

DESARROLLO DE UNA INTERFAZ GRÁFICA DE USUARIO PARA LA
CLASIFICACIÓN DEL TIPO DE FENOTIPO METABÓLICO QUE PRESENTA UN
PACIENTE UTILIZANDO LA BASE DE DATOS NHANES

LUIS ENRIQUE RAMÍREZ GUTIÉRREZ

UNIVERSIDAD SIMÓN BOLÍVAR

FACULTAD DE INGENIERÍA

INGENIERÍA DE SISTEMAS

SAN JOSÉ DE CÚCUTA, NORTE DE SANTANDER

2017

DESARROLLO DE UNA INTERFAZ GRÁFICA DE USUARIO PARA LA
CLASIFICACIÓN DEL TIPO DE FENOTIPO METABÓLICO QUE PRESENTA UN
PACIENTE UTILIZANDO LA BASE DE DATOS NHANES

LUIS ENRIQUE RAMÍREZ GUTIÉRREZ

Ing. JOSÉ GERARDO CHACÓN

Tutor Disciplinar

Trabajo de Investigación

Formación Investigativa III

UNIVERSIDAD SIMÓN BOLÍVAR

FACULTAD DE INGENIERÍA

INGENIERÍA DE SISTEMAS

SAN JOSÉ DE CÚCUTA, NORTE DE SANTANDER

2017

Tabla de Contenido

	Pág.
Introducción.....	10
1. Planteamiento del problema	12
1.1 Formulación del problema	13
1.2 Delimitación.....	13
1.2.1 Delimitación espacial.	13
1.2.2 Delimitación temporal.....	14
1.3 Objetivos.....	14
1.3.1 Objetivo general.	14
1.3.2 Objetivos específicos.	14
1.4 Justificación	14
2 Marco Referencial	16
2.1 Antecedentes.....	16
3 Marco Teórico	21
3.1 Interfaz Gráfica de Usuario (GUI).....	21
3.2 Fenotipo Metabólico:	22
3.3 Máquinas de Soporte Vectorial (SVM):	23
3.4 Encuesta Nacional de Salud y Nutrición NHANES:	24
3.5 Enfermedades No Transmisibles (ENT):.....	26
3.6 Aprendizaje supervisado:.....	26
3.7 Criterios de Evaluación.....	26

3.7.1 Verdadero positivo	27
3.7.2 Falso negativo	27
3.7.3 Sensibilidad	27
3.7.4 Especificidad	27
4. Revisión bibliográfica	29
4.1 La Salud	29
4.2 Los Fenotipos Metabólicos	29
4.3 Enfermedades no Transmisibles	30
4.4 Obesos enfermos	31
4.5 Obesos sanos	32
4.6 Delgado enfermo	33
5. Metodología de la investigación	35
5.1 Tipo de investigación	35
5.2 Instrumentos y técnicas	35
5.2.1 Ficha bibliográfica.	36
6. Desarrollo del proyecto	37
6.1 Apropiación de la información	37
6.2 Población y muestra	38
6.3 Libro de códigos	40
6.4 Verificación del funcionamiento de la GUI implementada computacionalmente	77
6.5 Fundamentación	82
6.6 Personas y Roles del Proyecto	83
7. Conclusiones	88
Bibliografía	89

Lista de Cuadros

	Pág.
Cuadro 1. Sexo: Sexo del encuestado.	42
Cuadro 2. Estado Civil: Estado civil del encuestado	42
Cuadro 3. Raza: Raza del encuestado	43
Cuadro 4. EstSocioec: Estatus socioeconómico por Graffar	44
Cuadro 5. EstEduc: grado educativo mayor	44
Cuadro 6. HabTab: Hábito tabáquico	45
Cuadro 7. ConAlcohol: Consumo de alcohol	46
Cuadro 8. Peso: Peso en Kg	46
Cuadro 9. Talla: Talla en metros	47
Cuadro 10. IMC: Índice de masa corporal	47
Cuadro 11. IMCclasifOMS: Clasificación ponderal según el IMC según la organización Mundial de la Salud (ordinal)	48
Cuadro 12. CABD: Circunferencia abdominal en centímetros	48
Cuadro 13. GBasal: Glicemia en ayuno	49
Cuadro 14. CTotal: Colesterol total	49
Cuadro 15. TAG: Triacilglicéridos	50
Cuadro 16. HDL: Colesterol de alta densidad o colesterol “bueno”	50
Cuadro 17. VLDL: Colesterol de muy baja densidad (TAG/5)	51

Cuadro 18. LDL: Colesterol de baja densidad o colesterol malo	51
Cuadro 19. PAS: Presión arterial sistólica	52
Cuadro 20. PAD: Presión arterial diastólica	52
Cuadro 21. PAM: Presión arterial media	53
Cuadro 22. ColNoHDL: Colesterol NO HDL	53
Cuadro 23. IndiceTAGHDL: División entre los TAG/HDL	54
Cuadro 24. HDLbajasIDF: HDL bajas	54
Cuadro 25. CABDaltaIDF: Circunferencia abdominal Alta	55
Cuadro 26. TAGaltaIDF: Triacilglicéridos altos	56
Cuadro 27. GLICEMIAaltaIDF: Glicemia elevada	56
Cuadro 28. PAaltaIDF Presión arterial elevada	57
Cuadro 29. GLICEMIAaltaATPIII2003: Glicemia alta según criterios del ATPIII (USA)	58
Cuadro 30. CABDaltaATPIII2003: Circunferencia abdominal alta	58
Cuadro 31. CABDaltaATPIII2005: Circunferencia abdominal alta ATPIII 2005	59
Cuadro 32. SMIDF2005: Diagnóstico de SM según la IDF 2005	60
Cuadro 33. SMATPIII2003: Diagnóstico de SM según ATPIII 2003	60
Cuadro 34. SMATPIII2005: Diagnóstico de SM según ATPIII 2005	61
Cuadro 35. RazaLast: Categorías de raza	61
Cuadro 36. METsTotalesAct	62

Cuadro 37. PatronActFisica	62
Cuadro 38. CAaltaNUEVOSTUNTOS: Circunferencia abdominal alta puntos de corte maracaibo	63
Cuadro 39. IR	64
Cuadro 40. CombinacionesTODAS Combinaciones de los criterios de síndrome metabólico	64
Cuadro 41. PAJNC7nueva: Clasificación de personas según su presión arterial	66
Cuadro 42. EstatusGlicemico: Estado de la glicemia en ayuno	67
Cuadro 43. SMIDF2009: Diagnóstico de SM según criterios del 2009 (los más nuevos)	67
Cuadro 44. TallaCM: Talla en cm	68
Cuadro 45. IndiceCinturaAltura: índice cintura/altura	69
Cuadro 46. IndiceCinturaAlturaLOG: Transformación logarítmica del anterior	69
Cuadro 47. AlteracionesLipidicas: Frecuencia de alteraciones en los lípidos sanguíneos	70
Cuadro 48. AFtrabajoMETs	71
Cuadro 49. AFtransporteMETs: Actividad física en METs esfera transporte	71
Cuadro 50. AFhogarMETs: Actividad física en METs esfera hogar	72
Cuadro 51. AFocioMETs: Actividad física en METs esfera ocio	73
Cuadro 52. VAI: Índice visceral adiposo (mide el riesgo de diabetes)	74
Cuadro 53. LogVAI	74
Cuadro 54. MRFA: Agregación de múltiples factores de riesgo (Menos de 2 y 2 o más)	75
Cuadro 55. TGI	75

