

INTELIGENCIA ARTIFICIAL PARA EL DIAGNÓSTICO MOLECULAR DE ENFERMEDADES EN MEDICINA GENÓMICA.

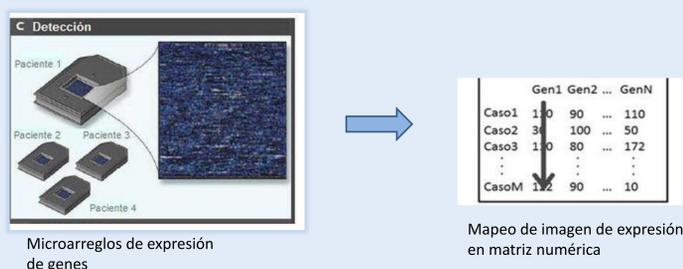


Edna Márquez, Valeria Barrón, Ana Espinosa

Hospital General de México "Eduardo Liceaga", Unidad de Medicina Genómica

INTRODUCCIÓN

Uno de los métodos de diagnóstico molecular que se lleva a cabo en el Hospital General de México, en la Unidad de Medicina Genómica, es a través de los microarreglos de RNA. Los microchips contienen la expresión de miles de genes en situación específica. La lectura de múltiples microarreglos proporciona una matriz numérica que representa la expresión de miles de genes. En este caso presentamos su aplicación para dos padecimientos de alta incidencia en la población mexicana y corresponden a problemas de salud pública actuales, el cáncer cervical (CaCu) y la diabetes mellitus tipo 2 (DM2).



El objetivo de este trabajo es la aplicación de métodos de Inteligencia Artificial para poder realizar el diagnóstico basado en la expresión molecular de los genes, como una alternativa al diagnóstico actual que en el caso del cáncer es por histopatología y en la diabetes es hasta que se presentan los síntomas y con pruebas de laboratorio.

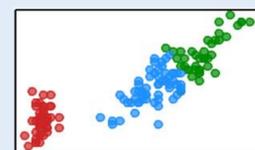
Los métodos de Inteligencia Artificial utilizados corresponden a la parte de Machine Learning (ML), que permite hacer clasificación y predicción basándose en información numérica, dichos métodos se ha probado en diversas áreas de aplicación como muy confiables. Identificar genes y poder diferenciar entre grupos de muestras a través de la expresión de genes es un objetivo de la medicina molecular para el diagnóstico médico y útil en la predicción de respuesta al tratamiento. Es una forma de explicar porque algunos casos que parecen ser iguales responden diferente a los tratamientos.

MATERIALES Y MÉTODOS

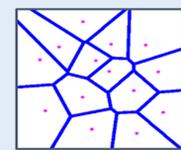
Se utilizaron 42 muestras de tumores y 12 de controles, para CaCu y en pruebas con diabetes 425 muestras de casos y 457 controles. En ambos estudios se aplicaron métodos de preprocesamiento para su normalización y filtrado de genes. En Inteligencia Artificial el ML se utiliza para encontrar nuevas formas de resolver problemas sin que el sistema tenga programada la forma exacta de encontrar la solución, lo que hace es descubrir cómo encontrar la solución, puede ser identificando patrones representativos de los grupos o posibles soluciones.

Los métodos de ML se utilizan para dos objetivos básicos con los microarreglos de genes: 1) para crear grupos de genes y 2) para hacer clasificación de muestras.

En el caso de selección de genes se hace por medio de aprendizaje no supervisado, dejando libremente que el modelo de aprendizaje descubra cómo agrupar los genes.



Grupos o clusters visualizados a través de análisis de componentes principales



Grupos o clusters formados como diagrama de Voronoi con cuantización vectorial.

Mientras que en la clasificación de muestras el objetivo es la separación de muestras, casos vs. controles, o entre las muestras del padecimiento diferenciando entre los diferentes tipos de casos. De los métodos de ML utilizados están las máquinas de soporte vectorial, c-medias, mapas auto-organizados y cuantización vectorial. También utilizamos análisis de componentes principales tanto para la separación de muestras como para la visualización de resultados.

Estos métodos de aprendizaje se programaron con C, java, Python y R.

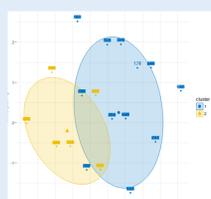
RESULTADOS

En el primer experimento se utilizaron los 42 tumores de cancer cervical había tanto de tipo escamoso como de adenocarcinoma, contra los 12 controles. Con el método de Mapas auto-organizados se obtuvo un grupo de 9 genes utilizando las muestras de cáncer cervical, en tabla 1 se muestra su relación con los procesos biológicos de acuerdo con el programa David (Database for annotation, visualization and integrated discovery).

Proceso biológico	% genes
Ciclo celular	39.8
Regulación biológica	38.8
Regulación celular	38.8
Replicación de DNA	20

Tabla 1. Procesos biológicos en los que participan 9 genes agrupados por ML para CaCu.

Con las muestras de expresión de genes de DM2 se presenta en table 2 la clasificación de muestras de casos vs. controles utilizando un grupo de genes creado también mediante ML.



Clusters de muestras de DM2 vs. controles creados con c-means.

Tabla 2. Clasificación Diabetes vs Controles			
Método	Correctos	Especificidad	Sensibilidad
MSV	0.865	0.83204	0.89473

Tabla 2. Con máquinas de soporte vectorial se logró una alta especificidad y sensibilidad en la clasificación de muestras.

CONCLUSIÓN

Los métodos de Inteligencia Artificial son una alternativa para crear grupos de genes a partir de su expresión y también para la diferenciación de muestras. La forma de clasificación de muestras por los métodos de ML es completamente objetiva al basarse en la matriz numérica, y muy confiable en cuanto se ha logrado obtener el modelo apropiado para ello.

Los métodos de máquinas de soporte vectorial y c-means obtuvieron buenos resultados para DM2, la sensibilidad de 86% y especificidad de 89% con MSV.

En CaCu se encontraron 9 genes sobresalientes en un mismo grupo (CNB2, CDKN3, NUSAP1, PRC1, SMC4, SYCP2, TYMS, CDKN2A, PCNA).

Una forma de comprobar si los grupos de genes creados son los adecuados es a través de bases de datos para verificar su función o procesos en los que.

Utilizando ML se verifica también la capacidad de los diferentes grupos de genes para separar correctamente las muestras de casos y controles, lo que se puede apreciar con los valores de especificidad y sensibilidad.