UNIVERSIDAD LATINA, S.C.

INCORPORADA A LA U.N.A.M.

CAMPUS CUERNAVACA 8344-48

FACULTAD DE INFORMÁTICA

Un Método de Minería de Datos para la rectificación del ácido fólico y su determinación en el Labio Paladar Hendido y/o Fisura Labiopalatina

T E S I S

PARA OBTENER EL TÍTULO DE:

LICENCIADO EN INFORMÁTICA

P R E S E N T A:

Francisco Javier López Riojas

ASESOR: DCC. MBA, MANI y L.I. Sergio Mauricio Martínez Monterrubio

CUERNAVACA MOR. 2014

Índice

LISTA DE FIGURAS

LISTA DE TABLAS

LISTA DE GRÁFICAS

Autorización de impresión de tesis (escanéo)

AGRADECIMIENTOS

RESUMEN

En la ciencia de la informática durante los últimos 30 años ha sido capaz de definir el futuro de varias empresas, pronosticar la salud, elecciones electorales y determinar el comportamiento de los consumidores, todo esto es posible a las herramientas que se han desarrollado. Hablando específicamente de la Minería de Datos (de sus siglas en inglés *Data Minning*); la cual es una herramienta informática cuyo objetivo es extraer mediante ciertos algoritmos, patrones de los cuales se pueden inferir cuestiones no evidentes.

"La minería de datos, es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto." (Fernández Aldana, 2005) Existen algoritmos especializados en la Minería de datos, debido a la complejidad de la información y sus múltiples variantes; sin embargo esto no deja de ser un proceso en el cual el principal objetivo es la extracción de información concreta a partir de un repositorio de datos.

Considerado para muchas otras disciplinas como una herramienta, la Minería de Datos está siendo usada cada vez más por el área médica, cuando su principal cliente fue el sector empresarial; cada vez es más el uso en las cuestiones médicas, no solo en la práctica sino en la investigación. Esta tesis es enfocada en el padecimiento o malformación bucofaríngea la cual es llamada Labio Paladar Hendido y/o Fisura Labiopalatina; la cual ciertos estudios científicos (De-Regil, Fernández-Gaxiola, & Peña-Rosas J., 2010) han determinado que el uso de ácido fólico en cierto periodo de concepción se puede evitar. Obteniendo una muestra significativa de pacientes procedentes de una clínica dental, se pretende obtener los mismos resultados que los estudios anteriores mencionados, utilizando la Minería de Datos.

CAPÍTULO 1. METODOLOGÍA DE LA INVESTIGACIÓN

1.1 Planteamiento del Problema

Recientemente dentro de las múltiples enfermedades que giran en torno al nacimiento de los seres humanos se tienen diferentes síndromes así como también enfermedades hereditarias, deformaciones físicas o malformaciones, deficiencias de algunas componentes principales, etc. Esto es causado en muchas ocasiones por factores hereditarios, multifactoriales, en otros casos por un mal cuidado durante la concepción del bebe, a su vez se pueden sumar también el uso de narcóticos antes y después del embarazo. Sin embargo uno de las malformaciones que ha venido aumentando desde hace ya unos 10 años, es el Labio Paladar Hendido (LPH) o Fisura Labiopalatina. La malformación multifactorial puede ser parte de un síndrome o puede venir de forma única en la mala formación del cráneo y cara del bebe. No se ha detectado la verdadera causa de esta malformación, sin embargo se tiene en el contexto que puede ser hereditario o en algunas ocasiones. En los últimos 5 años ha aumentado en México el número de casos, posiblemente por la falta de vitaminas necesarias en el embarazo o bien la carencia de ingesta del ácido fólico.

Recientes estudios (Tolarova, 1998) y (De-Regil, Fernández-Gaxiola, & Peña-Rosas J., 2010) indican que el ácido fólico en personas que no presentan LPH o FL de manera genética sus hijos recién nacidos han podido evitar este padecimiento incluso se lleva a cabo una pequeña muestra donde es administrado cierta cantidad de ácido fólico y vitaminas dando como resultado evitar totalmente el LPH. A pesar de estos estudios, no se tiene una veracidad científica completa de este hecho, incluso se ha tomado como medida de prevención en el embarazo; sin embargo con los estudios actuales no se ha demostrado que el ácido fólico evita el LPH.

Dentro del tema actual delimitado existe cierta bibliografía que recurre a soluciones de este problema tanto en el sentido quirúrgico una vez la concepción del bebe, sin embargo de acuerdo a la relación del ácido fólico con la intervención del LPH es escaza tanto médicas, como investigaciones por medio de Minería de Datos con relación a este tema.

1.2 Hipótesis

Por medio de un algoritmo de Minería de Datos utilizándolo en bases de datos médicas de pacientes con y sin LPH se demuestra que el ácido fólico es una forma de erradicar la malformación del LPH.

1.3 Objetivo General

Se pretende que por medio de la Minería de Datos, utilizando una base de datos de pacientes con y sin LPH, se pueda dar veracidad a estudios como el de (De-Regil, Fernández-Gaxiola, & Peña-Rosas J., 2010) y (Tolarova, 1998) los cuales aseguran que tomando ácido fólico durante el embarazo se previene el LPH de pacientes que no tienen padres con LPH u otro síndrome mayor.

1.3.1 Objetivo Particular

Por medio de la Minería de Datos utilizando los algoritmos ID3 y C4.5 se identificarán patrones en cuanto a los niños que ya hayan nacido con LPH, descartando aquellos que sean de manera hereditaria y tomando también dentro de la muestra aquellos que no tengan este padecimiento y sus cuidados prenatales. Utilizando una base de datos de pacientes de los cuales tendrán LPH y otros que no lo tengan, se pretende realizar la toma de patrones y pronosticar de alguna forma aquellos nuevos embarazos que se tengan si es que tendrán LPH o no.

Tomando como referencia los estudios acerca del ácido fólico y el LPH, su formación, concepción y situaciones que cometen a esa malformación, se pretende proponer el hecho de que el ácido fólico elimina el LPH.

1.4 Alcance

La investigación se verá basada principalmente en los artículos o estudios de (De-Regil, Fernández-Gaxiola, & Peña-Rosas J., 2010) y (Tolarova, 1998) de los cuales se han tomado diferentes decisiones en cuanto a México y la Secretaría de Salud y un análisis por medio de un cuestionario en una clínica dental con pacientes con LPH.

Se tomarán en cuenta pacientes de Rehabilitación Dental en una clínica especializada ubicada en la ciudad de Jiutepec, Morelos México que tiene un área de atención en pacientes con LPH y su planeación de estos niños. Solo se tomará en cuenta aquellos niños nacidos con LPH no sindrómico, pacientes de los cuales no tengan malformación de LPH pero tampoco cuenten con otro síndrome; serán descartados todos los demás de los cuales la toma de ácido fólico ayuda más no erradique la malformación del LPH.

CAPÍTULO 2. MARCO CONCEPTUAL

2.1 Labio Paladar Hendido y Minería de Datos

2.1.1 Labio Paladar Hendido (LPH) y/o Fisura Labiopalatina (FP)

"El complicado desarrollo de la cara a partir de los arcos branquiales produce muchas anomalías cráneo-faciales de las cuales el labio y paladar hendido es una malformación producidas en las estructuras orofaringonasales. Se denomina Labio Paladar y Hendido a las malformaciones congénitas producidas por defectos embriológicos en la formación de la cara. El término fisura se define como "apertura alargada", especialmente la que se produce en el embrión derivada de una falta de fusión de determinadas partes durante el desarrollo embrionario.

