo filtro son sometidos a una optimización para encontrar el mejor modelo del aprendizaje realizado. Como la optimización de encontrar el mejor modelo depende de la función objetivo, se han realizado estudios donde se utilizan nuevos enfoques para abordar este problema, como la OMO junto al principio de Pareto, que cambia la función objetivo al dominio múltiple (Vallejos, 2017).

Por otro lado, estudios sugieren utilizar optimizaciones metaheurísticas para resolver problemas más complejos, debido a la estrategia que utilizan para recorrer el espacio de búsqueda, por lo que una Optimización por Enjambre de Partículas (*Particle Swarm Optimization*, PSO) mejorada, basada en el cambio dinámico del peso inercial, supone una mejora de las habilidades de exploración y explotación del algoritmo de OMO, manteniendo la diversidad y convergencia de las soluciones óptimas de Pareto (Zhang et. al., 2013). También se han realizado estudios de optimizaciones con enfoques metaheurísticos como optimización por PSO basados en Recocido Simulado (*Simulated Annealing*, SA) para encontrar parámetros de modelos de AC (Sharma et. al., 2011).

2.2. ENFOQUES DE SOLUCIÓN

Los algoritmos metaheurísticos son métodos estocásticos de búsqueda global y pueden localizar eficientemente la vecindad del óptimo global la mayoría de las ocasiones en un tiempo computacional aceptable (Hernández et. al., 2013).

Una alternativa es utilizar la a Optimización por Colonias de Hormigas (OCH) que se inspira en el comportamiento y comunicación de las hormigas. El algoritmo se basa en la comunicación indirecta de una colonia de hormigas artificiales mediante los rastros de feromona artificial. Los rastros sirven como información utilizada por las hormigas para construir soluciones candidatas para el problema (Hernández et. al., 2013).

Otra posibilidad la constituye el método de PSO, que es un algoritmo inspirado en el comportamiento de la naturaleza, en particular del movimiento de algunos seres vivos como las bandadas de pájaros, los bancos de peces o las manadas de mamíferos, donde su actuación colectiva proporciona nuevas técnicas de búsqueda y optimización, modificando cada partícula en base a su experiencia y la experiencia de la partícula vecina (Jiménez et. al., 2016).

La PSO es reconocida como una herramienta útil para resolver problemas de optimización. Es un método de búsqueda estocástico poblacional donde cada partícula del enjambre encuentra un camino de acuerdo con la siguiente información:

1. Su propia posición.
2. La mejor posición global del enjambre completo.

A través de la cooperación entre los individuos, el grupo es capaz de conseguir su objetivo de manera eficaz y efectiva (Sharma et. al., 2011). Sin embargo, la PSO estándar posee las siguientes limitaciones:

1. Convergencia prematura, las razones son:
 - a. La velocidad de la partícula se hace cero después de algunas iteraciones.
 - b. Pérdida de diversidad.
2. Sensibilidad a parámetros de entrada y tamaño de la población.

Así, aparecen alternativas como el algoritmo de Optimización por Enjambre de Partículas basado en Recocido Simulado (*Simulated Annealing-based Particle Swarm Optimization*, SAPSO), que puede disminuir las desventajas de la PSO estándar mejorando su habilidad de encontrar óptimos globales. También en otros estudios se han introducido relaciones de mapeo entre pesos inerciales y temperatura de recocido, mejorando la velocidad y la precisión (Sharma et. al., 2011). Sin embargo, en la mayoría de los algoritmos SAPSO, SA ha sido implementado para encontrar la mejor solución global alrededor de una posición. Por lo tanto, la exploración se concentra solo en ese espacio, donde es posible que la partícula quede atrapada en un óptimo local, y que la nueva posición generada también caiga en el mismo espacio, por lo que aumentar la diversidad no es suficiente en algunas ocasiones.

Estudios se inspiran en la idea de que la información de distintas relaciones se puede combinar para producir otro tipo de relaciones, proponiendo los métodos de Optimización por Multi Enjambre de Partículas (*Multi-Swarm Particle Swarm Optimization*, MSPSO), que son propuestas estrategias para balancear la exploración y la explotación de los métodos PSO estándar (Xia et. al., 2018). A pesar de tener buen rendimiento en términos de precisión en funciones complejas, este método presenta rendimientos moderados en utilización de tiempo.

2.3. JUSTIFICACIÓN DEL ENFOQUE SELECCIONADO

El enfoque seleccionado es la PSO, debido al avance logrado por el aprendizaje de modelos con SVM y los resultados de la OMO sobre estos modelos entrenados, se podría implementar una estrategia de PSO con computación paralela para obtener buenos modelos en comparación a los estudios anteriores, realizando este procedimiento a los modelos entrenados para medir la eficacia y eficiencia de la solución en términos de rendimiento, ejecución y resultados, y posteriormente compararlos entre sí.

