**Optimización Multi-Objetivo por Enjambre de Partículas para selección automática de modelos de autorregulación cerebral**

**Informe N° 3**

**Propuesta final – versión interna del curso**

**UNIVERSIDAD DE SANTIAGO DE CHILE**

**FACULTAD DE INGENIERÍA**

**DEPARTAMENTO DE INGENIERÍA INFORMÁTICA**

Nombre: Benjamin Jorquera Jorquera

Rut: 19.182.719-8

Año estimado de egreso: 2022

Teléfono: 9 93362209

E-mail: [benjamin.jorquera@usach.cl](mailto:benjamin.jorquera@usach.cl)

Profesor Patrocinador: José Luis Jara

Fecha: 25 de Mayo de 2022

**RESUMEN**

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

**TABLA DE CONTENIDOS**

[1. DESCRIPCIÓN DEL PROBLEMA 1](#_Toc105030077)

[1.1. MOTIVACIÓN 1](#_Toc105030078)

[1.2. ENUNCIADO DEL PROBLEMA 2](#_Toc105030079)

[2. ANÁLISIS DE LA SOLUCIÓN 2](#_Toc105030080)

[2.1. ESTADO DEL ARTE 2](#_Toc105030081)

[2.2. ENFOQUES DE SOLUCIÓN 5](#_Toc105030082)

[**2.2.1.** **Optimización por Colonias de Hormigas** 5](#_Toc105030083)

[**2.2.2.** **Optimización por Enjambre de Partículas** 5](#_Toc105030084)

[**2.2.3.** **Variaciones de Optimización por Enjambre de Partículas** 6](#_Toc105030085)

[**2.2.4.** **Optimización Multi-Objetivo por Enjambre de Partículas** 7](#_Toc105030086)

[2.3. JUSTIFICACIÓN DEL ENFOQUE SELECCIONADO 7](#_Toc105030087)

[3. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA 8](#_Toc105030088)

[3.1. PROPÓSITOS DE LA SOLUCIÓN 8](#_Toc105030089)

[3.2. CARACTERÍSTICAS DE LA SOLUCIÓN 8](#_Toc105030090)

[3.3. ALCANCES Y LIMITACIONES DE LA SOLUCIÓN 9](#_Toc105030091)

[3.4. EVALUACIÓN DE LA SOLUCIÓN 10](#_Toc105030092)

[4. OBJETIVO DEL PROYECTO 10](#_Toc105030093)

[4.1. OBJETIVO GENERAL 10](#_Toc105030094)

[4.2. OBJETIVOS ESPECÍFICOS 10](#_Toc105030095)

[5. METODOLOGÍA, HERRAMIENTAS Y AMBIENTE DE DESARROLLO 11](#_Toc105030096)

[5.1. METODOLOGÍA PARA USAR 11](#_Toc105030097)

[5.2. HERRAMIENTAS DE DESARROLLO 12](#_Toc105030098)

[5.3. AMBIENTE DE DESARROLLO 12](#_Toc105030099)

[6. PLAN DE TRABAJO 13](#_Toc105030100)

[7. REFERENCIAS 14](#_Toc105030101)

# DESCRIPCIÓN DEL PROBLEMA

## MOTIVACIÓN

El Sistema de Autorregulación Cerebral (SAC) 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 sus actividades y también en situaciones críticas. Así que un mal funcionamiento o deterioro del SAC 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, 2016) y la segunda a nivel mundial (WHO, 2014). Por lo tanto, es importante realizar evaluaciones de la Autorregulación Cerebral (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 de este fenómeno realizados por el Departamento de Ingeniería Informática en la Universidad de Santiago de Chile corresponden a modelar el SAC 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 optimizadas, 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 del fenómeno de la AC. Además, aportar información valiosa sobre estos métodos de optimización para el problema planteado.

## ENUNCIADO DEL PROBLEMA

¿Cómo mejorar la búsqueda y selección de modelos de calidad de Autorregulación Cerebral utilizando métodos de optimización metaheurística, para diagnosticar a tiempo posibles afecciones cerebrales en las personas de manera no invasiva?