Se conocen fisuras de labio desde el siglo II y fue Galeno quien las identificó con el nombre de "Lagocheilos" que significa "labio de liebre". Se tiene registro de la primera cirugía para corregir esta anomalía realizada por el francés Le Monier en 1764. El LPH es la más común de las enfermedades congénitas presentándose en promedio 1 por cada 1,000 nacimientos, más frecuente en varones que en mujeres y varía en distintos grupos de población. Es menos frecuente el paladar hendido que la fisura labial teniendo como dato 1 por cada 2,500 nacimientos; y de mayor frecuencia en mujeres que en varones y no guarda relación con la edad materna. Parece ser que en la mujer las crestas palatinas se fusionan aproximadamente una semana más tarde que en el varón lo que explica la mayor frecuencia del paladar hendido aislado con mayor frecuencia en la mujer.

Si los padres son normales y tienen un hijo con LPH la probabilidad de que el siguiente la presente es del 2%; sin embargo si un familiar o uno de los padres y un hijo presentan paladar hendido la probabilidad aumenta al 7% en mujeres y 15% en hombres. El peligro de repetirse la anomalía en hijos del mismo matrimonio suele ser mayor cuanto más grave sea el defecto. Se dice que la fisura labial es tres veces más frecuente en los caucásicos que en la raza negra." (Flores, 2009)

2.1.2 Papel del ácido fólico en LPH

"Las investigaciones epidemiológicas y genéticas han determinado además una susceptibilidad genética la cual depende de la combinación de numerosos genes algunos de los cuales han sufrido mutación y pueden obrar conjuntamente con los factores el entorno dando origen a fisuras orofaciales y otras anomalías congénitas tales como defectos del tubo neural (NTD) y defectos congénitos del corazón. Por lo tanto los genes pueden obstaculizar e interferir en el mecanismo del ácido fólico, además en el consumo de importantes cantidades de ácido fólico durante la concepción y principio del embarazo puede darse lugar a que el infante nazca con FLP[13].

Las interacciones gen-gen y gen-entorno incluyendo exposiciones ambientales y deficiencias alimenticias producen efecto diverso sobre el riesgo de enfermedad o anomalía congénita con genotipos diferentes. El genoma humano puede clasificarse en 23 pares de cromosomas, 2,000 bandas cromosomales, 80,000 genes y aprox. 3 mil millones de pares base de proteínas. Cada uno de estos pares base, es un fragmento del código del ADN para una proteína específica. Una mutación en un solo par base (en un solo punto de la cadena) puede acarrear una enfermedad o anomalía congénita.

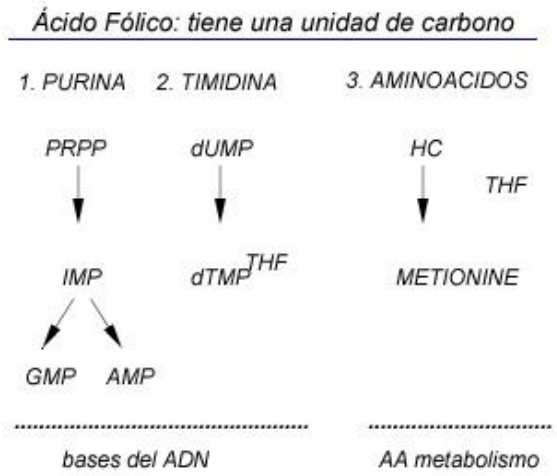


Ilustración 1. Ciclo del Ácido Fólico

Ilustración 2. Ácido Fólico en la embriología de las fisuras orofaciales

En el año 1930 el ácido fólico o vitamina B11 fue reconocido por primera vez como factor presente en el “marmite”, una preparación a base de extracto de levadura que pudo curar una anemia megaloblástica que ocurría entre las mujeres hinduistas (India), especialmente durante el embarazo. El término “ácido fólico” (derivado del latín folium, “hoja”) se usó para designar el ácido pteroilglutámico fue introducido por Michel en 1941 y la sustancia fue sintetizada con éxito en la forma de pteroilmonoglutamato por Angier en 1946. Ácido fólico y folato son los sinónimos preferidos para el ácido pteroilglutámico. Este está compuesto de una pteridina, el ácido p-aminobenzoico y el ácido glutámico. Los folatos están disponibles en parte como moléculas pequeñas provistas de una 1 a 3 cadenas laterales, es decir, mono y oligoglutamatos y en parte como molécula más grande es decir, poliglutamatos. Los folatos están presentes en todo tipo de tejido corporal. La mayoría de ellos se almacenan en forma de poliglutamato en el hígado, páncreas, riñón y cerebro tanto los mono como oligoglutamatos (folato libre) y los poliglutamatos representan el “folato total”.

La sangre procedente del cordón umbilical contiene enlazadores de folato tanto de baja como de alta afinidad. Las proteínas enlazadoras de folato están implicadas en la transferencia del folato contra un gradiente de concentración de madre a feto, lo que

sugiere el transporte activo por parte de la placenta, por lo cual el feto está capacitado para acumular el folato. Parece ser que la placenta proporciona cantidades suficientes de folato al feto sin embargo se sabe poco acerca del suministro del folato para el crecimiento antes del desarrollo de la placenta. La función principal del folato es la de proporcionar para la síntesis de 3 de las 4 bases del ADN, que son: guanina, adenina y timina además para la síntesis de otros compuestos. Estos son imprescindibles para la constitución de formatos para la síntesis de las purinas 5-formamidoimidazola-4-carboxamida ribonucleotida (FAICAR) y formilglicinamida-ribonucleotida (FAGR) metileno para la síntesis de deoxitimidina (dTMP) y el grupo metílico para síntesis de metionina. La síntesis de metionina requiere como cofactor a la vitamina B12 ósea a la metilcobalamina. La deficiencia tanto de folato como de vitamina B12 disminuye la formación de metionina lo que conlleva a una deficiencia funcional del folato aunque estén presentes niveles desde normales a elevados del folato en sangre." (Flores, 2009)

2.1.3 Estudios sobre el LPH y el ácido fólico

"Más de 20 años después de que los primeros estudios en animales de laboratorio indicaron que la deficiencia de vitamina en la madre pudiera originar malformaciones congénitas en los hijos, se mostros que la prueba de excreción del ácido formiminoglutamico indicador del metabolismo defectuoso del folato dio positivo con más frecuencia en mujeres embarazadas de un niño con defecto del tubo neural u otras anomalías congénitas que en los sujetos de control.

La provisión de suplementos multivitaminicos o ácido fólico en el período periconcepcional jugó un papel importante en la prevención de los NTD. No obstante la prevención de las anomalías congénitas parecía imposible como meta última de la teratología hasta que un ensayo aleatorizado, controlado y doble-ciego patrocinado por el Consejo de Investigaciones Médicas Británica mostro un descenso del 72% en la repetición de NTD cuando las mujeres ingirieron 4mg al día de ácido fólico desde el día de la aleatorizacion antes de la concepción y durante 12 meses después. Solo uno de todos los ensayos de intervención y de observación sugiere que el folato dietario y la utilización de suplementos

o multivitaminicos o del ácido fólico reducen el riesgo de que una mujer alumbre a un hijo con NTD.