Cuadro 56. TGIDico	76
Cuadro 57. Fenotipoclusters: Fenotipos metabólicos por análisis de clúster	76
Cuadro 58. Personas y roles del trabajo.	83
Cuadro 59. Historias de usuario y criterios de aceptacion.	84
Cuadro 60. Sprint 1 Historias de Usuario FM-001	85
Cuadro 61. Sprint 2 Historias de Usuario FM-002	85
Cuadro 62. Sprint 3 Historias de Usuario FM-003	85
Cuadro 63. Sprint 4 Historias de Usuario FM-004	86
Cuadro 64. Tabla de producto	86

Lista de Figuras

	Pág.
Figura 1. Interfaz Gráfica de Usuario (GUI)	22
Figura 2: Clasificación de las SVM	24
Figura 3. Logo de la Encuesta Nacional de la Salud y la Nutrición. NHANES.	25
Figura 4. Ficha Bibliográfica	36
Figura 5. Base de Datos NHANES, Data Original.	39
Figura 6. Base de Datos Seleccionada para el Clasificador	41
Figura 7. GUI Cargar Base de Datos.	78
Figura 8. GUI Base de Datos Cargada Exitosamente.	78
Figura 9. GUI Clasificador Entrenado.	79
Figura 10. GUI Validación del Clasificador.	80
Figura 11. Vector de Datos de un Paciente para Clasificar.	81
Figura 12. Clasificación del Tipo de Fenotipo Metabólico con un Paciente.	82
Figura 13. Grafica del producto	87

Introducción

La presente investigación, propone un prototipo de una interfaz gráfica de usuario mediante la cual se puede entrenar, validar y clasificar con máquinas de soporte vectorial, una matriz de datos recolectados en una base de datos medica realizada en estados unidos.

Betancourt (2005) afirma. “La teoría de las Máquinas de Soporte Vectorial (SVM por su nombre en inglés Support Vector Machines) es una nueva técnica de clasificación y ha tomado mucha atención en años recientes. La teoría de la SVM está basada en la idea de minimización de riesgo estructural”. Las máquinas de soporte vectorial nos permiten seguir un modelo de ayuda, la cual nos permite seguir un procedimiento de trabajo que favorece en la investigación, este modelo se basa en la clasificación y la regresión. En nuestro caso, el prototipo que se propone encaja con el modelo de clasificación, ya que es el objetivo principal de este trabajo.

Esta investigación se realizó con el fin de profundizar en el procesamiento de datos, siguiendo un modelo de máquinas de soporte vectorial que nos indica el procedimiento de tomar un conjunto de datos, utilizando una interfaz gráfica podemos, entrenar los datos para después poder simularlos y obtener unos resultados.

Además, este documento describe un planteamiento del problema donde se resalta el interés y para que se realiza esta investigación, la descripción de los datos tomados, y de donde se obtuvieron estos datos.

Seguido se mencionan algunos referentes teóricos, los cuales ayudaron a conocer los modelos que se pueden seguir utilizando máquinas de soporte vectorial, su funcionamiento y su debido y correcto procedimiento, con el fin de poder definir unos objetivos que definirán la dirección seguida del proyecto de investigación.

Para finalizar se muestra el producto realizado, su funcionamiento utilizando una interfaz gráfica, y siguiendo un modelo de trabajo con máquinas de soporte vectorial.

1. Planteamiento del problema

La salud es un tema de suma importancia a nivel global, por eso los países y más los países subdesarrollados invierten grandes cantidades de dinero en este campo y en investigaciones que ayuden a fortalecer ese conocimiento que se tiene sobre el impacto que causa la salud en los individuos y como combatir en caso de que surja alguna enfermedad que pueda poner en riesgo la sociedad.

En caso particular las enfermedades no transmisibles como los fenotipos metabólicos, han causado gran impacto en la salud de los individuos en la sociedad, ya que estas enfermedades causan un gran porcentaje de las muertes. Según la organización mundial de la salud “Las enfermedades crónicas son enfermedades de larga duración y por lo general de progresión lenta. Las enfermedades cardíacas, los infartos, el cáncer, las enfermedades respiratorias y la diabetes son las principales causas de mortalidad en el mundo, siendo responsables del 63% de las muertes. En 2008, 36 millones de personas murieron de una enfermedad crónica, de las cuales la mitad era de sexo femenino y el 29% era de menos de 60 años de edad”

En nuestro caso se quiere analizar un conjunto de datos alojados en una matriz compuesta por la información personal de cientos de pacientes encuestados por la NHANES (Encuesta nacional de salud y nutrición), allí se guardan resultados de muchos exámenes que algunos son complicados de realizar y otros son costosos como el examen de la insulina.

En esta investigación se propone un sistema que pueda descartar estos exámenes y variables que en algunos casos son difíciles de conseguir o que son costosas como la insulina, para esto se utiliza un modelo de máquinas de soporte vectorial que pueda tomar este conjunto de datos y con un entrenamiento supervisado pueda determinar nuestras variables de resultado que es este caso son los tipos de fenotipos metabólicos, sin tener que tomar el examen de la

insulina puesto que es costoso y muchos pacientes de bajos recursos no tendrán acceso a este examen.

En el transcurso del tiempo la sociedad se está preocupando por su salud, ya que hoy en día muchos factores están afectando y perjudicando su bienestar, ya se están realizando investigaciones y haciendo análisis correspondientes sobre cómo combatir estas enfermedades no transmisibles y lo que es en este caso particular lo que son los fenotipos metabólicos, pero lo que los individuos no tiene en cuenta es que estas enfermedades son a largo tiempo y ya para cuando descubran esta enfermedad puede ser un poco tarde, para esto se hace uso de la creación de interfaz de usuario (GUI), utilizando la herramienta Matlab 2015 donde nos permite trabajar con modelos de clasificación y editores de interfaz de usuario.

1.1 Formulación del problema

De acuerdo al planteamiento del problema descrito se puede considerar como pregunta de investigación la siguiente: ¿Como se podría clasificar los diferentes tipos de fenotipos metabólicos con los pacientes establecidos en la base de datos NHANES?

1.2 Delimitación

1.2.1 Delimitación espacial.

Este proyecto se realizará tomando como muestra la información de la base de datos certificada NHANES, teniendo en cuenta que se tomaran como criterio los datos de pacientes tomados al azar en una encuesta realizada en estados unidos, con la ayuda del método de máquinas de soporte vectorial se clasificara un paciente sin la necesidad de contener el examen de la insulina.

1.2.2 Delimitación temporal.

Esta investigación se desarrollará durante el periodo académico 2017-2, en el contexto de la asignatura Formación Investigativa III.

1.3 Objetivos

1.3.1 Objetivo general.

Desarrollar una interfaz gráfica de usuario para la clasificación del tipo de fenotipo metabólico que presenta un paciente utilizando la base de datos NHANES.

1.3.2 Objetivos específicos.

Realizar investigación sobre fenotipos metabólicos utilizando motores de búsqueda especializados para la identificación de los conceptos abordados en la revisión bibliográfica.

Identificar los datos relevantes, seleccionando los datos que se utilizaran de la base de datos NHANES, para la correcta clasificación en el modelo de máquinas de soporte vectorial.

Diseñar la interfaz gráfica con métodos de clasificación de máquinas de soporte vectorial para la identificación de los fenotipos metabólicos considerando la base de datos NHANES.

1.4 Justificación

Como se describe en el planteamiento del problema, la salud de las personas es de suma importancia, por lo que los países invierten grandes cantidades de dinero a este campo, su área investigativa es cada día más grande y de igual manera produce grandes soluciones. En esta investigación se plantea un problema que perjudica la salud de las personas, nos referimos a las enfermedades no transmisibles, como lo son los fenotipos metabólicos, estas alteraciones de salud, son enfermedades que se pronuncian a largo plazo, lo cual hace más difícil su detección.