Por otro lado, la simplicidad de la PSO y sus directas aplicaciones la han probado como un método eficiente para varios problemas de ciencias e ingeniería, constituyendo un método alternativo a los existentes para modelar la AC. Las principales ventajas de la PSO son:

1. Es un optimizador global libre de derivadas.
2. Rápido y fácil de paralelizar.
3. Necesita pocos parámetros de entrada.

Por último, el enfoque de optimización metaheurística sirve como referencia en la exploración de nuevos métodos para la selección eficiente de mejores modelos de la AC.

3. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

3.1. PROPÓSITOS DE LA SOLUCIÓN

El propósito de la solución es modelar de manera eficiente y no invasiva el fenómeno de la AC en comparación a los métodos existentes, para poder mejorar la oportunidad del diagnóstico de las personas y obtener información valiosa para el campo de investigación de este fenómeno, a través de los resultados de la búsqueda y selección eficientes de los modelos generados a partir del enfoque seleccionado, y que mejor se aproximen a estas señales biológicas.

3.2. CARACTERÍSTICAS DE LA SOLUCIÓN

La solución consiste en realizar un algoritmo PSO con computación paralela utilizando R (un lenguaje y entorno de programación para análisis estadístico) que permita guiar a una mejor búsqueda en el espacio de soluciones de los modelos de AC, y estudiar si es adecuado para resolver el problema de generar estos modelos de la hemodinámica cerebral de buena calidad en comparación con los otros métodos utilizados, y reducir el tiempo de procesamiento.

En primera instancia, se van a reproducir los trabajos de Miranda (2016 y Vallejos (2017) con un ambiente en R, y el conjunto de datos que se utilizará en esta investigación son los utilizados por Miranda (2016) y Vallejos (2017), a los cuales se les aplicarán los procedimientos de los algoritmos PSO con computación paralela desarrollados en R. Las características iniciales el algoritmo en primera instancia serán: inicializar las partículas, actualizar la velocidad y posición de las partículas, evaluar las partículas y obtener la solución óptima. Finalmente, se realizarán las pruebas estadísticas correspondientes.

3.3. ALCANCES Y LIMITACIONES DE LA SOLUCIÓN

Se contempla usar las mediciones de individuos utilizadas en las investigaciones de Miranda (2016) y Vallejos (2017). La línea base para comparar nuevos resultados serán los conseguidos en estos proyectos de titulación. También esta solución se encuentra limitada por el único método de optimización metaheurística que será desarrollado, donde algunas de las deficiencias que tiene la PSO estándar son en algunos casos la velocidad de búsqueda, y la precisión al encontrar óptimos locales debido a su diversidad. Además, el estudio estará fuertemente relacionado a las limitaciones computacionales de desarrollo y ejecución de pruebas, y los alcances y limitaciones de los estudios de la generación de modelos de AC presentados por Miranda (2016) y Vallejos (2017), donde los resultados obtenidos del aprendizaje deben ser filtrados mediante una inspección estándar de medidas fisiológicas, para calcular sus índices *Auto-Regulation Index* (ARI) y *model-free Auto-Regulation Index* (mfARI), y ser evaluados con las técnicas PSO.

3.4. EVALUACIÓN DE LA SOLUCIÓN

Se va a realizar un análisis estadístico para evaluar el desempeño y la eficacia de la selección de los modelos. Comparar la significación y representatividad del resultado de los modelos óptimos con PSO con los obtenidos con los métodos OMO (Vallejos, 2017), y con la función objetivo del estudio de Miranda (2016). Y determinar si existen

configuraciones óptimas e información relevante para la selección de modelos automáticos de AC utilizando el método PSO que cumplan con los objetivos propuestos.

4. OBJETIVO DEL PROYECTO

4.1. OBJETIVO GENERAL

Mejorar el rendimiento y calidad de los modelos del fenómeno de la AC utilizando un algoritmo PSO desarrollado en R, a partir de un conjunto de datos de señales de pacientes, tal que genere y seleccione los mejores modelos de la señal biológica, para evaluar la posibilidad de utilizar estos modelos como representativos del fenómeno de la AC, ayudando a mejorar su diagnóstico en pacientes enfermos.

4.2. OBJETIVOS ESPECÍFICOS

- Configurar el ambiente de desarrollo en R, y reproducir los trabajos de Miranda (2016) y Vallejos (2017).
- Desarrollar algoritmos que apliquen las técnicas de PSO paralelizados sobre el conjunto de datos.
- Aplicar los procedimientos desarrollados sobre los conjuntos de datos utilizados por Miranda (2016) y Vallejos (2017).
- Realizar pruebas estadísticas y evaluar la efectividad de la solución.