Era precisamente por la fisura labial y el paladar que se realizaron los primeros intentos de utilizar la terapia multivitaminica profiláctica incluyendo el ácido fólico para evitar la repetición del defecto en los seres humanos. Los primeros intentos de evitar las fisuras orofaciales los realizaron los cirujanos plásticos norteamericanos en los años 50's (Peer, Douglas y Conway en 1958). En estos estudios de observación el autor sugirió que las mujeres que utilizaban un suplemento multivitaminico que contenía ácido fólico tienen un riesgo menor de alumbrar hijos con fisura orofacial." (Flores, 2009)

2.2 Minería de Datos (MD)

Tradicionalmente los analistas han mejorado la tarea de extraer información útil de datos almacenados. Sin embargo el incremento de volumen de los datos tanto en los negocios como en las ciencias, ha requerido se tenga más aproximación a una base computarizada. Mientras los repositorios de los datos crecen más en tamaño y complejidad ha habido en gran cambio el manejo de la información de manera indirecta. El análisis automático de la información utiliza herramientas más complejas y sofisticadas; las tecnologías modernas organizan y colocan la información en categorías pero con un gran esfuerzo. Sin embargo los datos capturados necesitan convertirse en información y esta a su vez convertirse en conocimiento para que realmente pueda ser útil. En esta época es un mundo del manejo de los datos; datos que son analizados y procesados para convertirse en información, que informa, instruye, responde y ayuda al interpretación y toma de decisiones. Los datos son almacenados en diferentes sitios, sin embargo el uso de Data Warehouse[1] es más eficiente para la utilización de la minería de datos, teniendo todo dentro de un ambiente centralizado.

[1] Data Warehouse.- Repositorio de datos de muy fácil acceso, alimentado de numerosas fuentes, transformadas en grupos de información sobre temas específicos de negocios, para permitir nuevas consultas, análisis, reporteador y decisiones. (Simmon, 2009)

2.2.1 Definición y conceptos

Dentro del *"Knowledge Discovery in Databases"* conocido como KDD, se encuentra la minería de datos, campo de las ciencias de la computación especializado a encontrar patrones dentro de grandes volúmenes de datos. Utilizando diferentes métodos como inteligencia artificial, aprendizaje automático, estadístico y sistemas de base de datos, es que dentro de las metas de la MD es tomar aquel gran conjunto de datos y llevarla bajo ciertos procesos la cual transformaran la información en una estructura más usable para otros objetivos o decisiones. Todo esto sumado a las diferentes estrategias para los grandes repositorios de datos y tratamientos para la información es que se quiere llegar al "descubrimiento de algo nuevo" lo cual es preferible y objetivo principal dentro de la MD. (Kamber, 2006).

En muchos casos se puede identificar a la MD como el análisis automático o semi-automática de la información, mostrados principalmente en patrones de comportamiento de los datos; de los cuales hay diferentes modelos finales de donde se pueden extraer diferentes conclusiones en cuanto a la información base. La MD utiliza patrones los cuales son utilizados para la máquina de aprendizaje[2] y análisis predictivo. La MD utiliza grupos los cuales están alojados los datos haciendo que su predicción sea más precisa, sin embargo esto genera más datos más simplificados mejor utilizados y resumidos para terminar el proceso de KDD; por tanto se podría en algunas ocasiones mal interpretar que la MD, recolecta, prepara, interpreta los datos, porque estos pasos son directamente tareas de la KDD. La MD se relaciona específicamente en cuanto a la obtención, extracción y análisis de los datos por medio de diferentes métodos, los cuales utilizan los más grandes conjuntos donde la muestra suele ser tan pequeña que los métodos rutinarias estadísticos no son capaces de encontrar inferencias en cuanto a los patrones establecidos, creando entonces hipótesis para probar a grandes conjuntos.

2.2.2 Proceso de la MD

[2] Sistema de información utilizado para la predicción de comportamiento en patrones dentro de la MD

Dentro de la MD se tiene un proceso principal el cual es el siguiente:

1. **Selección de datos**

   Teniendo en cuenta los datos en bruto (base de datos), se determinan las variables necesarias para que los algoritmos puedan realizar la predicción

2. **Análisis de datos**

   Donde se analizarán histogramas, diagramas de dispersión, presencia de valores atípicos y ausencia de valores.

3. **Datos de entrada**

   A este paso también se le conoce como preprocesamiento de los datos, ya que los datos se les dará el formato adecuado y análisis dependiendo del método o algoritmo que se usará.

4. **Construcción del modelo**

   Se construye el modelo con el cual se va a trabajar, dentro de los cuales encontramos: predictivo, de clasificación o segmentación.

5. **Extracción de datos**

   Mediante un método de minería de datos, se genera un modelo donde se presentan los patrones de los datos.

6. **Interpretación de la información**

   Obtenido el modelo, se procede a sacar las conclusiones necesarias para responder a las dudas planteadas desde un principio.

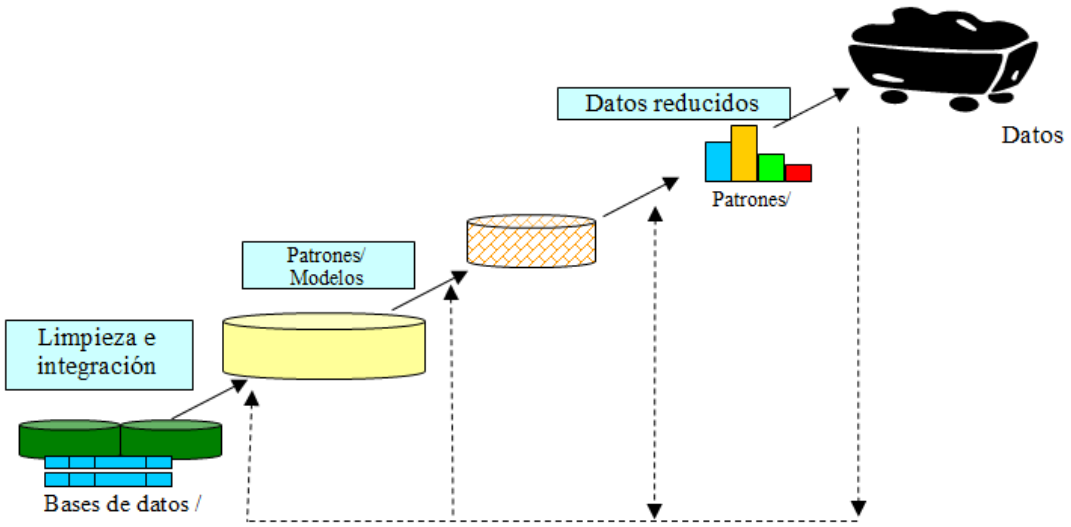


Ilustración 3. El KDD ha sido definido por (Fayyad et al, 1996) como "la identifiación no trivial de patrones válidos, nuevos, comprensibles y potencialmente útiles en los datos".

Una vez obtenido un modelo deseable, el cual contiene salidas y/o márgenes de error correctos, el modelo entonces se prepara para la explotación del mismo. Dentro de la toma de decisiones, procesos de validación y otro tipo de necesidades en varios ámbitos, los modelos desarrollados por la minería de datos se adjuntan en estos para obtener u optimizar el resultado y resolver aún más incógnitas.