Además, para poder identificar estas alteraciones de salud se requiere de exámenes que pueden ser costosos, como por ejemplo el examen de la insulina que tiene un costo aproximado de 200 mil pesos, un costo que para pacientes de recursos bajos es difícil de conseguir.

Esta investigación propone un sistema de clasificación de estas alteraciones metabólicas, tomando como base la información establecida en la base de datos NHANES, y haciendo un filtro de los datos que no serán considerados para la clasificación, también excluyendo como parte principal de la investigación el examen de la insulina para evitar este examen costoso y tener una viabilidad favorable en esta investigación.

2 Marco Referencial

2.1 Antecedentes

La investigación que se propone tiene en cuenta antecedentes en el campo de la salud a nivel local, nacional e internacional que ayudan a dar un concepto más claro a la idea que se propuso en este documento.

Título de la investigación: SISTEMA DE VIGILANCIA DE ENFERMEDADES NO TRANSMISIBLES EN CUBA

Autores: Dr. Ricardo Batista Moliner, Dr. Orlando Landrove Rodríguez

Año: 2000

Resumen: Las enfermedades no transmisibles constituyen las primeras causas de muerte en nuestro país, más de las 3/4 partes de los fallecidos corresponden a las llamadas "enfermedades crónicas no transmisibles". La vigilancia ha devenido una función básica de la Salud Pública y en la actualidad es una de las vías para alcanzar mejores resultados en la atención a este problema. El desarrollo del sistema de salud en Cuba y las experiencias internacionales en este campo permiten avanzar aún más en el terreno de la vigilancia y el control de estas entidades. La complejidad de la epidemiología de estas enfermedades crea dificultades para el desarrollo de la vigilancia, lo que requiere de la integración de todas las áreas desde el nivel primario de atención. Para la concepción y organización del sistema de vigilancia de las enfermedades no transmisibles en nuestro país se revisaron las experiencias internacionales y los resultados del desarrollo de la vigilancia de enfermedades infecciosas y de la salud pública en general. Asimismo, se realizaron consultas con expertos en el tema que aportaron elementos de valor al trabajo. Como resultados de esa labor se presentó una

propuesta que se discutió por los responsables de ejecutar las actividades del mismo, en todo el país. En la concepción de la vigilancia de estas afecciones quedaron establecidas 2 vertientes fundamentales para el monitoreo de los eventos relacionados con estas entidades: los factores de riesgo y los daños (morbilidad y mortalidad). Se definieron, además, los problemas a vigilar, los datos, las fuentes de los cuales se obtienen, la periodicidad, el mecanismo de disseminación y la retroalimentación de los resultados.

Título de la investigación: SISTEMA DE MONITOREO Y EVALUACIÓN DE LA CALIDAD DE LA ATENCIÓN MÉDICA EN LAS ENFERMEDADES NO TRANSMISIBLES

Autores: Dra. Milvia Ramírez Rodríguez, Lic. Diana Rodríguez Gabaldá, Dr. Orlando Landrove

Año: 2001

Resumen: Con esta metodología pretendemos que se evalúe la calidad de la atención médica que se brinda a los pacientes con enfermedades no transmisibles (asma bronquial, hipertensión arterial y diabetes mellitus) en las instituciones de salud a nivel primario y secundario, pues estas son las primeras causas de muerte y discapacidad en nuestro país. Esto permitirá determinar en qué medida las acciones de salud que se realizan cumplen el objetivo fundamental de mantener, restaurar o promover salud. Las enfermedades no transmisibles actualmente son causa del 60 % de la mortalidad y la discapacidad en el mundo. La mayoría de las personas afectadas (cerca del 70 %) tienen menos de 45 años y el 30 % menos de 15 años de edad. Todos los países desarrollados muestran esta tendencia incluyendo el nuestro, donde 3 de cada 4 muertes que ocurren

son atribuidas a las enfermedades no transmisibles, se destaca el hecho de que éstas constituyen 9 de las 10 primeras causas de muerte.

Título de la investigación: CLASIFICACIÓN DE DATOS DE IDS BASADA EN LS-SVM CON PSO

Autores: Eduardo Mauricio león García

Año: 2010.

Resumen: Muchas instituciones manejan enormes cantidades de información que circula por sus redes como paquetes de datos, que son vulnerables frente a ataques informáticos de diversa índole. Una opción para mitigar el riesgo de estos ataques es usar sistemas de detección de intrusiones (IDS), los que monitorean las redes constantemente en busca de anomalías. Sin embargo, los ataques son diversos y cambiantes, en consecuencia, siempre existirá la necesidad de mejorar los procesos de clasificación. En este contexto, el objetivo del presente proyecto de título es desarrollar e implementar la técnica LS-SVM con PSO para la detección de intrusiones. Para llevar a cabo la construcción y evaluación de cada modelo de clasificación, cada proceso es dividido en dos etapas: entrenamiento y pruebas. En la primera, el algoritmo de clasificación es optimizado por PSO, de tal manera, que, una vez construido el modelo, cuenta con la capacidad de responder bien frente a cualquier dato, incluyendo los desconocidos. En la segunda etapa, se realizan pruebas para medir la bondad del modelo en base a una serie de índices. Un punto fundamental en este desarrollo fue que los conjuntos de datos, de las dos etapas, tuvieran la menor cantidad de datos en común. De

esta manera se pudo evaluar efectivamente la capacidad de generalización de cada modelo.

Título de la investigación: CLASIFICACIÓN DE LOS FENOTIPOS DE SÍNDROME DE OVARIO POLIQUÍSTICO DE ACUERDO A LOS CRITERIOS DE ROTTERDAM: ¿UNA CONDICIÓN ESTÁTICA O VARIABLE?

Autores: Bárbara Echiburúa, Amanda Ladrón de Guevara, Cecilia Pereira.

Año: 2014.

Resumen: El síndrome de ovario poliquístico (SOP) es una enfermedad endocrino-metabólica común en la mujer en edad reproductiva¹ y afectaría aproximadamente 6-8% de ellas, dependiendo del criterio de diagnóstico utilizado. EL SOP se caracteriza por hiperandrogenismo, oligo-ovulación crónica y ovarios de aspecto poliquísticos, además, frecuentemente se asocia con resistencia insulínica (RI) y obesidad. Es probable que este síndrome tenga un origen genético^{3,4}, el cual sería influenciado por factores ambientales.

Título de la investigación: COMPORTAMIENTO DE LAS ENFERMEDADES CRÓNICAS NO TRANSMISIBLES EN ADULTOS MAYORES

Autores: Juan Carlos Llibre GuerraI; Milagros A. Guerra HernándezII; Elianne Perera Miniet.

Año: 2008.

Resumen: Con el incremento de la expectativa de vida y el consecuente aumento de la proporción de personas de 60 años y más, también se eleva el porcentaje de la población que padece de enfermedades crónicas no transmisibles, las que representan las primeras

causas de muerte en los países desarrollados y en Cuba. Se presenta un estudio descriptivo de corte transversal, realizado en el período comprendido entre marzo de 2007 y marzo de 2008 en pacientes de 65 años y más en el Policlínico "27 de noviembre" del municipio Marianao, con el objetivo de identificar el comportamiento de estas enfermedades. Se estudiaron 300 ancianos, la información se obtuvo de las bases de datos del estudio de investigación en demencias 10/66, con las tasas de prevalencia de las principales enfermedades no transmisibles que afectan a este grupo poblacional. La hipertensión arterial constituyó la enfermedad de mayor frecuencia con un 55 %, seguida de las enfermedades del corazón (32,3 %) y la diabetes mellitus (18,3 %). La primera fue más frecuente en el sexo masculino, mientras que las otras predominaron en el femenino. El síndrome demencial presentó una frecuencia del 12 %, y la depresión estuvo presente en el 6,7 % de los pacientes estudiados.