5. METODOLOGÍA, HERRAMIENTAS Y AMBIENTE DE DESARROLLO

5.1. METODOLOGÍA PARA USAR

La metodología general para utilizar en esta investigación es la del método científico, ya que los objetivos específicos del proyecto coinciden con sus principios fundamentales, además que se plantea resolver un problema con técnicas de la ingeniería como los métodos de optimización, utilizando procedimientos y herramientas informáticas. Las etapas del método científico que se pretenden utilizar son:

- Planteamiento del problema y objetivos: Sección 1 y 4 del presente informe respectivamente.
- Formulación de la hipótesis: Corresponde a lo que se intenta comprobar en base a la pregunta general que surge de las preguntas de la etapa anterior. Esta propuesta puede validar o invalidar el estudio, y para el caso de este, la hipótesis es la siguiente: *“Los modelos de AC obtenidos por la técnica PSO con computación paralela permiten obtener resultados similares o mejores que los obtenidos de forma semiautomática en los trabajos de Miranda (2016) y Vallejos (2017).”*
- Marco teórico: Donde se explicarán los fundamentos teóricos de la investigación.
- Estrategia del diseño de la solución: Se revisará en la literatura el desarrollo de los algoritmos PSO.
- Diseño de la solución: Se desarrollarán los procedimientos necesarios para alcanzar los objetivos planteados, donde se utilizará el conjunto de datos y los modelos entrenados.
- Verificación de la hipótesis: Se pondrá a prueba la hipótesis planteada en base a los diseños experimentales. Estableciendo una estrategia para asegurar las condiciones necesarias para probar la hipótesis planteada.
- Evaluación de la solución: Se presentarán los resultados y documentación de la investigación realizada.

La metodología del desarrollo del algoritmo o solución se divide en cuatro fases: la primera es la fase de investigación, donde se revisa la bibliografía necesaria obtenida del método científico. Luego la fase de análisis, donde se define el modelo matemático del problema para formular la estructura del algoritmo, sus variables y parámetros del modelo, y determinar la relación de estos. La fase de diseño, donde se desarrolla la arquitectura del procedimiento informático a partir de los requerimientos obtenidos en el análisis. Y la fase de ejecución, donde se realizan las pruebas y

posteriormente se ejecuta el procedimiento con los datos procesados. Estas fases son complementarias entre si y tienen naturaleza cíclica, tal como las metodologías ágiles de desarrollo.

5.2. HERRAMIENTAS DE DESARROLLO

Los datos que se utilizarán inicialmente en este estudio son las mediciones de 45 pacientes sanos, que corresponden a dos conjuntos de datos, el primero de 29 sujetos (Katsogridakis, et. al., 2013) y el segundo 16 sujetos (Mahony, et. al., 2000). Por otro lado, se tienen los datos generados por el *grid* de SVM de Miranda (2016) y utilizado junto con OMO por Vallejos (2017).

También se dispondrá de Notebook Samsung (12GB, 2.6GHz Core i5), *software* administrativo Microsoft Office para crear informes, presentaciones y carta Gantt del proyecto, como herramientas de trabajo para la investigación, y las herramientas disponibles en el laboratorio del Departamento de Ingeniería Informática de la Universidad de Santiago de Chile.

5.3. AMBIENTE DE DESARROLLO

El ambiente de desarrollo está conformado principalmente por el trabajo desde el hogar particular y el realizado en la Universidad de Santiago de Chile. En el ambiente particular se utilizará un Notebook de uso personal configurado con las siguientes tecnologías:

- R y sus paquetes, *software* estadístico para el manejo y análisis de datos.
- R Studio *software*, entorno de desarrollo integrado para el lenguaje de programación R, dedicado a la computación estadística y gráficos.
- Windows 10, sistema operativo para montar el *software* requerido.

Por otro lado, el ambiente de la Universidad se subdivide en dos:

- Departamento de Ingeniería Informática: se utilizará el espacio y las herramientas adecuadas que provee este ambiente, además del Notebook de uso personal mencionado anteriormente.

- Laboratorio del Departamento de Ingeniería Informática: disponible para procesar grandes cantidades de datos y ejecutar las pruebas de la solución.

Las personas que apoyaran este proyecto principalmente son el Dr. José Luís Jara, académico de la facultad y profesor guía de la presente memoria, también se contará con el apoyo y la ayuda de los académicos y funcionarios del Departamento de Ingeniería Informática de la facultad, y además se recibirá información y datos útiles del exalumno Diego Vallejos para la utilización de estos en la presente investigación, junto a los recibidos por el profesor Jara.

