

CONCLUSIONES

Para la realización de esta tesis en primer lugar hubo que comprender y estudiar a fondo el problema biológico y consultar las fuentes de información que sirvieron de antecedentes. Una vez comprendido el problema, se hizo una extracción de los datos originales recortando una pequeña ventana, de esta manera se estudio la estructura y presentación de los datos y se hizo un análisis visual para encontrar un catálogo de preguntas que diera una solución para esta pequeña ventana de datos. Después se generalizaron estas nociones al conjunto total de datos que se consideran en este trabajo y de acuerdo al modelo matemático se empezó a trabajar en la transformación de los datos para obtener la matriz de bialelos. Después de esto se tomó la decisión de utilizar el concepto de las preguntas en 20 posiciones para reducir el conjunto de soluciones sin perder información valiosa.

Una vez que se obtuvo el conjunto de las preguntas de 20 posiciones, se realizaron estas preguntas a la matriz de bialelos, considerando que si la pregunta es afirmativa para uno o los dos alelos se identifica con el dígito 1, si la respuesta es negativa, es decir, que ninguno de los bialelos tiene esa combinación de letras que da la pregunta, se identifica con el dígito 0. Así se construyó una matriz a la que se llama matriz de evaluación M_E de una dimensión de 38781 filas las cuales son los bialelos y 1117 columnas que representan al conjunto de preguntas.

Debido a la dimensión de los datos fue necesario utilizar herramientas computacionales para realizar a la matriz de bialelos M_b , la búsqueda de las preguntas en

posiciones de 20 y la construcción de M_E . Los programas se realizaron en lenguaje C y también se utilizó el procesador de base de datos Excel para almacenar y estudiar los datos originales.

La primera parte del trabajo consiste en la construcción de M_E , la segunda parte consistió en aplicar los conceptos de entropía de Shannon y ganancia de la información, la cuál incluye la heurística para la selección de preguntas y de esta forma encontrar una solución inicial al problema.

La solución inicial consiste en un catálogo de 65 preguntas seleccionadas del catálogo de 1117 preguntas, las cuáles dan una identificación única de 1's y 0's que permiten identificar a cada una de las 231 biclases del problema mediante una cadena de 65 posiciones de 1's y 0's. Es importante mencionar que a esta solución se le aplicó un algoritmo heurístico de mejora para tratar de reducir el tamaño del catálogo. Pero no fue posible encontrar una mejor solución, por lo tanto el catálogo de 65 preguntas que se obtuvo en la sección 4.2.2 se considera como la solución para el problema de clasificación de bialelos para esta tesis.

Una de las dificultades encontradas en la realización de este trabajo es la extensa cantidad de datos con las que se trabaja, lo cual implicó que al ejecutar el programa para la búsqueda de las 65 preguntas tomó un tiempo de 192 horas para encontrar esta solución. Otra complicación encontrada es que existen pocas fuentes de información sobre este problema, lo que hizo más difícil su comprensión al inicio del trabajo.

Una pregunta abierta, es que a partir de la clasificación de bialelos proporcionada en este trabajo de investigación se puede aplicar otros métodos heurísticos para reducir esta clasificación. Y finalmente, si se desea trabajar con los alelos HLA de otro tipo o grupo, se sugiere emplear el método desarrollado en este trabajo para obtener soluciones iniciales o para comparar los resultados obtenidos por otros métodos.