3 Marco Teórico

3.1 Interfaz Gráfica de Usuario (GUI)

La interfaz gráfica de usuario, conocida también como GUI (del inglés graphical user interface) es un programa Informático que actúa de interfaz de usuario, utilizando un conjunto de imágenes y objetos gráficos para representar la información y acciones disponibles en la interfaz. Su principal uso, consiste en proporcionar un entorno visual sencillo para permitir la comunicación con el sistema operativo de una máquina o computador.

Habitualmente las acciones se realizan mediante manipulación directa, para facilitar la interacción del usuario con la computadora. Surge como evolución de las interfaces de línea de comandos que se usaban para operar los primeros sistemas operativos y es pieza fundamental en un entorno gráfico. Como ejemplos de interfaz gráfica de usuario, cabe citar los entornos de escritorio Windows, el X-Windows de GNU/Linux o el de Mac OS X, Aqua.

En el contexto del proceso de interacción persona-ordenador, la interfaz gráfica de usuario es el artefacto tecnológico de un sistema interactivo que posibilita, a través del uso y la representación del lenguaje visual, una interacción amigable con un sistema informático.

GUIDE es un entorno de programación visual disponible en MATLAB para realizar y ejecutar programas que necesiten ingreso continuo de datos.

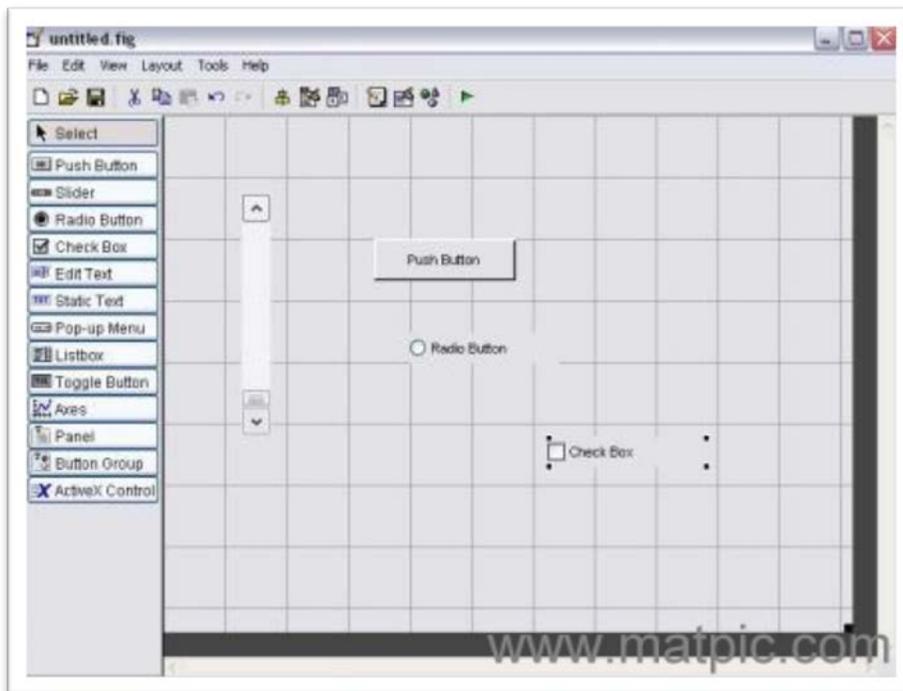


Figura 1. Interfaz Gráfica de Usuario (GUI)

3.2 Fenotipo Metabólico:

Según (Holmes E., Leng Loo R., 2008). “Los fenotipos metabólicos son los productos de las interacciones entre una variedad de factores dietéticos, otros estilos de vida / medio ambiente, intestino microbiano y genético. Utilizamos un enfoque analítico de exploración a gran escala para investigar la variación fenotipo metabólico a través y dentro de cuatro poblaciones humanas, con base en 1espectroscopia H RMN. Los metabolitos que discriminan entre las poblaciones se relacionan con los datos de los individuos sobre la presión arterial, un importante factor de riesgo de enfermedad coronaria y accidente cerebrovascular (las principales causas de mortalidad en todo el mundo)”.

3.3 Máquinas de Soporte Vectorial (SVM):

Según (Vapnik. V. 1995). “La red de vectores de soporte es una nueva máquina de aprendizaje para problemas de clasificación de dos grupos. La máquina implementa conceptualmente la siguiente idea: los vectores de entrada se asignan de forma no lineal a un espacio de características de muy alta dimensión. En este espacio de características se construye una superficie de decisión lineal. Las propiedades especiales de la superficie de decisión aseguran una alta capacidad de generalización de la máquina de aprendizaje. La idea detrás de la red de vectores de soporte se implementó previamente para el caso restringido donde los datos de entrenamiento se pueden separar sin errores”.

Según (BETANCOURT, 2005), La teoría de las Máquinas de Soporte Vectorial (SVM por su nombre en inglés Support Vector Machines) es una nueva técnica de clasificación y ha tomado mucha atención en años recientes. La teoría de la SVM está basada en la idea de minimización de riesgo estructural (SRM). En muchas aplicaciones, las SVM han mostrado tener gran desempeño, más que las máquinas de aprendizaje tradicional como las redes neuronales y han sido introducidas como herramientas poderosas para resolver problemas de clasificación.

Las SVM son básicamente clasificadores para 2 clases.

Se puede cambiar la formulación del algoritmo QP para permitir clasificación multiclase. Más comúnmente, los datos son divididos “inteligentemente” en dos partes de diferentes formas y una SVM es entrenada para cada forma de división. La clasificación multiclase es hecha combinando la salida de todos los clasificadores.

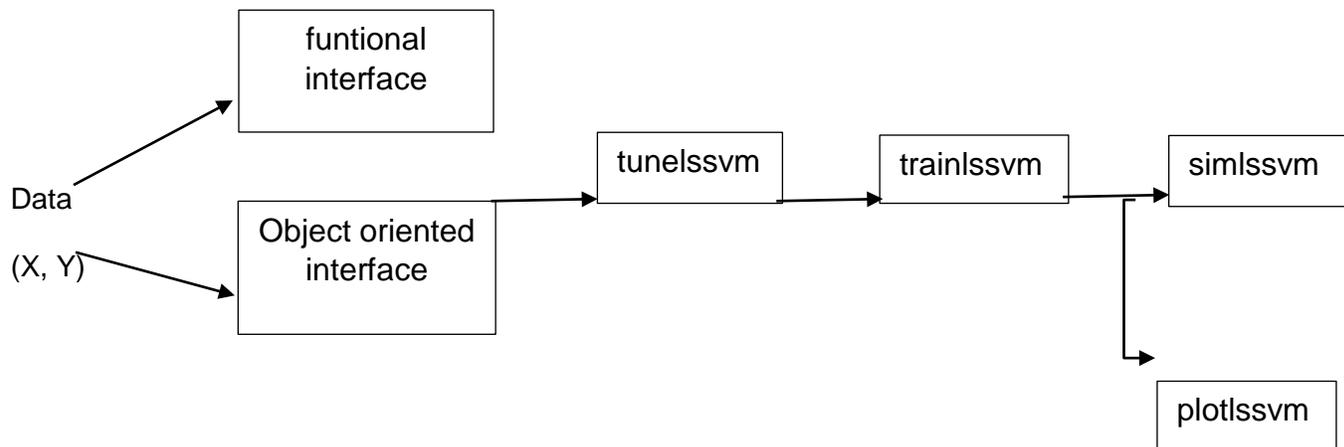


Figura 2: Clasificación de las SVM

3.4 Encuesta Nacional de Salud y Nutrición NHANES:

Nacional de salud y de la encuesta de examen de nutrición (NHANES) es un programa de estudios diseñados para evaluar la salud y el estado nutricional de adultos y niños en los Estados Unidos. La encuesta es el única que combina entrevistas y exámenes físicos. NHANES es un importante programa del centro nacional para estadísticas de salud (NCHS). NCHS es parte de los centros para el Control y prevención de enfermedades (CDC) y tiene la responsabilidad de producir vitales y estadísticas de salud de la nación.