6. PLAN DE TRABAJO

A continuación, la Carta Gantt del proyecto, con la dedicación efectiva aproximada en horas por semana, sumando un total de 680 horas a lo largo de 17 semanas. Se considera también que existen sub-metodologías dentro del plan de trabajo, donde pueden existir tareas que se repitan a lo largo del ciclo de vida del proyecto, a medida que se cuente con la retroalimentación necesaria del profesor guía.

Memoria de Titulación	Agosto	Septiembre		Octubre		Noviembre		Diciembre
Tarea	2° - 3° - 4 ° Semana	1° - 2 ° Semana	3° - 4 ° Semana	1° - 2 ° Semana	3° - 4 ° Semana	1° - 2 ° Semana	3° - 4 ° Semana	1° - 2 ° Semana
Escritura de memoria	60 hrs	20 hrs	20 hrs	10 hrs	20 hrs	20 hrs	60 hrs	80 hrs
Corrección memoria								
Especificación de requerimientos	20 hrs	10 hrs						
Investigación del método y preparación del ambiente de desarrollo								
Análisis								
Reproducción de procedimientos y preparativos	40 hrs	50 hrs	20 hrs	20 hrs	10 hrs	10 hrs		
SVM								
Diseño								
Desarrollo del procedimiento en R			40 hrs	10 hrs	10 hrs	10 hrs		
Algoritmos PSO								
Computación paralela								
Implementación								
Procesamiento de datos				40 hrs	20 hrs	20 hrs		
Ejecución del procedimiento								
Testeo								
Pruebas y evaluación del método					10 hrs	10 hrs	10 hrs	
Mantenimiento					10 hrs	10 hrs	10 hrs	
Modificaciones necesarias								

Figura 1: Carta Gantt del proyecto

Fuente: Elaboración propia

7. REFERENCIAS

- Chacón, M., Araya, C., y B. Panerai, R. (2011). *Non-linear multivariate modeling of cerebral hemodynamics with autoregressive Support Vector Machines*. Medical Engineering & Physics, 33, 180-187.
- Czosnyka, M., Smielewski, P., Piechnik, S., Steiner, L., y Pickard, J. (2001). *Cerebral autoregulation following head injury*. Journal of Neurosurgery, 95(5), 756-763.
- Galindo, H., Camaño, P., Fontalvo, M. y col. (2020) *Autorregulación cerebral: fisiología y fisiopatología esenciales para el manejo neurocrítico*. Rev Arg Med 8(4):305-310
- Hernández, T., Irizar, M., Llanes, O., y col. (2013) *Comparación de diferentes algoritmos metaheurísticos en la estimación de parámetros del modelo relacional general de cromatografía líquida en columna*. Ingeniare. Revista chilena de ingeniería, vol. 22 N° 1 (2014) 14-25.
- Jimenez, J., Arango, R. y Jimenez, L. (2016). *Métodos de búsqueda usando los algoritmos de enjambre de partículas y genético*, Lámpsakos, no. 16, pp 52-60.
- Lassen, N. (1959). *Cerebral Blood Flow and oxygen consumption in man*. Physiological Reviews, 183-238.
- MINSAL. (2017, Octubre 2017). Ministerio de salud de Chile. Recuperado en Junio 8, 2022, desde *Ataque Cerebrovascular*: <https://goo.gl/X8ayCc>
- Miranda, R. (2016). *Evaluación de la Autorregulación en Sujetos Normocápnicos e Hipercápnicos usando SVM*. Memoria de Título profesional de Ingeniero Civil en Informática, Universidad de Santiago de Chile, Santiago.
- Muñoz, H. (2009). *Modelamiento no lineal del sistema de autorregulación cerebral con retroalimentación por distensibilidad intracraneal*. Tesis conducente al grado de Ingeniería Civil Informática, Universidad de Santiago de Chile, Santiago.

Panerai, R. (1998). *Assessment of cerebral pressure autoregulation in humans—a review of measurement methods*. *Physiological Measurement*, 19(3), 305–338.

Sharma, S., Patnaik, R., Sharma, N., y Tiwari, J.P. (2011) *Simulated annealing-based particle swarm optimization with adaptive jump strategy for modelling of dynamic cerebral pressure autoregulation mechanism*. *Bio-Inspired Computation*, Vol. 3 (4) 225–237.

Tiecks, F., Lam, A., Aaslid, R., y Newell, D. (1995). *Comparison of Static and Dynamic Cerebral Autoregulation Measurements*. *American Heart Association*, 26(6), 1014-1019.

