

Modelo Computacional para la identificación de endofenotipos y clasificación de pacientes con Artritis Reumatoide a partir de datos genéticos y clínicos utilizando técnicas de Inteligencia Computacional

Luis Morales Muñoz

Estudiante Maestría en Ingeniería Biomédica Facultad de Medicina, Universidad Nacional, Bogotá D.C. Colombia, lamoralesmu@unal.edu.co

Luis Fernando Niño Phd.

Facultad de Ingeniería Industrial y Sistemas Universidad Nacional, Bogotá D.C. Colombia, lfninov@unal.edu.co

Gerardo Quintana Md. MS

Facultad de Medicina Departamento de Medicina Interna Universidad Nacional, Bogotá D.C. Colombia

ABSTRACT

Rheumatoid arthritis (RA) is an autoimmune disease that affects 1% to 3% of the human population, there are genetic markers involved in this disease, we used computational intelligence models for classification and identification of endophenotypes (relationship between phenotype and markers genetic) in patients with rheumatoid arthritis and healthy controls, this genetic information is the HLA DRB1 (Human Leukocyte Antigen) and Shared Epitope theory, which refers to the association between RA and HLA-DRB1, alleles mainly containing a common motif amino acid sequences QKRAA, RRRAA QRRRAA or in the HLA-DRB1, at positions 70 to 74 DRB1 chain which confers susceptibility to particular disease individuals. It was possible to develop a computational model for classification using computational models such as neural networks, Bayesian networks and kernel methods for obtaining interesting results.

RESUMEN

La Artritis Reumatoide (AR) es una enfermedad autoinmune que afecta del 1% al 3% de la población humana, existen marcadores genéticos implicados en esta enfermedad, se utilizaron modelos de inteligencia computacional para la clasificación e identificación de endofenotipos (relación entre fenotipo y marcadores genéticos) en pacientes con Artritis Reumatoide y controles sanos, esto a partir de información genética principalmente el HLA DRB1 (Antígeno Leucocitario Humano) así como la teoría del Epítoto Compartido, la cual hace referencia a la asociación entre la AR y el HLA-DRB1, principalmente los alelos que contienen un motivo común de aminoácidos las secuencias QKRAA, QRRRAA o RRRAA en el HLA-DRB1, en las posiciones 70 a 74 de la cadena del DRB1 los cuales les confiere una susceptibilidad particular de la enfermedad a los individuos. Se logró desarrollar un Modelo Computacional para clasificación utilizando modelos computacionales como Redes Neuronales, Redes Bayesianas y métodos de Kernel obteniendo resultados interesantes.

Palabras clave: HLA Antígeno Leucocitario Humano, Epítoto Compartido, Inteligencia computacional, Artritis Reumatoide.

1. INTRODUCCIÓN

La Artritis Reumatoide (AR) es una enfermedad autoinmune que está asociada con discapacidad progresiva, complicaciones sistémicas, muerte temprana y altos costos socioeconómicos. La causa es desconocida y su pronóstico es reservado.

La AR se caracteriza por la inflamación crónica de las articulaciones y afecta del 0.5-1% de la población adulta. Las principales articulaciones afectadas son pequeñas articulaciones de manos y pies.

El más importante factor de riesgo genético para AR se encuentra en el HLA. En particular los que contengan las secuencias de aminoácidos QKRAA, QRRAA o RRRAA en las posiciones 70-74 de la cadena DRB1. Esto es conocido como Epítoto Compartido (EC).

El curso clínico de la enfermedad difiere de un individuo a otro y los factores genéticos ligados a HLA DR han surgido como potencial factor pronóstico de esta enfermedad. El EC ha sido asociado con la predisposición y progresión de la enfermedad.

Los criterios para el diagnóstico de AR están definidos por el Colegio Americano de Reumatología y el EULAR 2010.

2. PLANTEAMIENTO DEL PROBLEMA

La AR afecta aproximadamente del 1% al 3% de la población humana con predominio femenino en una relación mujer hombre de 3:1, la herencia de la artritis reumatoide es poligénica y no sigue un patrón mendeliano. La importancia de encontrar factores genéticos asociados con la artritis reumatoide radica en la contribución a la comprensión de los mecanismos patogénicos de la enfermedad y su posible aplicación clínica para la identificación de marcadores, diagnóstico y pronóstico. Es posible a través de una herramienta computacional que use técnicas de Sistemas inteligentes identificar endofenotipos como son la severidad de la enfermedad o el daño óseo a partir de la tipificación del HLA Clase II en Pacientes con Artritis Reumatoide Colombianos.

