

Capítulo 5

CONCLUSIONES

El desarrollo de la presente investigación requirió de varias técnicas de análisis de información. En primer lugar, hubo que comprender a fondo el problema biológico y consultar las fuentes de información que sirvieron de antecedente. Una vez comprendido el problema, se empleó el análisis visual de una extracción de los datos originales para adquirir nociones sobre su estructura y presentación. Después se generalizaron estas nociones al conjunto total de datos que consideramos en nuestro estudio. Y, posteriormente, se efectuó una transformación de los mismos al lenguaje matemático para poder generar un modelo y obtener una solución numérica.

Debido a la enorme cantidad de datos con la que contábamos originalmente, nos vimos en la necesidad de diseñar mecanismos para compactarlos sin perder información valiosa. De otra manera hubiera sido mucho más complicado trabajar con ellos, realizar operaciones y detectar los errores. No obstante, aún habiendo compactado los datos, fue indispensable el empleo de herramientas computacionales para aplicar los algoritmos de transformación de la información y de búsqueda de soluciones. Para esto se utilizó el lenguaje de programación Java y el procesador de bases de datos Excel.

La solución encontrada para la clasificación de códigos genéticos de que trata esta tesis consta de dos etapas. En la primera etapa es posible ubicar en su clase a la mayoría de los datos, aplicando simultáneamente una lista de 60 preguntas que representa una solución parcial al problema. La segunda etapa es complementaria, pues sólo es necesario considerarla cuando el código genético bajo consideración logre eludir la clasificación de la primera etapa. Para estos casos, se determinó un criterio bajo el cual se elige un conjunto pequeño de preguntas pertenecientes a la segunda etapa que nos permiten alcanzar la clasificación deseada.

Por otro lado, edificamos una base de datos que debe ser consultada para aplicar los resultados obtenidos en esta investigación. En el capítulo 4 del presente documento se explica el procedimiento que debe seguirse para obtener la clasificación de un código genético particular, utilizando nuestra base de datos. Ahí se buscan un par de identificadores construidos en la primera etapa y, dependiendo de los resultados de la búsqueda, decidimos si es necesario pasar a la segunda etapa. Debido a las grandes

CAPÍTULO 5. CONCLUSIONES

dimensiones de dicha base de datos, ésta fue anexada como apéndice en un disco de 3.5 pulgadas, bajo un archivo de Excel.

A lo largo de este trabajo de tesis encontramos diversas dificultades. Una de ellas es la extensa cantidad de datos con los que trabajamos, lo cual implicó que la búsqueda de errores y excepciones fuera difícil y tomara un tiempo considerable. Por otro lado, no podemos dejar de mencionar la inexistencia de una solución general. Como se explicó en el capítulo 3, encontramos que aproximadamente el 1.79% de los datos totales no pueden ser clasificados bajo las hipótesis establecidas en esta tesis. Aunque, afortunadamente, para el resto de los casos nos fue posible obtener una clasificación. Otra complicación que nos encontramos fue la escasez de información de base. Existen muy pocas fuentes de información sobre este problema, lo que hizo más difícil su comprensión al inicio de nuestro trabajo.

Por último, a toda persona que desee indagar a mayor profundidad y extensión el problema de clasificación de alelos HLA, se sugiere consultar como referencia, además de este trabajo, la tesis de Beatriz Luna [5] y el artículo por Rafael Arguello [1]. Otra recomendación que podemos dar es la de utilizar lenguajes de programación especializados en el manejo de matrices o bases de datos de grandes dimensiones, pues esto reduce significativamente el tiempo análisis de datos y de aplicación de algoritmos de solución. Y, finalmente, si se desea trabajar con los alelos HLA del grupo B, se recomienda emplear los métodos desarrollados en este trabajo de tesis para obtener soluciones iniciales o para comparar los resultados obtenidos por otros métodos.