Vallejos, D. (2017). *Optimización multi-objetivo para seleccionar modelos de señales biológicas*. Memoria de Título profesional de Ingeniero Civil en Informática, Universidad de Santiago de Chile, Santiago.

WHO. (2020, Diciembre). World Health Organization. Recuperado en Junio 8, 2022, desde *The 10 leading causes of death in the world, 2000 and 2019*: <https://goo.gl/m1h1tx>

Xia, X., Gui, L. y Zhan, Z. (2018). *A multi-swarm particle swarm optimization algorithm based on dynamical topology and purposeful detecting*. *Applied Soft Computing* 67 (2018) 126-140.

Zhang, J., Huang, C., Xu, J. y Lu, J. (2013). *A novel particle swarm algorithm for multi-objective optimization problem*. *Modelling, Identification and Control*, Vol. 18 (4) 380–386.

ANEXOS: GLOSARIO

AC	Autorregulación de la presión sanguínea a nivel cerebral.
ARI	Autoregulatory Index. Índice de autorregulación, definido por Aaslid y Tiecks, los cuales estipulan 10 niveles para la medición de la autorregulación, con valores del 0 al 9, siendo 0 ausencia total de autorregulación y 9 autorregulación perfecta. Con la finalidad de generar mayor precisión las curvas de estos valores fueron interpoladas para obtener 91 curvas.
FSC	Flujo sanguíneo cerebral. Corresponde al flujo de sangre en los vasos cerebrales.
mfARI	Model-Free Auto-regulation Index. Índice autorregulatorio, definido por Chacón, Jara y Panerai. Se utiliza para caracterizar numéricamente la autorregulación cerebral.
MSPSO	Multi-Swarm Particle Swarm Optimization, en español optimización por multi enjambre de partículas. Método de optimización que utiliza múltiples enjambres de partículas que cooperan entre si dinámicamente (Xia et. al., 2018).
OCH	Optimización por Colonias de Hormigas, método de optimización metaheurística que utiliza una colonia de hormigas artificial, donde cada hormiga sigue el rastro de feromonas artificiales de las demás para explorar el espacio de búsqueda (Hernández et. al., 2013).
OMO	Optimización Multi-Objetivo. Técnica de los métodos de optimización enfocada en utilizar múltiples funciones objetivo a modo de encontrar un conjunto solución homogéneo que presente rendimientos iguales.
PAM	Presión sanguínea arterial media.
PPC	Presión de perfusión cerebral, corresponde a la diferencia entre la PAM y la PIC (Galindo et. al., 2020).

PSO	Particle Swarm Optimization, en español optimización por enjambre de partículas. Técnica de optimización inspirada en el movimiento de las partículas en la naturaleza, que utiliza metaheurísticas para ampliar el espacio de búsqueda y encontrar soluciones casi óptimas (Jiménez et. al., 2016).
SA	Simulated Annealing, en español recocido simulado, algoritmo de búsqueda metaheurística para problemas de optimización global (Sharma et. al., 2011).
SAPSO	Simulated Annealing-based Particle Swarm Optimization, en español optimización por enjambre de partículas basada en recocido simulado. Algoritmo híbrido que combina SA con PSO, donde SA disminuye las desventajas de PSO (Sharma et. al., 2011).
SVM	Support Vector Machine, en español máquinas de vectores soporte. Corresponde a un algoritmo de aprendizaje automático del tipo supervisado para modelar sistemas, su base se sustenta en la teoría de aprendizaje estadístico, son utilizadas para clasificación (SVC) y regresión (SVR).
VEP	Variaciones espontáneas de presión, corresponde al estado de relajación del sujeto.
VFSC	Velocidad de flujo sanguíneo cerebral.

ANEXOS: REFORMULACIÓN

Antecedentes

Los estudios indican que la mejor opción es utilizar modelos con SVM ya que presentan mejores resultados en términos de precisión y cálculo de la reactividad de los vasos sanguíneos cerebrales a la presión arterial de CO_2 (EtCO_2), además de presentar menor varianza y un muy bajo error cuadrático medio en la línea base de presión (). Es aquí donde aparecen las dificultades, debido a que los algoritmos de entrenamiento se prueban con muchas combinaciones de hiperparámetros para conseguir un buen modelo, generando cientos de miles de modelos para cada individuo, donde se debe escoger al que mejor capture la relación PAM/VFSC (sin estar sobreentrenado) de forma manual ().

Miranda (2016) presentó un entrenamiento de modelos dinámicos con SVM utilizando diferentes configuraciones de parámetros de entrada sobre pacientes en régimen de Normocapnia e Hipercapnia, aplicando a los resultados un nuevo filtro mejorado con respecto al método clásico mencionado anteriormente. El nuevo filtro asigna valores numéricos a comportamientos esperados en la señal de la respuesta autorregulatoria, seleccionando 100 posibles modelos, y escogiéndolo el de más alta correlación.