3. METODOLOGIA

1. Se caracterizó una muestra de pacientes Colombianos con AR y un grupo control, datos obtenidos con consentimiento informado de cada uno de las personas implicadas en el estudio.
2. Se seleccionó y clasificó los datos de los pacientes con Artritis Reumatoide y la información de los controles para ser utilizados por el modelo computacional.
3. Se estudiaron y diseñaron varios modelos computacionales basados en técnicas de inteligencia computacional para la clasificación e identificación de endofenotipos tales como daño óseo, grado de severidad en pacientes Colombianos con Artritis Reumatoide.
4. Implementación de un prototipo de herramienta computacional a partir del modelo propuesto.

4. DATOS

Se inicia el estudio con una muestra de 149 personas de las cuales 74 corresponden a pacientes con artritis reumatoide y 75 son controles sanos. Dentro de los atributos que se seleccionaron para la realización del clasificador son: edad, género, tabaquismo, antecedente de hipertensión arterial, uso de hipolipemiantes, articulaciones inflamadas, articulaciones dolorosas, rigidez matinal, fatiga, además cada uno de los 2 alelos tipificados en el HLA DRB1 en el caso de pacientes y controles, para la identificación de endofenotipos en paciente con artritis se utilizaron además variables como: HAQ, Cuadro Hemático, Factor Reumatoide, Anti CCP, VSG, PCR, ANAS, DAS28, tratamientos utilizados DMART, esteroides, biológicos, etc.

5. PROCEDIMIENTO

Se trabajó con una muestra de datos de pacientes colombianos con Artritis Reumatoide y casos control. La figura muestra la tabla con los datos utilizados.

Se utilizó los datos tipificados en el HLA DRB1 y la información clínica de los pacientes con Artritis Reumatoide y de los controles para realizar análisis estadístico.

Se realizó la búsqueda de las secuencias de aminoácidos para cada uno de los distintos alelos para paciente y controles en el EBI (European Bioinformatics Institute).

Se realiza un alineamiento múltiple con las secuencias de aminoácidos obtenidas, además se realiza una búsqueda de motivos.

Se trabaja con datos de pacientes y controles para diseñar modelos de inteligencia computacional para lograr clasificar entre pacientes y control.

Se trabaja con los datos de pacientes para realizar identificar las relaciones entre datos genéticos y datos clínicos de severidad en Artritis Reumatoide.

Se comparan los distintos modelos se selecciona los mejores modelos computacionales por medidas de desempeño.

6. RESULTADOS OBTENIDOS

Se utilizó los datos tipificados en el HLA DRB1 y la información clínica de los pacientes con Artritis Reumatoide y de los controles para realizar un análisis de los datos.

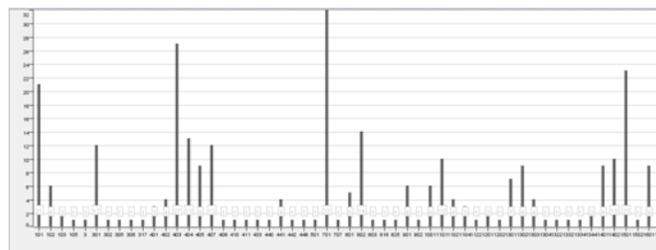


Figura No. 4: Histograma frecuencia de alelos DRB1

Los alelos más frecuentes en los datos estudiados son *0701 con una frecuencia de 32, *0403: 27, *1501: 23, *0101: 21, *0802: 14.

A partir de esta información encontramos las siguientes secuencias para cada uno de los alelos en estudio:
Frecuencia por secuencia:

Tabla No. 2 Tabla de frecuencias por secuencia de aminoácidos.

| Secuencia | n |
|-----------|----|
| QKRAA | 4 |
| QRRAA | 93 |
| RRRAA | 6 |
| DERAA | 39 |
| DRRGQ | 32 |
| QARAA | 23 |
| DRRAL | 14 |
| QKRGR | 12 |
| QRRAE | 12 |
| DRRAA | 10 |

Las secuencias QKRAA, QRRAA y RRRAA se repiten 103 veces en los datos analizados, estos se relacionan con el epítopo compartido.