# ANÁLISIS DE LA SOLUCIÓN

## ESTADO DEL ARTE

Se tiene constancia de que los primeros estudios del SAC 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). Pero este método tenía la desventaja de utilizar drogas para inducir los cambios bruscos de presión necesarios para cubrir las mediciones, provocando molestias y riesgos a los pacientes (Panerai, 1998). Con el paso del tiempo, se empezó a utilizar el Ultrasonido Doppler Transcraneal, que utilizaba un enfoque no invasivo para la toma de muestras.

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 y los manguitos.

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. También 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.

Sin embargo, estos métodos de cambios espontáneos de presión también 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 otro método (Miranda, 2016).

Por otro lado, en 2003 se modeló el SAC utilizando este enfoque de VEP y un modelo de arquitectura de Redes Neuronales Recurrentes (Recurrent Neural Network, RNN), que obtuvo buenos resultados y se demostró la no linealidad del SAC bajo VEP, pero los modelos no lograron buenos coeficientes de correlación (Blanco, 2003).

Actualmente el estudio de la AC consiste en medir la presión sanguínea arterial media (PAM) y la velocidad de flujo sanguíneo cerebral (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 optimas 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).

## ENFOQUES DE SOLUCIÓN

### **Optimización por Colonias de Hormigas**

La Optimización por Colonias de Hormigas (OCH) se inspira en el comportamiento y comunicación de las hormigas. Análogamente 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. Inicialmente los métodos OCH trataban con problemas combinatorios, luego fueron aplicados a problemas con variables continuas. El rastro de feromonas es modelado como una cantidad de feromona dejada en el camino, pudiendo ser alcanzado con una función de distribución probabilística normal.

### **Optimización por Enjambre de Partículas**

El método de PSO 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. Desde su propia posición.
2. Desde 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:
   1. La velocidad de la partícula se hace cero después de algunas iteraciones.
   2. Pérdida de diversidad.
2. Sensibilidad a parámetros de entrada y tamaño de la población.

### **Variaciones de Optimización por Enjambre de Partículas**

Optimización por Enjambre de Partículas basado en Recocido Simulado

El algoritmo de Optimización por Enjambre de Partículas basado en Recocido Simulado (*Simulated Annealing-based Particle Swarm Optimization*, SAPSO) 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.

Optimización por Multi Enjambre de Partículas

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). Donde 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, presenta rendimientos moderados en utilización de tiempo.

### **Optimización Multi-Objetivo por Enjambre de Partículas**

Varios grupos de investigadores de distintos países han realizado estudios relacionados a la Optimización Multi-Objetivo por Enjambre de Partículas (Multi-Objetive Particle Swarm Optimization, MOPSO), utilizando distintas técnicas de optimización para resolver problemas. En uno de estos estudios se introdujo un peso dinámico inercial en la PSO estándar para mejorar las habilidades de explotación y exploración del algoritmo, utilizando esta estrategia el algoritmo MOPSO puede encontrar soluciones diversas y convergentes. Los resultados de este estudio demuestran que ese enfoque genera aproximaciones satisfactorias al frente de Pareto con soluciones distribuidas de manera uniforme (Zhang et al., 2013).

## JUSTIFICACIÓN DEL ENFOQUE SELECCIONADO

El enfoque seleccionado es la MOPSO, 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 junto con PSO para obtener buenos modelos en comparación a los estudios ya realizados, y paralelamente realizar este procedimiento a los modelos no 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.

Además, la OMO es un tema de investigación importante para científicos e ingenieros. En las últimas décadas varios estudios se han enfocado en este tipo de optimización, logrando varios descubrimientos. Así, se comenzaron a utilizar algoritmos evolucionarios para la OMO, dando como resultado una gran variedad de nuevos algoritmos, siendo PSO una de las técnicas de computación evolucionarias más significativas (Zhang et al., 2013).

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.

# DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

## PROPÓSITOS DE LA SOLUCIÓN

El propósito de la solución es poder mejorar el diagnóstico de la AC 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.