2.2.3 Técnicas dentro de la MD

Principalmente la MD utiliza algoritmos, los cuales utilizan los patrones detectados en los datos y teniendo una entrada de datos, estos algoritmos generan una salida, la cual generalmente es un análisis completo de estos patrones, dando entonces resultados determinantes y que se pueden utilizar para generar otro tipo de datos más maleables (conocimiento).

Este tipo de algoritmos son básicamente provenientes de la inteligencia artificial y la estadística aplicada, estás técnicas optimizadas para poder arrojar los resultados necesarios para usarlos después y obtener excelentes resultados. Dentro de estos algoritmos existen diferentes tipos de ellos, todo va a depender del tipo de datos que se estén manejando en el preprocesamiento y claro está en el resultado deseado; dentro de esta investigación se manejarán los algoritmos basados en árboles de decisión.

Un árbol de decisión no es más que modelos de inteligencia artificial los cuales funcionan mediante conjeturas lógicas que por medio de reglas, actúan de manera sucesiva creando categorías dentro de sus condiciones para resolver problemas.

ID3 (Iterative Dichotomiser 3)

Ross Quinlan es el inventor del algoritmo ID3, este algoritmo básicamente genera arboles de decisión por medio de una conjunto de datos. ID3 utiliza el cálculo de la entropía [3]para verificar cual es el atributo más "débil" eliminándolo del conjunto, de esta forma va creando las ramificaciones del árbol, al quitar los atributos más débiles los deja como el final de una ramificación y así avanza hasta llegar al final del conjunto; al final utiliza la recursión para recrear pequeños conjuntos de datos de aquellas variables que no se han utilizado antes.

Este algoritmo no garantiza una solución óptima y definitiva. Sin embargo ID3 puede ser entrenado, lo cual garantiza que durante el entrenamiento con datos que se sabe su resultado puede dar aun árboles no tan grande y más explícitos; de esta manera se evita el sobre entrenamiento, el cuál crearía árboles demasiado grandes y recursivos. Con datos continuos, ID3 no es muy recomendable usarlo, debido a que puede haber un sinfín de posibilidades de las cuales se puede ampliar el atributo, el cálculo de este tipo de árboles con este tipo de atributos puede llevar demasiado tiempo obtener un resultado. El uso más adecuado de este algoritmo es utilizado para mediante el entrenamiento de los datos desconocidos, se crean árboles de decisión los cuales se usarán para formar caso de uso en ejemplos y ver la fuente de ciertos conjunto de datos.

C4.5

Como extensión del algoritmo ID3, es conocido como un clasificador estadístico debido a su uso principal basado en la clasificación. Este algoritmo funciona de la misma forma que ID3, utilizando en primeras instancias la entropía, sin embargo, al momento de analizar cada nodo del árbol, analiza la capacidad de dividirse en otros criterios albergados dentro del criterio principal, esto conocido como la ganancia de información normalizada; de tal forma que aquel atributo con la mayor ganancia de información normalizada es el que toma la decisión para el camino del árbol.

Dentro de esta investigación se llevarán a cabo los modelos con estos dos algoritmos, debido a que la hipótesis es referida en base a un juicio de decisión sobre datos conocidos y además de que las variables dentro del conjunto no son muchas y son de índole binarias; tanto ID3 como C4.5 son la mejor opción para obtener un resultado más acertado para apuntar a una u otra hipótesis. Ya que ambos algoritmos manejan lo que es la predictibilidad del conjunto y de futuros registros, dejando entonces no solo la efectividad de la respuesta basada en el árbol de decisión sino también la posible respuesta teniendo nuevos casos.

[3] Medida de la incertidumbre de una variable aleatoria.

2.2.4 Software utilizado para la MD

Para la realización tanto del modelaje como del minado de datos y la obtención de resultados ya se encuentran diferentes herramientas (software) que no solamente han sido programadas para llevar a cabo los diferentes pasos para la MD, sino también analizar los datos de entrada y salida, etc.

Weka

Dentro de estas herramientas podemos encontrar el software WEKA, programado en java por la Universidad de Waikato y bajo la licencia de software libre (GNU-GLP). Esta herramienta tiene una colección de algoritmos y herramientas de visualización, incluso maneja algunas herramientas de pre-procesamiento de los datos.

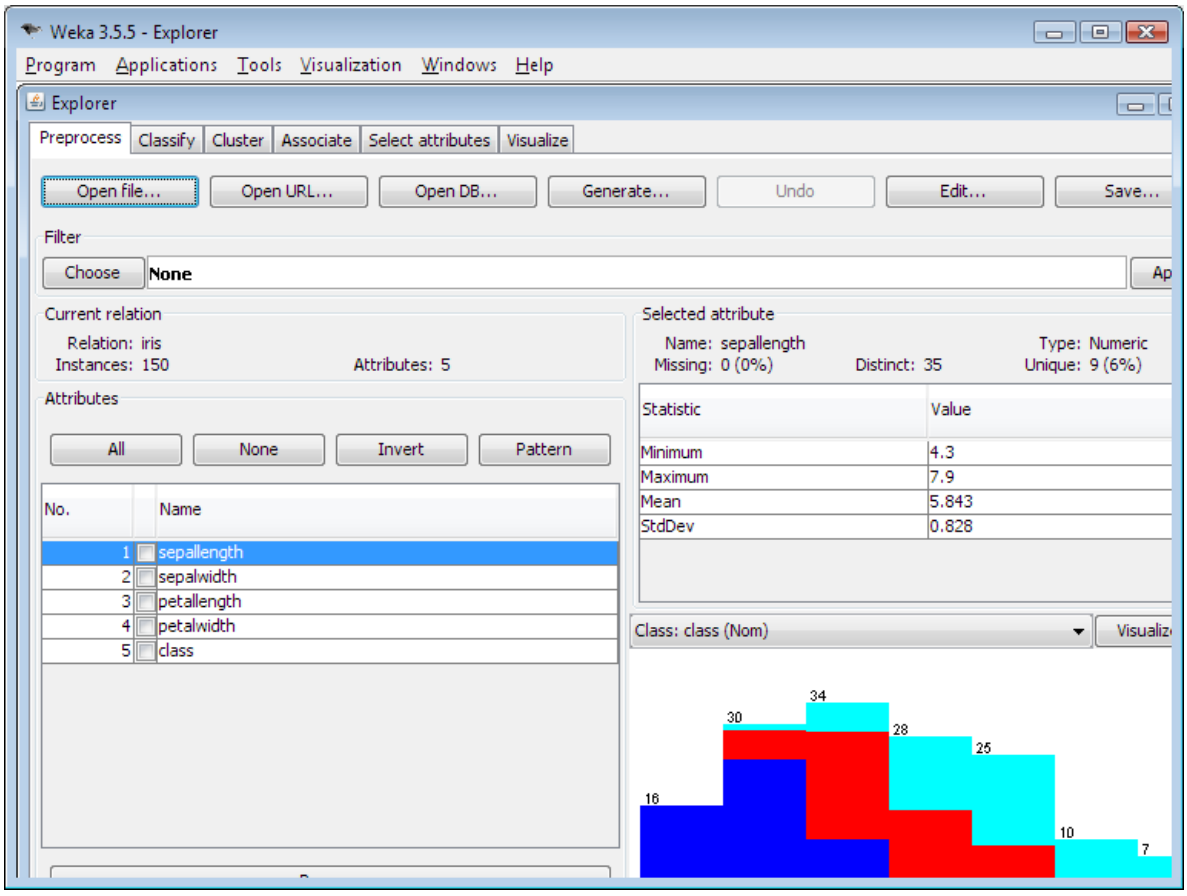


Ilustración 4. Programa en uso (WEKA)

RapidMiner

Desarrollado en la Universidad de Dortmund, RapidMiner utiliza un entorno gráfico de desarrollo por medio de operadores, de los cuales son utilizados en procesos los cuales generan salidas. Además de su facilidad con el entorno, maneja una gama sumamente alta de algoritmos y herramientas para la MD, estadística e inteligencia artificial cubriendo muchos aspectos de los cuales no solo en la MD se utilizan sino también en cualquier instancia de aprendizaje de la información. Se puede manejar el pre procesamiento de los datos, como incluso la preparación de estos, pase anterior a cualquier proceso de la MD, además de incluir múltiples recursos de entrada para los conjuntos de datos y salidas para los mismos.