Alternativamente, Vallejos (2017) propone un método de optimización multi objetivo que evalúa la información disponible para seleccionar un modelo representativo para los individuos del estudio, con el fin de proponer guías que permitan reducir el espacio de búsqueda.

Todavía existen varios enfoques que se podrían aplicar a este problema, los más prometedores son la continuación de la utilización de técnicas de optimización para obtener una serie de modelos óptimos y poder realizar un análisis de los factores de cada modelo, donde si los resultados resultan significativos, podrían acotar inicialmente el espacio muestral sobre el que se realiza el estudio de la autorregulación, obteniendo resultados igual de precisos de forma más rápida y eficiente.

Índices de autorregulación cerebral

De acuerdo con los estudios, se han definido índices para evaluar la autorregulación, los más relevantes para este estudio son el ARI (modelando de una ecuación diferencial de segundo orden) y el mfARI (sin modelo específico). Entre estos dos sistemas de medida se ha demostrado que mfARI presenta mayor robustez que ARI. Para el caso de las VEP se requiere modelar primero y luego medir los índices, y también según estudios, si se entrena un modelo con las VEP y se le pasa un escalón a ese modelo se pueden encontrar resultados válidos para el fenómeno.

Estructura de los modelos SVM

Se utilizarán 4 modelos de entrenamiento multivariado.

- Finite Impulse Response (FIR)
- Non-Linear Finite Impulse Response (NFIR)
- Auto-Regression with Exogenous Variables (ARX)
- Non-Linear Auto-Regression with Exogenous Variables (NARX)

Se entrenarán estos modelos utilizando SVM. Las entradas utilizadas en los modelos son la PAM y etCO₂, y la salida del modelo corresponde a la VFSC estimada. Para los modelos autorregresivos se utiliza también la VFSC como variable de entrada. Para la validación de los modelos se utilizará validación cruzada de la señal presente en los datos. Para la selección de parámetros es posible obtener variados resultados en base a los hiperparámetros escogidos para los entrenamientos de las SVM. Los modelos lineales ocupan los hiperparámetros ν y costo (C). Los modelos no lineales utilizan un Kernel radial por lo que agregan el hiperparámetro sigma (σ).

Posibles valores iniciales paramétricos de las SVM

Estos valores se irán ajustando de acuerdo avanza la investigación, pero en general los siguientes fueron utilizados en el estudio de Vallejos (2017).

Parámetro	Valor de parámetro en FIR / NFIR	Valor de parámetro en ARX / NARX
Retardos PAM	1,2,3,...,8	1,2,3,...,8
Retardos EtCO ₂	1,2,3,...,6	1,2,3,...,6
Retardos VFSC	1,2,3,...,6	1,2,3,...,6
Hiperparámetro C (Exponencial)	0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096	0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096
Hiperparámetro C (Lineal)	0.25, 292.8, 585.36, 877.91, 1170.46, 1463.02, 1755.57, 2048.13, 2340.68, 2633.23, 2925.79, 3218.34, 3510.89, 3803.45, 4096	0.25, 292.8, 585.36, 877.91, 1170.46, 1463.02, 1755.57, 2048.13, 2340.68, 2633.23, 2925.79, 3218.34, 3510.89, 3803.45, 4096
Hiperparámetro nu	0.1, 0.2, 0.3, ..., 0.9	0.1, 0.2, 0.3, ..., 0.9
Hiperparámetro σ	2 ⁻⁴ , 2 ⁻² , 2 ⁰ , ..., 2 ¹⁰	2 ⁻⁴ , 2 ⁻² , 2 ⁰ , ..., 2 ¹⁰

Tabla 2: Parámetros SVM

Fuente: Vallejos (2017)

Solución propuesta reformulada

Se propone implementar un nuevo método de optimización utilizando metaheurísticas para una mayor eficiencia y selección automática de los mejores modelos, considerando inicialmente diferentes criterios de aceptación como calidad predictiva (precisión de los índices obtenidos respecto a los índices reales obtenidos de los sujetos), parsimonia (simplicidad de los modelos), y verosimilitud fisiológica (que la respuesta al escalón se presente dentro de los rangos normales del comportamiento humano).

El propósito de este estudio es proponer un modelo de optimización por enjambre de partículas en computación paralela utilizando el algoritmo “PSO” o “metaheuristicOpt-PSO” del paquete CRAN, en lenguaje R, a partir de las señales entrenadas con máquinas de vectores de soporte para evaluar tales señales de autorregulación cerebral, y comparar los resultados con los métodos propuestos por el filtro avanzado de Miranda (2016) y la optimización multi objetivo de Vallejos (2017).