## CARACTERÍSTICAS DE LA SOLUCIÓN

La solución consiste en realizar un algoritmo MOPSO 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, 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 de datos.

El conjunto de datos que se utilizará en esta investigación corresponde a las mediciones de individuos utilizadas en la investigación de Miranda (2016), donde, por un lado, se utilizará el algoritmo MOPSO sobre estos datos para obtener los mejores modelos, y por otro, los modelos preprocesados y seleccionados serán entrenados utilizando SVM sobre estructuras lineales y no lineales, para que luego se apliquen diversos filtros al resultado de estos entrenamientos para obtener los índices de autorregulación y seleccionar con un algoritmo los mejores modelos de acuerdo con diversos factores presentados por Vallejos (2017). Luego el algoritmo MOPSO se encargará de guiar la búsqueda de los mejores modelos, definiendo una función multiobjetivo para la optimización y las reglas matemáticas asociadas.

## ALCANCES Y LIMITACIONES DE LA SOLUCIÓN

Los alcances de la solución serán los resultados obtenidos de la muestra de datos de los pacientes utilizando el algoritmo MOPSO, estos resultados corresponderán a modelos de AC los cuales podrían modelar una buena aproximación al comportamiento real del fenómeno. Otro alcance puede ser el impacto que tendrá esta investigación al área del modelamiento de la AC con respecto a los resultados, la información y las técnicas utilizadas. Y si se cumple el objetivo, se podrán obtener resultados eficientes de forma más rápida, mejorando y acelerando los procesos de diagnóstico, monitoreo y control de los pacientes con problemas cerebrales.

Por otro lado, la solución se encontrará limitada por el tamaño y tipo de las muestras utilizadas para el estudio, donde los datos están sujetos al historial clínico de los pacientes y estos deben dar su consentimiento voluntario para ser parte del estudio, el cual debe ser aprobado por los comités de ética correspondientes. 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 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 MOPSO.

## 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 MOPSO con los obtenidos con los métodos OMO (Vallejos, 2017), y con la función objetivo del estudio de Miranda (2016). Determinar si existen configuraciones óptimas e información relevante para la selección de modelos automáticos de AC utilizando el método MOPSO que cumplan con los objetivos propuestos.

# OBJETIVO DEL PROYECTO

## OBJETIVO GENERAL

Mejorar el rendimiento y calidad de los modelos del fenómeno de la AC utilizando un algoritmo MOPSO 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.

## OBJETIVOS ESPECÍFICOS

* Configurar el ambiente de desarrollo en software R, cargar los datos y procesarlos.
* Desarrollar algoritmo en software R que aplique las técnicas de MOPSO sobre el conjunto de datos de muestra.
* Utilizar los procedimientos y datos generados por la investigación de Vallejos (2017) para actualizar el modelo de optimización a MOPSO.
* Ejecutar los procedimientos desarrollados sobre los conjuntos de datos en el laboratorio de Ingeniería Informática de la Universidad de Santiago de Chile.
* Realizar test de hipótesis y analizar con pruebas estadísticas los resultados y evaluar la efectividad de la solución.

# METODOLOGÍA, HERRAMIENTAS Y AMBIENTE DE DESARROLLO

## METODOLOGÍA PARA USAR

La metodología 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 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 MOPSO son adecuados para representar de manera objetiva el fenómeno en cuestión y ser utilizados para predecir posibles problemas cerebrales en las personas*.”
* 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 MOPSO.
* 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 reproducir el fenómeno y probar la hipótesis planteada.
* Evaluación de la solución: Se presentarán los resultados, y se realizarán análisis estadísticos y documentación de la investigación realizada.

## HERRAMIENTAS DE DESARROLLO

Los datos que se utilizaran 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), y *software* administrativo Microsoft Office para crear informes, presentaciones y carta Gantt del proyecto, como herramientas de trabajo para la investigación

## AMBIENTE DE DESARROLLO

El ambiente de desarrollo está conformado por:

* 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.

# 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.



Figura 1: Carta Gantt del proyecto

Fuente: Elaboración propia

# REFERENCIAS

**ANEXOS: GLOSARIO**