Debido a módulos de integración con WEKA para los algoritmos de árboles de decisión ID3 y C4.5 este es el software que se utiliza en la investigación para todo el proceso de la MD. No solo por la robustez, sino también la facilidad y amplio espectro a datos a utilizar, ya muchos otros software no solo tienen un mínimo a usar en el conjunto de datos sino también hay limitaciones en la salida de los diferentes procesos.

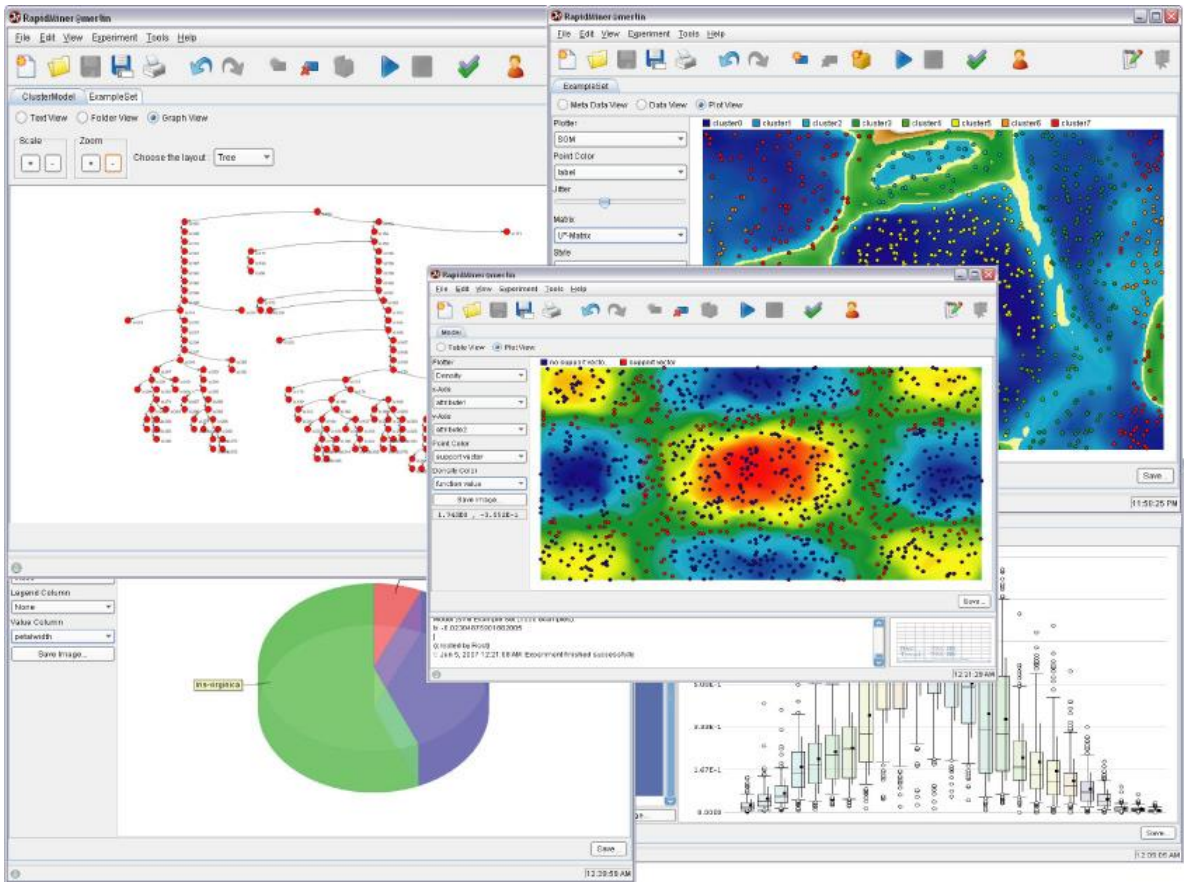


Ilustración 5. Ejemplos y resultados en Rapidminer

CAPÍTULO 3. ESTADO DEL ARTE

3.1 Investigaciones actuales

Dentro del campo de la investigación en medicina, se tienen una fuente de recursos que día a día va creciendo, segundo a segundo cambiando y a cada momento es más precisa en algunas cosas. Dentro de este campo, el análisis de la información por medio de técnicas basadas en la informática está apenas siendo explorado por muchos y aceptado por pocos, pero su precisión ha dejado a varios con expectativas muy altas y proyecciones muy prometedoras.

Dentro del campo de los problemas bucales, específicamente en alteraciones fisiológicas, no se tiene una investigación dirigida al 100% a los factores que producen estas malformaciones, como son el LPH. Muchos de los estudios de los cuales se han dedicado a tratar de identificar los factores exactos que provocan la malformación son realmente escasos. Se tienen muchos más avances en cuanto al tratamiento posterior al nacimiento y durante los primeros años de vida, sin embargo estos avances y métodos no están al alcance de todos y mucho menos en cualquier centro de salud. Enfocándonos en los pocos estudios que se dedican a establecer un factor del LPH, en este caso centrándose en la falta de ingesta de ácido fólico antes y durante el embarazo, tenemos dos estudios:

3.1.1 Primera Reunión Internacional de ROTAPLAST Internacional y FUNDAPAFI (Tolarova, 1998)

Investigación realizada por la Dra. Marie M. Tolarova dentro del proyecto de investigación genética y prevención ROTAPLAST Internacional para LPH, Caracas Venezuela. Dentro de esta investigación, por medio del proyecto, se toma como hipótesis que la falta de folatos (ácido fólico) en la dieta de las mujeres antes de embarazarse y durante la concepción produce que en la formación del bebe haya esta falta de componentes, creando la malformación del LPH.

Mediante una muestra controlada de 600 mujeres de las cuales el 50% solamente se quedó con una dieta estructurada rica en ácido fólico y vitaminas necesarias para evitar la presencia de LPH mientras que el porcentaje restante quedó sin ningún cuidado específico. El estudio concluye que mediante aquellos cuidados que se le dieron a las 300 mujeres embarazadas previenen de manera total la presencia del LPH.

3.1.2 Efectos y seguridad de la administración periconcepcional de suplementos de folato para la prevención de los defectos congénitos. (De-Regil, Fernández-Gaxiola, & Peña-Rosas J., 2010)

Dentro de este artículo de la base de datos Cochrane, se realiza la comparación de artículos respeto a las deficiencias del tubo neural (DTN), siendo un total de 5 ensayos de los cuales hablan totalmente de la toma de ácido fólico antes y durante los primeros tres meses de embarazo para poder logar un porcentaje de desaparición de estos defectos (de un 54% a un 72%) (De-Regil, Fernández-Gaxiola, & Peña-Rosas J., 2010). Sin embargo dentro del estudio se trata de prevenir el DTN, ignorando el efecto que tiene en otras deformaciones como el LPH o espina bífida.

