



## CAPÍTULO 6

### RESULTADOS

En este capítulo se presentan los resultados que se obtuvieron al aplicar la heurística de la entropía de Shannon y la Ganancia de Información al problema de la clasificación de alelos HLA así como de los problemas a los que se enfrentó la realización de este proyecto.

#### ***6.1 ASPECTOS TEÓRICOS***

Algunos problemas que en muchas ocasiones crean confusiones es la manera de explicar o querer dar a entender alguna situación. Este detalle se tuvo que enfrentar en la realización de este proyecto, debido a que al existir trabajos relacionados con la clasificación de alelos y otras investigaciones enfocadas a la tipificación hubo muchas confusiones en cuestión de términos biológicos y matemáticos, por lo cual se tuvo que recurrir a personas relacionadas con el tema para entender de mejor manera el problema, el grupo que aportó gran ayuda es “Kupachaloti” de la Universidad de las Américas, Puebla.

Otros problemas que se enfrentaron fue que al analizar el trabajo de Peña [P2004], el cual fue la base principal de este proyecto, no se contaba con una documentación del programa que realizó en lenguaje C para la clasificación lo cual creó una confusión y complicó la obtención de datos por medio de su programa. Por lo cual en este proyecto se creó una documentación del programa pero en lenguaje Java, la cual se localiza en el Apéndice **F** y de igual manera un manual para facilitar el uso del programa y el cual se encuentra en el



Apéndice **B**, esto es con el fin de que personas interesadas en el tema de la clasificación de alelos HLA y que conozcan sobre programación o únicamente tienen interés por alguna parte de la metodología que fue aplicada por medio del programa lo puedan consultar. Además de que esto permite, si el lector lo desea, convertirlo a otro lenguaje de programación.

## **6.2 ASPECTOS PRÁCTICOS**

Una de las decisiones que fueron de gran impacto para el proyecto fue la de definir la herramienta computacional que sería utilizada para aplicar la metodología. Se decide programar en Java debido a que esta herramienta está orientada a objetos, por lo cual el programa está modelado de acuerdo a los pasos que se van realizando en la entropía de Shannon y la ganancia de información, es decir, si en la metodología hay una parte que calcula los bialelos en el programa existe una clase que realiza este proceso, y así para cada paso que se va realizando en la metodología.

Otro resultado obtenido, y relacionado con el punto anterior, es la creación de la metodología en un lenguaje de programación Java. Es decir, la entropía ya se convirtió de un lenguaje matemático a un lenguaje computacional.

## **6.3 APORTACIONES**

Una de las aportaciones importantes obtenidas en el transcurso de este proyecto es que se comprobó que la aplicación de la metodología basada en la entropía de Shannon y la ganancia de información, realizada en el trabajo de Peña [P2004], sí proporciona un resultado para la clasificación de los alelos HLA permitiendo obtener soluciones iniciales con esta metodología, lo cual “deja la puerta abierta” para la aplicación de nuevas heurísticas basándose en estos resultados con la posibilidad de mejorarlos, en caso de que



sea posible.

Otra aportación importante es que el programa realizado en Java permite clasificar cualquier catálogo de alelos HLA a diferencia del trabajo realizado por Peña [P2004], lo cual es una gran herramienta para aquellas personas que deseen obtener soluciones iniciales a través de la heurística de la entropía de Shannon y de la ganancia de información.

En la siguiente sección se propone una heurística para aplicarla en el futuro al catálogo de preguntas generadas por la entropía con la finalidad de buscar una mejor solución, en caso de que pueda mejorarse.

### **6.3.1 HEURÍSTICA PROPUESTA**

Al contar con un catálogo inicial de preguntas ya es posible realizar una mejora a dicha solución. Peña [P2004] propone ir eliminando preguntas del catálogo inicial de 65 preguntas que obtuvo y aplicar la heurística de cumplir la igualdad de la entropía del atributo con la entropía del sistema, así hasta reducir al mínimo el número de preguntas del catálogo. Al aplicar esta heurística no pudo encontrar una mejor solución que las 65 preguntas seleccionadas al inicio.

La metodología que se propone en esta tesis para mejorar la solución inicial obtenida de la entropía es la siguiente:

1. Definir el valor de la Entropía de Shannon del sistema  $H(C)$ .
2. Obtener un conjunto de preguntas que conformen la solución inicial  $SI = \{I_1, I_2, I_3, \dots, I_j\}$  donde  $I_j$  es el número total de preguntas seleccionadas en base a la ganancia de información que cada una proporciona.



3. Se obtiene un conjunto de preguntas no seleccionadas  $NS = \{N_1, N_2, N_3, \dots, N_i\}$  donde  $N_i$  es el número total de preguntas que no fueron elegidas para la solución inicial.
4. Tomar dos preguntas del catálogo  $SI$  y cambiarla por una sola pregunta no seleccionada del conjunto  $NS$ .
5. Calcular la ganancia de información que proporciona dicho intercambio a través de la entropía del atributo intercambiado.
6. En caso de que se obtenga una ganancia de información mayor con el intercambio, se genera un nuevo catálogo tomando éste como inicial.
7. Si no se obtiene una mejora en la ganancia de información se regresa al paso 4, tomando otra pregunta diferente del conjunto  $NS$ .
8. Repetir estos pasos hasta obtener una mejor ganancia de información, en caso de que no se mejore el catálogo se queda con la solución la generada en el paso 2.
9. Al terminar los pasos del 2-8 se define el catálogo de la solución final  $SF = \{F_1, F_2, F_3, \dots, F_t\}$  siendo  $F_t$  el número total de preguntas seleccionadas como solución al problema de la clasificación de bialelos HLA.

Con los pasos descritos con anterioridad se puede llegar posiblemente a reducir el número de preguntas contenidas en el catálogo. Estos catálogos van a variar de acuerdo al catálogo de alelos HLA que se trate debido a la cantidad de información que cada uno contenga.