

*Málaga, junio de 2011*

## Informe Ejecutivo

**TÍTULO:** BIO-3.0-2011: Parallel Multi-Swarm for Gene Selection in DNA Microarrays  
**RESUMEN:** En este informe se presenta el trabajo realizado en el entorno de DIRICOM referente a un algoritmo paralelo multi-swarm para la selección y clasificación de genes en Microarrays de ADN. Se trata de una propuesta en la que múltiples algoritmos de particle swarm se ejecutan en islas independientes y colaboran intercambiando las mejores soluciones (subconjuntos de genes) entre ellas. Los experimentos se realizaron con grandes bases de datos de expresiones genéticas bien conocidas en la literatura: Leukemia, Colon, Lymphoma y Lung.

**OBJETIVOS:**

1. Estrategia de multi-swarm paralela para la resolución eficiente del problema.
2. Experimentación y análisis respecto al conte computacional.
3. Experimentación y análisis respecto a la calidad de la solución.

**CONCLUSIONES:**

1. Los resultados obtenidos serán puestos a disposición de los especialistas con el objetivo de que se trabaje con genes más manejables pero con el mismo poder de clasificación, descartando la información genética errónea y redundante.  
Este problema se resolverá en el ámbito de DIRICOM empleando una combinación de técnicas de optimización (metaheurísticas) y clasificadores de máquinas de soporte (SVM).
2. La propuesta presenta importantes ganancias en cuanto al coste computacional (con una eficiencia del 85%) y en cuanto a la precisión de los subconjuntos de genes seleccionados a la hora de clasificar.

**RELACIÓN CON ENTREGABLES:**

- PRE: BIO-2.0-2010 (anterior o necesario de leer)  
PRE: BIO-1.0-2009 (anterior o necesario de leer)  
CO: PSO-1.0-2008 (anterior o necesario de leer)

*Málaga, June, 2011*

## Executive Summary

**TITLE:** BIO-3.0-2011 Parallel Multi-Swarm for Gene Selection in DNA Microarrays

**ABSTRACT:** In this report, a parallel Particle Swarm Optimization (PSO) is used for gene selection of high dimensional Microarray datasets. The proposed algorithm, consists of running a set of independent PSOs following an island model. A feature selection mechanism is embedded in each subalgorithm for finding small samples of informative genes amongst thousands of them. PMSO has been experimentally assessed with different population structures on four well-known cancer datasets: Leukemia, Colon, Lymphoma y Lung.

**GOALS:**

1. Defining the parallel optimization strategy.
2. Experimental analysis concerning the computational cost.
3. Experimental analysis concerning the subset accuracy.

**CONCLUSIONS:**

1. Our parallel approach is able to improve sequential algorithms in terms of computational time/effort (Efficiency of 85%), as well as in terms of accuracy rate, identifying specific genes that our work suggests as significant ones for an accurate classification.
2. Additional comparisons with several recent state the of art methods also show competitive results with improvements of over 100% in the classification rate and very few genes per subset.

**RELATION WITH  
DELIVERABLES:**

PRE: BIO-2.0-2010 (mandatory reading)

PRE: BIO-1.0-2009 (mandatory reading)

CO: PSO-1.0-2008 (advisable reading)

# Parallel Multi-Swarm for Gene Selection in DNA Microarrays

DIRICOM

June 2011

## 1. Referencia del trabajo

J. García-Nieto and E. Alba, **Parallel Multi-Swarm Optimizer for Gene Selection in DNA Microarrays**. *Applied Intelligence*, In Press, Online first, 2011 URL <http://www.springerlink.com/content/m7w68182911p2233/> JCR impact f. = 0.893pos 75/108 in CS-AI (Q3). ISSN 0924-669X

### 1.1. Paper Abstract

The execution of many computational steps per time unit typical of parallel computers offers an important benefit in reducing the computing time in real world applications. In this work, a parallel Particle Swarm Optimization (PSO) is used for gene selection of high dimensional Microarray datasets. The proposed algorithm, called PMSO, consists of running a set of independent PSOs following an island model, where a migration policy exchanges solutions with a certain frequency. A feature selection mechanism is embedded in each subalgorithm for finding small samples of informative genes amongst thousands of them. PMSO has been experimentally assessed with different population structures on four well-known cancer datasets. The contributions are twofold: our parallel approach is able to improve sequential algorithms in terms of computational time/effort (Efficiency of 85%), as well as in terms of accuracy rate, identifying specific genes that our work suggests as significant ones for an accurate classification. Additional comparisons with several recent state the of art methods also show competitive results with improvements of over 100% in the classification rate and very few genes per subset.

## 2. Citas en la actualidad

Yinghuan Shi, Yang Gao, Ruili Wang, Ying Zhang and Dong Wang. **Transductive cost-sensitive lung cancer image classification**. *Applied Intelligence*. Online first, URL <http://dx.doi.org/10.1007/s10489-012-0354-z> JCR impact f. = 0.893pos 75/108 in CS-AI (Q3). ISSN 0924-669X