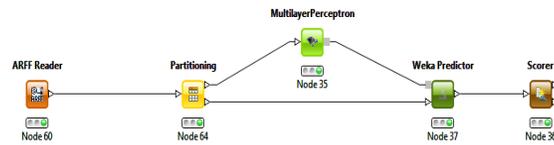


Figura No.6 Esquema Knime de Red Neuronal

A partir de los datos seleccionados de paciente con AR y de los controles se crea una red neuronal tipo Perceptrón Multicapa Red Neuronal para clasificación entre paciente con AR y casos control, se entreno con el 80% de los datos y se valido con el 20% de los datos obteniendo los siguientes resultados: se clasificaron adecuadamente 118 instancias de un total de 120 instancias lo que corresponde al 98.3% y se clasificaron inadecuadamente 2 instancias para un 1.6% incorrectamente clasificados.

| Evaluation: | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 118 | 98.3333 % |
| Incorrectly Classified Instances | 2 | 1.6667 % |
| Kappa statistic | 0.9667 | |
| Mean absolute error | 0.0346 | |
| Root mean squared error | 0.1037 | |
| Relative absolute error | 6.9191 % | |
| Root relative squared error | 20.7367 % | |
| Total Number of Instances | 120 | |

Figura No. 6 Resultados de desempeño de la Red Neuronal para clasificación entre AR y Control

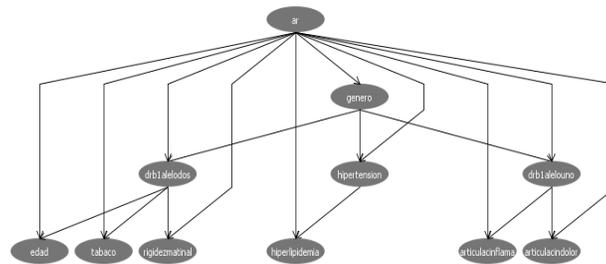


Fig. 7. Red Bayesiana obtenida para la clasificación de pacientes con AR.

Se utiliza un Red Bayesiana para la clasificación entre paciente con AR y casos control, se obtuvieron los siguientes resultados: se clasificaron adecuadamente 107 instancias de un total de 120 instancias lo que corresponde al 89.1% y se clasificaron inadecuadamente 13 instancias para un 10.8% incorrectamente clasificados.

| Evaluation: | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 107 | 89.1667 % |
| Incorrectly Classified Instances | 13 | 10.8333 % |
| Kappa statistic | 0.7833 | |
| Mean absolute error | 0.1408 | |
| Root mean squared error | 0.3013 | |
| Relative absolute error | 28.1168 % | |
| Root relative squared error | 59.8888 % | |
| Total Number of Instances | 120 | |

Figura No. 8 Resultados de desempeño de la Red Bayesiana para clasificación entre AR y Control

También se utilizo métodos de kernel para lograr dividir los 2 grupos obteniendo los siguientes resultados, los cuales reportan que se logro obtener 2 grupos

El desempeño de la Red Bayesiana para la clasificación de los paciente de acuerdo con su predisposición genética no fue el espero, esperamos mejor su desempeño en trabajos futuros.

REFERENCIAS

- [1] I. B. McInnes and G. Schett, "The pathogenesis of rheumatoid arthritis.," *The New England journal of medicine*, vol. 365, no. 23, pp. 2205–19, Dec. 2011.
- [2] M. Bax, J. van Heemst, T. W. J. Huizinga, and R. E. M. Toes, "Genetics of rheumatoid arthritis: what have we learned?," *Immunogenetics*, vol. 63, no. 8, pp. 459–66, Aug. 2011.
- [3] L. Massardo, N. Gareca, M. a Cartes, V. Cervilla, a González, and S. Jacobelli, "The presence of the HLA-DRB1 shared epitope correlates with erosive disease in Chilean patients with rheumatoid arthritis.," *Rheumatology (Oxford, England)*, vol. 41, no. 2, pp. 153–6, Mar. 2002.
- [4] D. L. Matthey, W. Thomson, W. E. R. Ollier, M. Batley, P. G. Davies, a K. Gough, J. Devlin, P. Prouse, D. W. James, P. L. Williams, J. Dixey, J. Winfield, N. L. Cox, G. Koduri, and a Young, "Association of DRB1 shared epitope genotypes with early mortality in rheumatoid arthritis: results of eighteen years of followup from the early rheumatoid arthritis study.," *Arthritis and rheumatism*, vol. 56, no. 5, pp. 1408–16, May 2007.
- [5] A. M. Wasserman, "Diagnosis and management of rheumatoid arthritis.," *American family physician*, vol. 84, no. 11, pp. 1245–52, Dec. 2011.

Authorization and Disclaimer

Authors authorize LACCEI to publish the paper in the conference proceedings. Neither LACCEI nor the editors are responsible either for the content or for the implications of what is expressed in the paper.