CAPÍTULO 4. ADQUISICIÓN DE LOS DATOS

4.1 Recolección de la información

La clínica Rehabilitación Dental Especializada, tiene como especialidades médicas las siguientes: Ortodoncia, Ortopedia Dentofacial y Maxilofacial. Se resuelven diferentes tipos de problemas dentales y muchos de los pacientes que son atendidos llevan un seguimiento, en ocasiones, de las 3 áreas. Sin embargo, debido a que la doctora dueña de la clínica, la Dra. Cecilia Riojas Flores, se especializa desde hace más de 15 años en niños de LPH, también atiende a personas con esta malformación. Para poder generar un ambiente de muestra parecido al que se ha hecho en anteriores estudios de manera clínica (De-Regil, Fernández-Gaxiola, & Peña-Rosas J., 2010) se tomaron todos los pacientes de la clínica que sin problemas de LPH y aquellos que presentan LPH sin ningún otro síndrome o que sea hereditario, tomando en consideración lo siguiente el total es de aprox. 500 pacientes. No hubo una distinción de género, ni tampoco de edad; sin embargo de igual forma a aquellas familias se tomaron en cuenta también siempre y cuando ambos hermanos fueran atendidos en la clínica y se tomaron de manera independiente cada uno como un paciente más.

4.2 Limpieza de la información

Para evitar como resultado final datos inconsistentes e incoherentes de acuerdo a la hipótesis; se realizó un cuestionario para poder filtrar desde la toma de datos, aquellos pacientes con LPH pero que tienen algún síndrome o que tienen familiares con su misma deformación. Las preguntas de las cuales se formó y filtro la información de la base de datos, fueron las siguientes:

1. ¿Dentro de la familia se conoce algún miembro con LPH?
2. ¿La madre tomó ácido fólico antes y durante los primeros tres meses de embarazo?
3. ¿El paciente tiene LPH?

La base se compone de 4 columnas las cuales son: GÉNERO, P1, P2, P3. El género no se toma como parte del proceso en total ya que la hipótesis abarca de manera general la muestra sin haces distinción entre hombres o mujeres, además salvo que la muestra es demasía pequeña no se dispondría de un margen recomendable para poder realizar todo el proceso de minería de datos. El total de registros es de 514, el cual se redujo a 436 los cuales son aquellos pacientes con o sin LPH pero sin antecedentes heredofamiliares.

4.3 Minado de datos

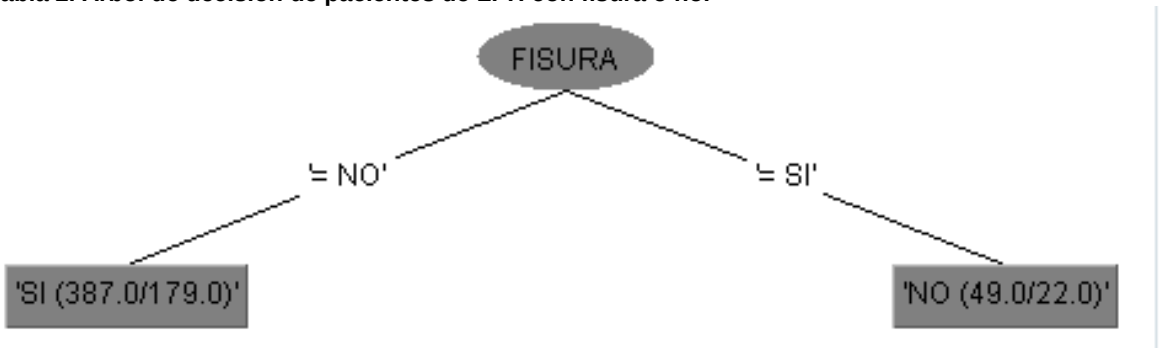
Una vez que se limpió la base de datos y acto seguido, se corrió el algoritmo de minería de datos C4.5 el resultado que arrojo fue dos clasificaciones en este caso fueron a y b donde a significa que Si había tomado el ácido fólico antes del embarazo y b significa que No había tomado el ácido fólico antes ni durante el embarazo. En la tabla 1 se muestra una referencia cruzada de los resultados arrojados por los algoritmos, donde se resaltan aquellas anomalías o casos específicos donde a pesar de haber tomado ácido fólico, se tiene presencia de la malformación del LPH y a su inversa aquellos casos en los cuales aun sin haber tomado ácido fólico no hay presencia de la malformación

Tabla 1. Matriz de resultados predictivos.

CON LPH	SIN LPH
208	22
183	23

En la tabla 1 se explican aquellos casos aislados de los resultados dentro del minado de datos los cuales generan confusión y son descartados por salir del patrón inicial que generan los demás resultados por ejemplo, factores genéticos. La matriz que el resultado predictivo usó, en este caso para conocer si saldrá con LPH aun tomando ácido fólico tiene 22 casos que no encajan correctamente en el patrón y por lo tanto se descartan estos datos por otras causas posibles.
En cuanto a los resultados de a 208 pacientes que tomaron el AF solo 22 se encontró el LPH y de los 183 que tomaron AF durante el embarazo, 23 si presentaron el LPH. A partir de estos resultados se hizo el siguiente árbol de decisión que indica si los pacientes presentaban una fisura de labio paladar hendido o bien, no la presentaban.

Tabla 2. Árbol de decisión de pacientes de LPH con fisura o no.



La figura 2 se presenta un árbol de decisión en el cual se indica que de los 436 registros se tienen 387 que no tienen LPH y si tomaron ácido fólico y dentro de esta misma rama se tiene que 179 no tienen LPH pero no tomaron ácido fólico; mientras que al observar la otra parte del esquema se puede observar que 49 casos que si presentaron LPH pero no tomaron ácido fólico y 22 elementos

que también presentaron LPH pero si tomaron ácido fólico. Por lo tanto, el 52% indica que no existe una fisura debido a que la madre tomo ácido fólico antes y durante el embarazo. El 47% asume que tienen la malformación por no tomar ácido fólico. El 1% si presentó la malformación a pesar de haber tomado ácido fólico antes y durante el embarazo.

CAPÍTULO 5. RESULTADOS Y SU INTERPRETACIÓN

Dentro de la siguiente tabla se encuentra la matriz de confusión[4], la cual nos determina que en el caso de aquellos predicciones que se realizan con las clases antes determinadas en el minado de datos. Podemos apreciar que las predicciones que se realizan en el minado tienen un margen de error, el cual nos indica que tan exactas son las predicciones basadas en lo que se ha aprendido de los patrones extraídos de los datos.

En este caso es un 53% que ha acertado en las predicciones que se han hecho dándonos que hay un 53% de posibilidad que el tomar ácido fólico no se presente el LPH.

Tabla 3. Matriz de confusión

CON LPH	SIN LPH	
208	22	CON LPH
183	23	SIN LPH

En la siguiente gráfica, podemos apreciar que se tienen 2 columnas (Fisura y No Fisura) las cuales indican en estos casos aquellos datos que resultan positivos en cuanto a la hipótesis que se ha planteado, en este caso, la parte señalada en color azul son aquellos casos que cumples al 100% y la parte en rojo nos indica aquellos que definitivamente no cumplen con esta regla, de la misma forma ocurre con los datos de aquellos casos de los cuales no hubo ingesta de ácido fólico.