Los sujetos, mediciones y preprocesamiento de datos serán los mismos utilizados por el estudio de Vallejos (2017). El método de paralelización

Principios de la PSO estándar

En PSO, una cantidad de partículas con forma de enjambre vuelan a través del espacio de búsqueda para encontrar soluciones óptimas o casi óptimas. Cada coordenada de la partícula representa una posible solución al problema de optimización. El movimiento de la partícula depende de su mejor posición y la mejor posición de la vecindad de las partículas del enjambre.

Descripción matemática de PSO

Suponiendo que $x_i = (x_{i1}, x_{i2}, \dots, x_{id})$ es la posición actual de la i-ésima partícula en el espacio de búsqueda d-dimensional.

$P_{ibest} = (p_{i1}, p_{i2}, \dots, p_{id})$ la mejor posición encontrada por la i-ésima partícula

$G_{best} = (g_1, g_2, \dots, g_d)$ la mejor posición encontrada por el enjambre

En cada iteración, la velocidad y posición (coordenadas) de la partícula se actualiza de acuerdo con las siguientes ecuaciones:

$$v_i^{t+1} = wv_i^t + c_1r_1(p_{ibest} - x_i) + c_2r_2(g_{best} - x_i) \quad (1)$$

$$x_i^{t+1} = x_i^t + av_i^{t+1} \quad (2)$$

Donde $i = 1$ a n , corresponde al número de partículas en el enjambre, $t =$ corresponde al número de iteración, $v_i = (v_{i1}, v_{i2}, \dots, v_{id})$ vector velocidad de la partícula,

w = peso inercial, $c1$ = constante cognitiva, $c2$ = coeficiente social, $r1$, $r2$ números aleatorios entre 0 y 1, y a = coeficiente de velocidad escalar.

Algoritmo PSO estándar

Paso 1: Inicializar población con partículas generadas aleatoriamente. Velocidad inicial de cada partícula inicializada en 0. Inicializar $c1$, $c2$ y w .

Paso 2: La adaptación de cada partícula es calculada utilizando la función objetivo.

Paso 3: En base al paso 2, $gbest$ y $pbest$ son seleccionados para el enjambre y para cada partícula respectivamente.

Paso 4: La velocidad de la partícula para la próxima iteración es calculada utilizando la ecuación (1)

Paso 5: Se actualiza la posición de la partícula utilizando la ecuación (2)

Paso 6: Evaluar criterio de parada (si el máximo número de iteraciones es alcanzado) si $t = itermax$ parar, si no ir al paso 2, con los valores actualizados de los contadores de la iteración ($t = t + 1$)

Paso 7: El valor final de $gbest$ es la solución

Parámetros PSO

Peso inercial (w)

Describe la confianza de la partícula en movimiento, representando el impacto de la velocidad anterior. Diferentes estudios mantienen inicialmente alto este valor para apoyar la exploración, luego es disminuido para apoyar la explotación, implementándolo como:

$$w = w_{\max} - \left(\frac{w_{\max} - w_{\min}}{iter_{\max}} \right) \times t$$

Donde w_{\max} es el máximo valor del peso inercial = 0.9, w_{\min} es el mínimo valor del peso inercial = 0.4, $iter_{\max}$ es el máximo número de iteraciones y t es la iteración actual.

Constantes cognitiva y social (c_1 y c_2)

También llamadas constantes de aceleración, mantienen el balance entre la partícula y el enjambre. Por lo general se utiliza 2 como valor, mientras que en algunos estudios se calculan con las siguientes ecuaciones:

$$c_1 = 2.5 - 2.0 \times \frac{t}{iter_{\max}}$$

$$c_2 = 0.5 + 2.0 \times \frac{t}{iter_{\max}}$$

Tamaño del enjambre

Mientras más grande este valor, mayor es la habilidad de exploración en el espacio de búsqueda, pero aumenta el tiempo de computación, debido a este compromiso, el tamaño del enjambre ideal es de 20 a 30 partículas, suficiente para resolver la mayoría de los problemas.

Metodología propuesta

Se propone una optimización por enjambre de partículas con computación paralelizada en R, es decir, para cada función objetivo del problema a optimizar, se realizará una PSO para la selección de los mejores modelos de autorregulación. También se evaluará la posibilidad de adaptar la metodología de desarrollo al modelo CRISP-DM de Wirth and Hipp (2000).