El programa NHANES comenzó en la década de 1960 y ha sido llevado a cabo como una serie de estudios enfocándose en diferentes grupos de población o temas de salud. En 1999, la encuesta se convirtió en un programa continuo que tiene una orientación cambiante en una variedad de medidas de salud y nutrición a las necesidades emergentes. La encuesta examina una muestra nacionalmente representativa de aproximadamente 5.000 personas cada año. Estas personas se encuentran en los condados de todo el país, de los cuales 15 son visitados cada año.

La NHANES entrevista incluye preguntas demográficas, socioeconómicas, dietéticas y relacionados con la salud. El componente de la examinación consiste en medidas médicas, dentales y fisiológicas, así como pruebas de laboratorio administradas por personal médico altamente capacitado.

Los resultados de esta encuesta se utilizarán para determinar la prevalencia de las enfermedades más importantes y factores de riesgo para enfermedades. Información se utilizará para evaluar el estado nutricional y su asociación con salud promoción y prevención de enfermedades. Resultados de NHANES son también la base para normas nacionales para tales medidas como altura, peso y presión arterial. Datos de esta encuesta serán utilizados en los estudios epidemiológicos y la investigación en Ciencias de la salud, que ayudan a desarrollar políticas de salud pública sana, dirigir y diseñar programas de salud y servicios y expandir el conocimiento de la salud de la nación.

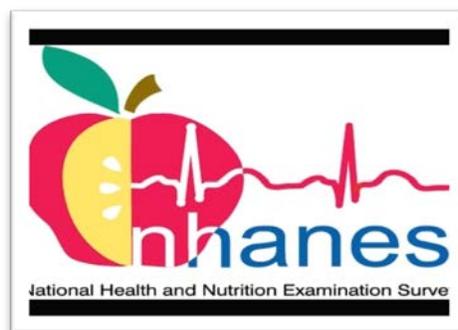


Figura 3. Logo de la Encuesta Nacional de la Salud y la Nutrición. NHANES.

3.5 Enfermedades No Transmisibles (ENT):

Según la Organización Mundial de La Salud, Las enfermedades no transmisibles - o crónicas - son afecciones de larga duración con una progresión generalmente lenta. Los cuatro tipos principales de enfermedades no transmisibles son:

Las enfermedades no transmisibles, o ENT, representan con diferencia la causa de defunción más importante en el mundo, pues acaparan un 63% del número total de muertes anuales.

Las enfermedades no transmisibles (ENT) matan a más de 36 millones de personas cada año. Cerca del 80% de las muertes por ENT se concentran en los países de ingresos bajos y medios.

3.6 Aprendizaje supervisado

El aprendizaje supervisado se caracteriza porque el proceso de aprendizaje se realiza mediante un entrenamiento controlado por un agente externo (supervisor, maestro) que determina la respuesta que debería generar la red a partir de una entrada determinada.

El supervisor controla la salida de la red y en caso de que ésta no coincida con la deseada, se procederá a modificar los pesos de las conexiones, con el fin de conseguir que la salida obtenida se aproxime a la deseada.

3.7 Criterios de Evaluación

Para medir la eficiencia del proceso de clasificación de fenotipos metabólicos se utiliza un conjunto de conceptos, los cuales definirán el tipo de método utilizado y cómo es clasificado por el sistema implementado.

3.7.1 Verdadero positivo.

(True Positive o Verdaderos Positivos): Conocido como “acierto”, es cuando realmente hay un paciente de un tipo “x” y es detectada correctamente.

3.7.2 Falso negativo.

(False Negative o Falsos Negativos): Conocido como “fallo”, es cuando existe un paciente de un tipo “x”, pero el clasificador no tuvo la capacidad para detectarla.

3.7.3 Sensibilidad.

Conocida como Tasa de Verdaderos Positivos, mide la proporción de datos que han sido clasificados correctamente como tal. Es decir, el número de tipos que son identificados por el clasificador correctamente.

3.7.4 Especificidad

Conocida como Tasa de Falsos Negativos mide cuán específico es el clasificador. Es decir, mide la proporción de datos que son identificados de modo incorrecto. Esto es, el número de datos que son clasificados como tal incorrectamente.

3.8 Clasificación de los Fenotipos Metabólicos

Nuestra investigación hace énfasis en estas alteraciones metabólicas conocidas como fenotipos o enfermedades no transmisibles. Para poder realizar la correcta clasificación se debe tener en cuenta lo que significa cada uno de estos tipos. Estos se clasifican de la siguiente manera:

Primer Tipo: su valor valido numéricamente corresponde al uno, Delgado Sano. Este hace referencia a todos aquellos pacientes que según sus análisis realizados fue diagnosticado en un estado delgado y que está sano.

Segundo Tipo: su valor valido numéricamente corresponde al dos, hace referencia a todos aquellos pacientes que según sus análisis realizados fue diagnosticado en un estado delgado, pero a diferencia del anterior este se establece como enfermo.

Tercer Tipo: su valor valido numéricamente corresponde al tres, hace referencia a todos aquellos pacientes que según su análisis fue diagnosticado en un estado de sobrepeso y en forma sana.

Cuarto Tipo: su valor valido numéricamente corresponde al cuatro, hace referencia a todos aquellos pacientes que según los análisis realizados fue diagnosticado en un estado de sobrepeso, y a diferencia del anterior este se encuentra en un estado enfermo.

Quinto Tipo: su valor valido numéricamente corresponde al cinco, este hace referencia a todos aquellos pacientes que fueron diagnosticado según los análisis realizados como obeso en un estado sano.

Sexto Tipo: su valor valido numéricamente corresponde al seis, este hace referencia a todos aquellos pacientes que fueron diagnosticados como obesos según los análisis realizados, pero a diferencia del anterior este se encuentra en un estado enfermo.

Para la ejecución de nuestro sistema de clasificación, con el modelo de máquinas de soporte vectorial, se creó primeramente un prototipo utilizando solo dos clases de afectaciones de salud. Para prueba de nuestro sistema se utilizó el fenotipo clase uno y la clase seis. Esto se hace con el fin de trabajar nuestro modelo primeramente probándolo con solo dos clases, y para la verificación de estos resultados.

4. Revisión bibliográfica

Durante la realización del primer objetivo específico se hizo una investigación en diferentes motores de búsqueda especializados, con el fin de encontrar información que nos ayude a contemplar de una forma más clara los conceptos y temas que estamos tratando, el impacto que tienen en la sociedad, y que se ha hecho hasta hoy en este campo temático.

4.1 La Salud

El concepto más importante en esta investigación es la salud, es ahí donde se quiere intervenir para dar un aporte investigativo a la sociedad en torno a la prevención de enfermedades. Según la OMS, Organización Mundial de la Salud “La salud es un estado de completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades”.

En la presente investigación se hace énfasis en el tema relacionado con fenotipos metabólicos que están afectando la salud de las personas, y encontrar las características que establecen que un individuo tiene un riesgo alto, medio o bajo, en padecer esta enfermedad.

4.2 Los Fenotipos Metabólicos

Según Holmes E., Leng Loo R., (2008). “Los fenotipos metabólicos son los productos de las interacciones entre una variedad de factores dietéticos, otros estilos de vida / medio ambiente, intestino microbiano y genético. Utilizamos un enfoque analítico de exploración a gran escala para investigar la variación fenotipo metabólico a través y dentro de cuatro poblaciones humanas, con base en 1espectroscopia H RMN. Los metabolitos que discriminan entre las poblaciones se relacionan con los datos de los individuos sobre la presión arterial, un

importante factor de riesgo de enfermedad coronaria y accidente cerebrovascular (las principales causas de mortalidad en todo el mundo)”.