[4] Herramienta visual de la inteligencia artificial, la cual se utiliza para aprendizaje supervisado; sin embargo permite ver si la herramienta está confundiendo clases dentro de las predicciones de los resultados.

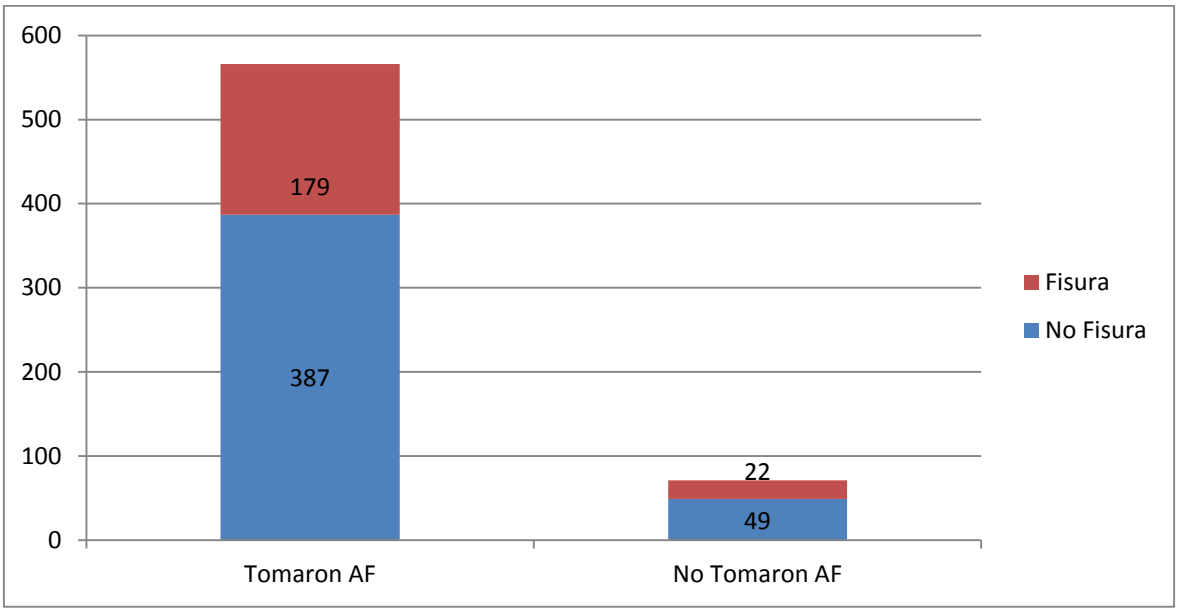


Gráfico 1. Pacientes que tomaron Ácido Fólico (AF) antes del embarazo y pacientes que no lo tomaron

Como resumen tenemos la siguiente gráfica, que nos determina en porcentajes los casos que representan de forma total y confirman la hipótesis, como a su vez tenemos también los que la rechazan por completo, todo esto da el total de los casos y el conocimiento que se obtiene después de realizar la limpieza de la información y el minado de datos.

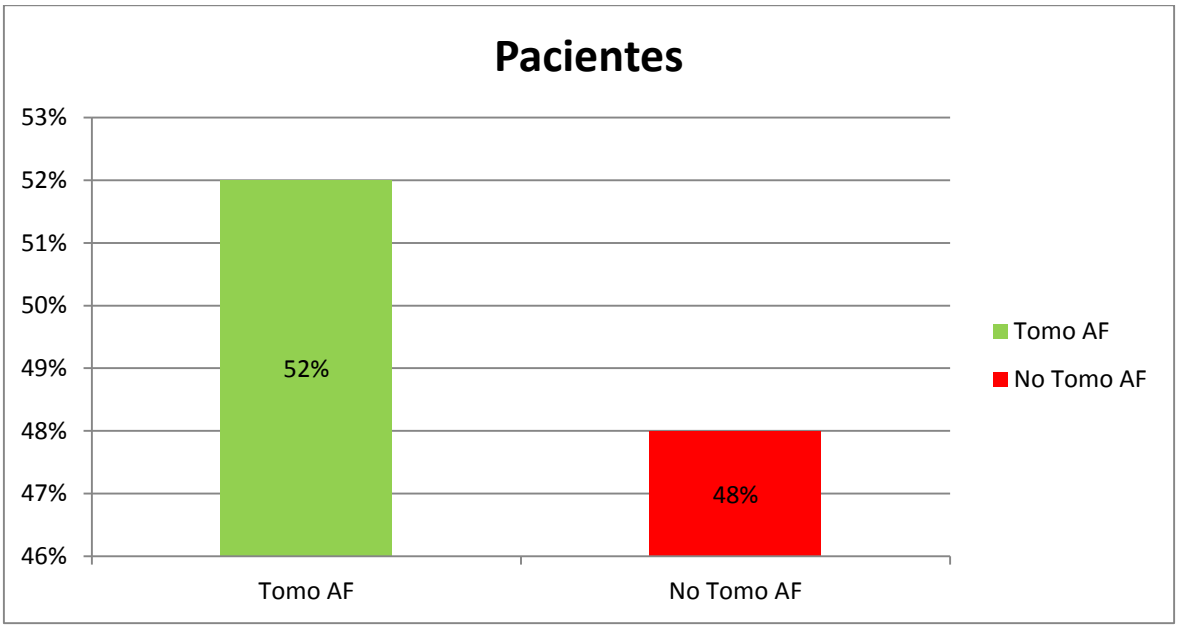


Gráfico 2. Conjunto total de los pacientes con presencia de LPH y sin LPH

Por lo tanto, la investigación se encuentra en un punto medio donde la hipótesis no puede ser confirmada pero tampoco la hipótesis nula. A pesar de quizás tener un porcentaje levemente elevado en casos positivos, en realidad la diferencia de estos dos es por muy

poco parecida, haciendo que ninguna de las dos hipótesis puedan ser confirmadas con seguridad.

Esto se demuestra de la misma forma en las predicciones que el minado realiza en donde las diferencias son iguales en base a los últimos resultados; donde inclusive realizando más predicciones con diferentes configuraciones o patrones, se llega a un punto donde es totalmente a la mitad, es decir 50% de ambas opciones.

CONCLUSIONES

En base a la pregunta de investigación se demostró con este trabajo de que la ingesta del ácido fólico evita el defecto congénito de labio paladar hendido. Se utilizó la minería de datos para extraer la información de estos pacientes mediante dos algoritmos: ID3 y C4.5 para comprobar esta hipótesis. Al término de este estudio se encontró que no es posible determinar a un margen considerable esta hipótesis, dejando a la investigación en exactamente un pun intermedio sin conclusión final la hipótesis o la nulificación de esta.

Como trabajo futuro se encuentra el utilizar esta investigación de base, de esta forma se modificarán variables posiblemente, como también se aumentarán los registros de la base de datos. Se pretende de igual forma utilizar diferentes algoritmos, implementar métodos genéticos y valores relacionados a estos y formar colaboraciones conjuntas entre bancos de datos genéticos y otros investigadores.