Criterios de optimización

1. Maximizar Coeficiente de correlación (CC)

Mide el grado de relación entre dos señales y se define mediante la siguiente ecuación:

$$CC = \frac{\sum_i (v(i) - \bar{v}) * (\hat{v}(i) - \bar{\hat{v}})}{\sqrt{\sum_i (v(i) - \bar{v})^2} * \sqrt{\sum_i (\hat{v}(i) - \bar{\hat{v}})^2}}$$

Los valores de CC varían entre [-1, 1], donde -1 representa una correlación inversa, 0 que no existe correlación, y 1 que las señales poseen el mismo comportamiento.

2. Minimizar Error cuadrático medio normalizado ajustado por amplitud (ECMNa)

Otra forma de evaluar las diferencias entre el valor real y el estimado. Se utiliza para identificar el error generado en la estimación de la VFSC y su valor real sin considerar su componente continua. Su cálculo es:

$$ECMN_a = \frac{\sum_{i=1}^N (v_{scc}(i) - K * \hat{v}_{scc}(i))^2}{N * Var(v_{scc})}$$

$$K = \frac{\sum_{i=1}^N v_{scc}(i) * \hat{v}_{scc}(i)}{\hat{v}_{scc}(i)^2}$$

Con k un factor de normalización, N el largo de las señales, v_scc es la señal real de VFSC, v'_scc la señal estimada de VFSC (ambas VFSC sin su componente continua) y Var la varianza. El resultado de esta ecuación posee un comportamiento similar al cálculo de $1 - CC^2$.

3. Maximizar Coeficiente de variación (CV)

Mide el grado de homogeneidad, un valor pequeño indica homogeneidad de la variable. Este valor es usado en porcentaje y su cálculo es:

$$CV = \frac{\sigma}{\bar{x}} * 100$$

Con los estadísticos desviación estándar y media aritmética.

4. Maximizar Puntuación de escalones

La respuesta al escalón de presión es una medida de evaluación de los modelos que verifica si han capturado el fenómeno de la autorregulación. La puntuación de escalones es una metodología para obtener un puntaje para las señales, para obtener un ranking de las respuestas. La metodología utiliza un rango de [0, 10] para medir su puntaje, siendo el valor 10 las respuestas con características excelentes y 0 las respuestas con características deficientes. El procedimiento consiste en que todas las respuestas

comienzan con un puntaje de 10 y a medida que pasan por tres criterios van recibiendo penalizaciones acumuladas, para luego obtener el puntaje final de la respuesta:

$$Penalización = \sum_{i=1}^3 C_i$$
$$Score = 10 - Penalización$$

Los criterios son:

Criterio 1 (C1): Se verifica que la respuesta en su tramo de estabilización (15 a 30 [s]) no sea una señal creciente o decreciente, en el caso de que no se cumpla se le aplica una penalización acorde a la distancia recorrida desde el peak máximo al último punto final de la señal. En general las pendientes tienden a valores pequeños (rango de las centésimas), por lo que la penalización es ponderada por 100, con esto se logra suavizar el puntaje.

Criterio 2 (C2): Se verifica que la caída que está antes de la estabilización sea como máximo un 45% del tramo de subida de la señal. El valor fue establecido por las 91 curvas de Aaslid-Tiecks Decimal, ya que la señal con ARI =9 cae hasta un 45%, con respecto al tramo de subida. Si la respuesta no cumple con ser menor o igual a 45%, es penalizada utilizando un factor con valor 10.

Criterio 3 (C3): Se verifica que la estabilización de las señales ocurra al menos al mismo nivel del peak de la caída. Ya que basándose en las 91 curvas de Aaslid-Tiecks Decimal, se aprecia que ninguna respuesta se estabiliza bajo este valor, por lo tanto, la penalización adquiere un rol fundamental en este paso y su valor es calculado a través de la distancia entre puntos y se pondera por 10.

Posibles objetivos específicos adicionales

- Automatizar el proceso de descarte de escalones
- Aumentar alcance de PSO utilizando SA

Criterios de evaluación

Están los estadísticos, como el coeficiente de correlación donde principalmente se evalúa la similitud entre la señal predicha (en base a la señal de entrenamiento) y sección contigua de validación. Además de las pruebas para contrastar la normalidad e independencia de los conjuntos de datos.

Respuesta al escalón de presión: Para evaluar si efectivamente se obtuvieron resultados que reflejen el fenómeno de la autorregulación es necesario realizar un análisis sobre los modelos entrenados. Este análisis parte de una base simple la cual es que al realizar variaciones en la PAM se observan reacciones por parte de la VFSC. Esto quedó demostrado en los estudios realizados previamente. También están el filtro avanzado propuesto por Vallejos (2017), los índices autorregulatorios (descritos más arriba), y análisis de varianza.