Se tiene en cuenta que para el trabajo y el desarrollo de esta investigación se contemplaron los seis tipos de fenotipos metabólicos, los cuales se clasifican de la siguiente manera: delgado sano, delgado enfermo, sobrepeso sano, sobrepeso enfermo, obeso sano, obeso enfermo. Estos seis tipos de afectaciones son las que se buscan tener en cuenta para la clasificación con ayuda de un sistema de aprendizaje supervisado en máquinas de soporte vectorial.

4.3 Enfermedades no Transmisibles

Estas enfermedades hacen parte de un grupo grande, clasificadas como enfermedades no transmisibles que considerarse crónicas, y que son aquellas que afectan de manera lenta y progresan lentamente. La OMS, Organización Mundial de la Salud define a las enfermedades no transmisibles como: “Enfermedades de larga duración, lenta progresión, que no se resuelven espontáneamente y que rara vez logran una curación total. A nivel mundial, son responsables del 63% de las muertes equivalente a 36 millones de muertes por año, un 25% de estas en menores de 60 años por lo que la detección precoz y el tratamiento oportuno de estas patologías es prioritario”.

En un aspecto general, refiriéndonos a estas enfermedades no transmisibles, cuyo impacto ha generado un gran porcentaje de mortalidad en el mundo y afecta continuamente a las personas, demostrado en estadísticas notables hechas por la OMS. La organización mundial de la salud afirma que:

- a) Las enfermedades no transmisibles (ENT) matan a 40 millones de personas cada año, lo que equivale al 70% de las muertes que se producen en el mundo.

- b) Cada año mueren por ENT 15 millones de personas de entre 30 y 69 años de edad; más del 80% de estas muertes "prematuras" ocurren en países de ingresos bajos y medianos.
- c) Las enfermedades cardiovasculares constituyen la mayoría de las muertes por ENT (17,7 millones cada año), seguidas del cáncer (8,8 millones), las enfermedades respiratorias (3,9 millones) y la diabetes (1,6 millones).
- d) Estos cuatro grupos de enfermedades son responsables de más del 80% de todas las muertes prematuras por ENT.
- e) El consumo de tabaco, la inactividad física, el uso nocivo del alcohol y las dietas malsanas aumentan el riesgo de morir a causa de una de las ENT.
- f) La detección, el cribado y el tratamiento, igual que los cuidados paliativos, son componentes fundamentales de la respuesta a las ENT.

A lo largo de los años se han venido realizando investigaciones correspondientes a estos grupos de enfermedades que están dentro del concepto de no transmisible. Uno de los factores que puede afectar el fenotipo metabólico, es la obesidad ya que esta hace parte de ese grupo de factores que perjudican la salud de las personas.

4.4 Obesos enfermos

Una de las investigaciones encontradas sobre este grupo de factores que pueden afectar la salud de los individuos es la del señor, S. Karger GmbH, Freiburg. (2017) “La actual epidemia de obesidad plantea un importante problema de salud pública ya que la obesidad predispone a varias enfermedades crónicas. El IMC y la adiposidad total se correlacionan positivamente con el riesgo de enfermedad cardio metabólica a nivel de población. Sin embargo, la distribución

de grasa corporal y una función deteriorada del tejido adiposo, en lugar de la masa total de grasa, mejor predecir la resistencia a la insulina y las complicaciones relacionadas a nivel individual. La disfunción del tejido adiposo se determina por una alteración de la capacidad de expansión del tejido adiposo, hipertrofia de los adipocitos, alteración del metabolismo de los lípidos e inflamación local. Recientes estudios en humanos sugieren que la oxigenación del tejido adiposo puede ser un factor clave en la presente. Un subgrupo de personas obesas - el 'obeso metabólicamente sano' (MHO) - tienen una mejor función del tejido adiposo, menos almacenamiento de grasa ectópica, Y son más sensibles a la insulina que las personas obesas metabólicamente insalubres, haciendo hincapié en el papel central de la función del tejido adiposo en la salud metabólica”.

4.5 Obesos sanos

De cierta manera el termino obeso se asocia con un estado grave de salud, pero actualmente muchas investigaciones están hablando de lo que se puede determinar como “obesos sanos” es un caso particular que se presenta en muchas personas que presentan obesidad, y es importante dejar claro que no solo están los obesos enfermos, sino que también se puede estar en un estado obeso sano.

Según Salabert, E. (2017) afirma que:

Los obesos metabólicamente sanos no sufren hipercolesterolemia o diabetes tipo 2. Los expertos explican que, además del IMC, el tipo de grasa, su distribución corporal, y el estilo de vida, afectan al perfil metabólico. El concepto de obesos metabólicamente sanos se refiere a las personas que a pesar de presentar obesidad (índice de masa corporal (IMC) igual o superior a 30 kg/m²) no sufren complicaciones metabólicas, como diabetes tipo 2, hipertensión, hipercolesterolemia... Según los expertos, no solo el IMC, sino también el tipo de grasa y su distribución corporal, influyen en la salud del individuo y en el riesgo a largo plazo de desarrollar determinadas enfermedades asociadas a la obesidad.

4.6 Delgado enfermo

También en estudios e investigaciones realizadas se encuentra que estar delgado no siempre es sinónimo de una buena salud, y es lo que la sociedad asume. Se asume que un individuo estando en un estado delgado quiere decir que es saludable, pero varias investigaciones presumen que no siempre estar delgado es estar saludable.

“En la última década se ha caracterizado la existencia de individuos ‘falsos delgados’. Se trata de personas que tienen un peso normal de acorde a su estatura, pero que presentan características relacionadas con la obesidad, como deposición de grasa visceral, y un mayor riesgo de padecer diabetes mellitus tipo 2 y enfermedad cardiovascular. Este síndrome, conocido en la literatura científica como obesidad normopeso, va asociado a la ingesta de dietas desequilibradas, ricas en grasas o con exceso de azúcares sencillos y a un descenso en la actividad física, y constituye un serio problema de salud pública” (García P. 2016).

Actualmente existen varios estudio e investigaciones que tratan de aclarar y darle una solución a este problema que está afectando la sociedad, y trae con ella un impacto de mortalidad alto. La mayoría de los estudios encontrados tratan de el efecto que tiene el fenotipo con respecto al metabolismo, y cuáles son los avances tecnológicos que están ayudando en la investigación.

Según SH Zeisel , H. C. Freake , DE Bauman (2005) “El concepto de fenotipo nutricional se propone como un conjunto definido e integrado de factores genéticos, proteómicos, metabolómicos, funcionales y de comportamiento que, cuando se miden, forman la base para la evaluación del estado nutricional humano. El fenotipo nutricional integra los efectos de la dieta en la enfermedad / bienestar y es la indicación cuantitativa de los caminos por los cuales los genes y el medio ambiente ejercen sus efectos sobre la salud. Los avances tecnológicos y los conocimientos biológicos fundamentales permiten definir y medir con exactitud el fenotipo

nutricional en una sección transversal de individuos con diversos estados de salud y enfermedad. Esta creciente base de datos y conocimientos podría servir como un recurso para todas las disciplinas científicas implicadas en la salud humana. Las ciencias de la nutrición deben ser un motor principal en la toma de decisiones clave que incluyen: Qué insumos ambientales (además de la dieta) son necesarios; Qué genes / proteínas / metabolitos deben ser medidos; Qué fenotipos de punto final deben ser incluidos; Y qué herramientas informáticas están disponibles para hacer preguntas nutricionalmente relevantes. La nutrición debe ser la disciplina principal que establece cómo los elementos del fenotipo nutricional varían en función de la dieta”.