De igual forma se cierran vertientes para poder resolver otro tipo de preguntas relacionadas a esta investigación como es: ¿Qué factor determina la presencia de esta malformación?. De tal modo que se pueda utilizar esta investigación para considerar otras variables aun más grande y con mayor dificultad de cálculo y determinación.

BIBLIOGRAFÍA

Abdul, J., Sibi, S., & Aswin, R. B. (2012). Artificial Neural Network Based Detection of Skin. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering , 1*.

Alan, R. (2011). *Hendiduras Orales-Faciales.* EBSCO Publishing.

American Cancer Society. (2014). Cancer Facts and Figures 2014. *American Cancer Society* .

Asociación Mexicana de Cirugía Plástica, Estética y Reconstructiva, AC. (2003). Análisis de la incidencia, prevalencia y atención del labio y paladar hendido en México. *Cirugía Plástica , 13* (1), 35-39.

Braga, L. P. (2009). *Introducción a la Minería de Datos (Spanish Edition).* Río de Janeiro: E-papers Servicios Editoriales.

Braga, Vieira, L., & Ortiz Valencia, L. (2005). *Data Mining: Practical Machine Learning Tools and Techniques.* San Francisco: Morgan Kaufmann Publishers.

Celebi, M. E., Kingravi, H. A., Uddin, B., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., et al. (2007). A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics , 31(6)*, 362-373.

Charles M. Woolf, R. M. (1980). *A Genetic Study of Cleft Lip and Palate in Utah.* Tempe, Arizona: Deaprtment of Zoology, Arizona State University.

Chung, G. H. (1974). A Genetic Study of Cleft Lip and Palate in Hawai. *Am J Hum Genet* , 26:162,176.

Cobourne, M. T. (2012). Cleft Lip and Palate: Epidemiology, Aetiology and Treatment. *Frontiers of Oral Biology , 16*, 1-156.

DANIEL BARBARA, S. J. (2002). *APPLICATIONS OF DATA MINING IN COMPUTER SECURITY.* Virginia: Kluwer Academic Publishers.

De-Regil, I., Fernández-Gaxiola, A., & Peña-Rosas J., D. T. (2010). Efectos y seguridad de la administración periconcepcional de suplementos de folato para la prevención de los defectos congénitos. *Cochrane Database Systematic Reviews 2010* , Issue 10. Art. No.: CD007950. DOI: 10.1002/14651858.CD007950.

Ercal, F., Chawla, A., Stoecker, W. V., Lee, H. C., & Moss, R. H. (1994). Neural network diagnosis of malignant melanoma from color images. *Biomedical Engineering, IEEE Transactions , 41(9)*, 837-845.

Fernández Aldana, L. (27 de Junio de 2005). Principios de Data Mining. Puebla, Puebla, México.

Flores, D. C. (2009). *MODIFICACIÓN DEL ANCLAJE INTRABUCAL.* D.F., México: Asociación Odontológica Mexicana para la Enseñanza y la Investigación.

Ignacio Zarante, L. F. (2010). Frecuencia de malformaciones congénitas: evaluación y pronóstico de 52.744 nacimientos en tres ciudades colombianas. *Biomédica* , 65-71.

Janert, P. K. (2011). *Data Analysis with Open Source Tools.* California: O'Reilly Media.

Kamber, J. H. (2006). *Data Mining:Concepts and Techniques.* San Francisco: Services Manager Simon Crump.
Kantardzic, M. (2003). *Data Mining: Concepts, Models, Methods, and Algorithms.* The Institute of Electrical and Electronics Engineers.
Karlind T. Moller, C. D. (1990). *A Parent´s Guide to Cleft Lip and Palate.* Minneapolis, USA: University of Minnesota.
Kimball, R., Reeves, L., Warren, M. R., Mundy, J., & Becker, B. (2007). *The Data WareHouse Lifecycle Toolkit.* Wiley.
Lee, T. K., & Claridge, E. (2005). Predictive power of irregular border shapes for malignant melanomas. *Skin Research and Technology , 11(1)*, 1-8.
Lee, T. K., McLean, D. I., & Stella Atkins, M. (2003). Irregularity index: a new border irregularity measure for cutaneous melanocytic lesions. *Medical image analysis , 7(1)*, 47-64.
Linoff, G. S. (2008). *Data Analysis Using SQL and Excel.* Indiana: Wiley Publishing.
Maglogiannis, I., & Doukas, C. N. (2009). Overview of advanced computer vision systems for skin lesions characterization. *Information Technology in Biomedicine, IEEE Transactions on , 13(5)*, 721-733.
Maynor Alfonso García-López, M. d.-R.-E. (2010). Diagnóstico prenatal de paladar hendido mediante ultrasonografía 3D. *78* , 626-632.
Michael Mars, D. S. (2008). *Management of Cleft Lip and Palate in the Developing World.* England: John Wiley & Sons Inc.
Moore, P. S. (2004). Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. *Human Molecular Genetics* , 73-81.
Ogorzałek, M., Nowak, L., Surówka, G., & Alekseenko, A. A. Modern techniques for computer-aided melanoma diagnosis. (DOI: 10.5772/23388).
Pang-Ning Tan, M. S. (2006). *Introduction to Data Mining.* Pearson Addison-Wesley.
Penagos, F. M. (2012). *CONSTRUCCION DE CARTILLA EDUCATIVA PARA LA PROMOCION DE LA SALUD ORAL Y PREVENCION DE LA APARICION DE NUEVOS CASOS DE LABIO Y/O PALADAR HENDIDO NO SINDROMICO, EN LA POBLACION DE TARAPACÁ, DEPARTAMENTO DEL AMAZONAS, COLOMBIA.* Bogotá, Colombia: Universidad Nacional de Colombia.
Reynolds, J., & Holbrook, P. (1 de Enero de 1991). *The Internet Engineering Task Force RFC 1244*. Retrieved 2013 de Febrero de 2013 from http://www.ietf.org/rfc/rfc1244.txt?number=1244. Julio 1991
Rubio, N. L. (2009). *Factores genéticos en el labio leporino.* Eroski Consumer.
Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representation by back-propagation errors. *NATURE , 323*.
Russell, M. A. (2011). *Mining the Social Web.* California: O'Reilly Media.
Salud, B. (21 de 10 de 2012). Arranca el mayor estudio sobre la causa del labio leporino. New York, USA.
Simmon, T. C. (2009). *DataWarehousing for Dummies.* Indianapolis: Wiley Publishing.
Sociedad Mexicana de Oncología. (2005). *Gaceta Mexicana de Oncología* (Vol. 4). (D. M. Sánchez, Ed.) México, DF: MASSON DOYMA MÉXICO.

Tolarova, D. M. (1998). *Fisura Labial - Hendidura de Paladar.* Caracas, Venezuela: ROTAPLAST Internacional.
Torre, E. L., Caputo, B., & Tommasi, T. (2010). Learning methods for melanoma recognition. *International Journal of Imaging Systems and Technology , 20(4)*, 316-322.
Vesna Kozel J. D.D.S., P. D. (1999). Changes Produced by Presurgical Orthopedic Tratement Before Cheiloplasty in Cleft Lip and Palate Patients. *Cleft Palate-Craniofacial Journal , 36* (6), 517-521.
Werbos, P. J. (1990). Backpropagation Through Time: What it Does and How to Do it. *proceeding of the IEEE , 79* (10).
Wyszynski, D. F. (2002). *Cleft Lip & Palate from origin to treatment.* New York: Oxford University Press.