Estos tipos mencionados anteriormente son los más tomados en cuenta en el campo de la salud, ya que son los que más impacto causan en la sociedad. Por esta razón es que la presente investigación requiere de conceptos e ideas que ayuden a conocer su concepto y su impacto, para así poder estimar la viabilidad del sistema de clasificación de estas enfermedades con ayuda de un modelo de máquinas de soporte vectorial, el cual será un aporte investigativo en el área de la salud y la ciencia.

5. Metodología de la investigación

Dobles, Zúñiga. (1998). Afirma. *“la teoría de la ciencia que sostiene el positivismo se caracteriza por afirmar que el único conocimiento verdadero es aquel que es producido por la ciencia, particularmente con el empleo de su método”*. La metodología para las variables de investigación que se realizará es positivista, ya que se busca explicar, controlar y predecir el impacto generado por la variedad de afectaciones a la salud ocasionado por los fenotipos metabólicos.

5.1 Tipo de investigación

El diseño del proyecto es de tipo no experimental ya que las variables no se pueden manipular y estas se dan en su contexto natural para después ser analizadas, al estar proporcionando una visión sobre la necesidad de usar herramientas computacionales que permitan utilizar máquinas de soporte vectorial para la clasificación de los diferentes tipos de fenotipos metabólicos.

5.2 Instrumentos y técnicas

Para la elaboración del proyecto se realizará una ficha bibliográfica donde se registra y se resumen los datos extraídos de fuentes bibliográficas. Allí se establecen los conocimientos previos que se tomarán en cuenta para la elaboración de este proyecto de investigación, ya que servirán de guía para una visión más clara de la situación actual del problema y cuáles son los avances investigativos que se han formulado hasta la fecha. En la ficha bibliográfica se abarcará un aspecto específico que es tomado en la presente investigación, se refiera a las investigaciones y conceptos que definen y explican los fenotipos metabólicos, para conocer y determinar el impacto y la viabilidad de implementar un software que implemente máquinas

de soporte vectorial, con aprendizaje supervisado para la clasificación de los fenotipos metabólicos.

5.2.1 Ficha bibliográfica.

TITULO	AUTOR	AÑO	RESUMEN
SISTEMA DE VIGILANCIA DE ENFERMEDADES NO TRANSMISIBLES EN CUBA	Dr. Ricardo Batista Moliner, Dr. Orlando Landrove Rodríguez	2000	Las enfermedades no transmisibles constituyen las primeras causas de muerte en nuestro país, más de las 3/4 partes de los fallecidos corresponden a las llamadas "enfermedades crónicas no transmisibles". La vigilancia ha devenido una función básica de la Salud Pública y en la actualidad es una de las vías para alcanzar mejores resultados en la atención a este problema. El desarrollo del sistema de salud en Cuba y las experiencias internacionales en este campo permiten avanzar aún más en el terreno de la vigilancia y el control de estas entidades. La complejidad de la epidemiología de estas enfermedades crea dificultades para el desarrollo de la vigilancia, lo que requiere de la integración de todas las áreas desde el nivel primario de atención. Para la concepción y organización del sistema de vigilancia de las enfermedades no transmisibles en nuestro país se revisaron las experiencias internacionales y los resultados del desarrollo de la vigilancia de enfermedades infecciosas y de la salud pública en general. Asimismo, se realizaron consultas con expertos en el tema que aportaron elementos de valor al trabajo. Como resultados de esa labor se presentó una propuesta que se discutió por los responsables de ejecutar las actividades del mismo, en todo el país.
EL FENOTIPO METABÓLICO EN LA OBESIDAD: MASA GRASA, DISTRIBUCIÓN DE GRASA CORPORAL Y FUNCIÓN DEL TEJIDO ADIPOSO	S. Karger GmbH, Freiburg.	2017	La actual epidemia de obesidad plantea un importante problema de salud pública ya que la obesidad predispone a varias enfermedades crónicas. El IMC y la adiposidad total se correlacionan positivamente con el riesgo de enfermedad cardio metabólica a nivel de población. Sin embargo, la distribución de grasa corporal y una función deteriorada del tejido adiposo, en lugar de la masa total de grasa, mejor predicen la resistencia a la insulina y las complicaciones relacionadas a nivel individual. La disfunción del tejido adiposo se determina por una alteración de la capacidad de expansión del tejido adiposo, hipertrofia de los adipocitos, alteración del metabolismo de los lípidos e inflamación local. Recientes estudios en humanos sugieren que la oxigenación del tejido adiposo puede ser un factor clave en la presente.
SISTEMA DE MONITOREO Y EVALUACIÓN DE LA CALIDAD DE LA ATENCIÓN MÉDICA EN LAS ENFERMEDADES NO TRANSMISIBLES	Dra. Milvia Ramírez Rodríguez, Lic. Diana Rodríguez Gabaldá, Dr. Orlando Landrove	2001	Con esta metodología pretendemos que se evalúe la calidad de la atención médica que se brinda a los pacientes con enfermedades no transmisibles (asma bronquial, hipertensión arterial y diabetes mellitus) en las instituciones de salud a nivel primario y secundario, pues estas son las primeras causas de muerte y discapacidad en nuestro país. Esto permitirá determinar en qué medida las acciones de salud que se realizan cumplen el objetivo fundamental de mantener, restaurar o promover salud.
OBESOS METABÓLICAMENTE SANOS*	Salabert, E.	2017	Las enfermedades no transmisibles actualmente son causa del 60 % de la mortalidad y la discapacidad en el mundo. La mayoría de las personas afectadas (cerca del 70 %) tienen menos de 45 años y el 30 % menos de 15 años de edad. ¹ Todos los países desarrollados muestran esta tendencia incluyendo el nuestro, donde 3 de cada 4 muertes que ocurren son atribuidas a las enfermedades no transmisibles, se destaca el hecho de que éstas constituyen 9 de las 10 primeras causas de muerte.

Figura 4. Ficha Bibliográfica

6. Desarrollo del proyecto

6.1 Apropiación de la información

En la actual investigación que se está realizando sobre factores de riesgo de salud, se tiene en cuenta la información recolectada por La Encuesta Nacional de Examen de Salud y Nutrición (NHANES). Con los datos recolectados en esta base de datos, se realizará el análisis correspondiente de la información para así poder identificar y apropiarnos de aquellos datos importantes que nos servirán para el clasificador de los tipos de fenotipos metabólicos.

La Encuesta Nacional de Examen de Salud y Nutrición (NHANES) es un programa de estudios diseñado para evaluar el estado de salud y nutrición de los adultos y de los niños en Estados Unidos. La encuesta es única en cuanto a que combina entrevistas y exámenes físicos. El programa de NHANES empezó al comienzo de los años 60 y ha sido conducido como una serie de encuestas enfocadas en diferentes grupos de la población o tópicos de salud. Las detalladas entrevistas de NHANES incluyen preguntas relacionadas con demografía, socioeconomía, dieta y salud. El componente de examen consiste en exámenes médicos y dentales, mediciones fisiológicas, y pruebas de laboratorio administradas por personal médico altamente entrenado. Los resultados de esta encuesta serán usados para determinar la prevalencia de enfermedades graves y los factores de riesgo de las enfermedades.

Para la apropiación de la información y selección de los datos relevantes que afectan la salud de los individuos se hace uso del análisis de un especialista en fenotipos metabólicos quien nos indicara cuales variables establecidas en la base de datos NHANES son las que identifican los factores de riesgo de padecer uno de los fenotipos metabólicos, y asimismo obtener una apropiación adecuada del funcionamiento de los datos.

6.2 Población y muestra

Para la clasificación de los tipos, se hace uso de la base de datos NHANES quienes recolectan cierta cantidad de datos de pacientes que se escogen al azar en los estados unidos.

Para el caso de esta investigación, no se hace uso que técnicas de recolección de datos ya que se tuvo en cuenta para la elaboración del presente trabajo una base de datos certificada y aprobada por especialistas en estados unidos, alterar el valor de estas variables ocasiona una falsa calificación de los resultados establecidos por los participantes de la encuesta nacional. Esta recolección de datos abarca desde pruebas físicas, psicológicas y algunas preguntas de localización social. Esta matriz de datos normalmente cuenta con 67 variables, las cuales algunas serán utilizadas para la clasificación, y para elaboración del sistema clasificar que se propone en la presente investigación. de estas 67 variables se hace la limpieza de aquellos datos que pueden dañar el sistema, por ejemplo, los datos NULL, que son estos datos que no se encontraron, o no fueron diligenciadas. El número de pacientes encuestados en esta prueba fue de 1824. Así se compone nuestra matriz de datos que se está tomando en cuenta para la elaboración de este sistema de clasificación.

Para la exitosa clasificación del sistema, se utilizó la ayuda de un experto en salud, y experto en estas alteraciones de salud como lo son los fenotipos metabólicos.

The image displays a screenshot of an Excel spreadsheet titled "A1175". The spreadsheet contains a data table with columns labeled C through AD and rows numbered 2 through 11. The data is organized into columns representing different variables, with some columns having multiple sub-columns. The bottom of the image shows the Windows taskbar with the search bar and system tray.

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD		
2	Sexo	EnteOrCivil	Placa	EnteSocSec	EnteEder	MaTab	ConAlcohol	Puro	Talla	BMC	BMCcliv	FORMA	CAED	Insulas	HOMA2B	HOMA2S	HOMA2R	GBoral	CTotal	TAG	MDL	YLDL	LDL	PAS	PAD	PAM	ColReHCL	IndecTAGH	HDLbeta	CAE
3	1	0	1	2	3	2	2	88,30	1,56	26,20	5	107,00	8,00	10,30	31,00	3,20	250,00	253,00	144,60	50,00	28,32	100,08	110,00	80,00	90,00	203,00	2,89	0		
4	2	2	1	4	5	1	2	124,20	1,86	35,90	5	121,00	23,50	303,80	25,70	3,30	73,00	221,00	339,00	42,00	78,20	100,80	120,00	80,00	93,33	173,00	3,31	0		
5	2	1	1	3	7	2	2	103,50	1,75	35,76	5	112,50	10,40	139,00	66,00	1,50	86,22	262,00	245,00	45,00	43,00	164,00	120,00	80,00	93,33	217,00	5,44	0		
6	2	2	2	4	3	1	1	73,20	1,60	30,54	4	113,00	12,40	127,60	53,70	1,90	96,00	243,00	220,00	45,00	44,50	160,00	120,00	70,00	86,67	204,00	4,53	0		
7	2	3	5	4	3	2	2	18,40	1,55	32,63	4	102,00	22,40	18,40	27,90	2,60	127,00	165,00	153,00	41,00	26,90	117,40	100,00	80,00	90,00	124,00	3,24	0		
8	2	2	1	3	8	2	2	104,50	1,71	35,67	5	111,00	10,80	111,70	61,20	1,60	97,00	204,00	170,00	43,00	34,00	127,00	120,00	70,00	86,67	161,00	3,95	0		
9	1	2	1	3	4	2	1	63,40	1,51	30,44	4	91,00	12,20	89,00	51,30	1,90	116,30	225,00	135,80	55,00	27,36	142,84	104,00	66,00	117,00	2,47	0			
10	1	5	2	3	8	1	2	30,40	1,67	32,41	4	83,00	11,00	129,50	51,20	1,60	91,00	273,00	173,00	58,00	34,60	196,40	110,00	70,00	93,33	221,00	2,98	0		
11	2	1	1	4	5	2	1	83,00	1,69	31,96	4	94,00	23,80	244,20	20,60	4,30	98,00	164,00	160,00	41,00	32,00	111,00	120,00	70,00	86,67	143,00	3,90	0		
12	2	2	1	3	5	2	2	86,20	1,61	33,25	4	97,00	10,60	111,50	62,20	1,60	97,40	203,00	148,00	40,00	29,60	133,40	140,00	70,00	90,00	163,00	3,70	0		
13	1	5	1	4	4	2	1	16,60	1,53	32,32	4	102,00	30,70	217,30	22,60	4,40	96,00	125,00	147,00	26,00	29,40	69,60	100,00	70,00	83,33	193,00	5,65	1		
14	1	2	1	2	3	2	2	91,50	1,66	33,21	4	107,00	11,20	123,90	53,70	1,70	94,00	187,00	145,00	55,00	23,00	103,00	115,00	70,00	83,33	132,00	2,64	0		
15	2	1	1	3	4	2	2	124,40	1,79	38,83	5	127,00	26,70	228,00	26,00	3,80	93,00	276,00	145,00	48,00	38,00	233,00	140,00	60,00	96,67	178,00	3,82	1		
16	1	2	1	2	7	2	2	120,90	1,95	33,90	4	126,00	22,70	178,60	29,70	3,40	100,00	226,00	133,00	48,00	26,60	117,40	110,00	60,00	83,33	178,00	2,71	0		
17	2	2	1	4	3	2	1	173,00	1,60	67,58	6	104,00	10,80	111,70	61,00	1,60	97,40	203,00	148,00	40,00	29,60	133,40	140,00	70,00	90,00	163,00	3,70	0		
18	2	2	1	3	5	3	2	103,50	1,73	36,32	5	126,00	25,90	232,40	27,00	3,70	91,00	189,00	144,00	43,00	28,80	111,20	130,00	80,00	103,33	140,00	2,94	0		
19	1	1	1	3	7	2	2	106,40	1,87	30,43	4	101,50	3,30	109,20	71,50	1,40	94,02	171,00	143,00	25,00	27,80	130,20	110,00	70,00	83,33	136,00	3,97	1		
20	2	2	1	2	7	3	2	105,50	1,79	33,30	4	109,00	10,50	129,80	64,40	1,60	90,00	183,00	145,00	48,00	23,00	106,00	110,00	70,00	83,33	135,00	3,82	0		
21	1	2	1	3	2	3	1	108,00	1,85	39,67	5	118,00	25,50	208,30	26,90	3,70	96,14	215,00	135,20	54,00	27,04	133,96	140,00	90,00	106,67	160,00	2,50	0		
22	2	1	1	3	2	4	1	92,50	1,72	31,27	4	110,00	26,20	221,20	26,40	3,80	94,00	164,00	148,00	40,00	27,62	86,38	120,00	80,00	93,33	194,00	3,45	0		
23	1	1	1	4	2	1	2	79,70	1,54	33,38	4	100,00	4,30	111,40	163,60	0,60	88,00	198,00	145,00	31,00	23,00	138,00	120,00	80,00	93,33	187,00	4,66	1		
24	2	2	1	3	8	2	1	91,70	1,74	30,29	4	108,00	11,90	116,40	53,80	1,70	96,00	229,00	133,00	43,00	26,32	139,40	120,00	80,00	93,33	226,00	3,85	0		
25	1	2	1	4	3	2	2	83,00	1,48	31,50	4	98,00	8,50	96,60	71,20	1,90	97,00	217,00	123,00	40,00	25,80	131,20	120,00	80,00	93,33	217,00	3,23	1		
26	1	2	2	4	5	2	2	104,50	1,66	48,81	6	146,00	12,00	122,20	15,20	1,80	97,00	203,00	129,80	48,00	25,80	129,20	122,00	70,00	106,67	185,00	2,49	0		
27	1	5	1	2	7	1	1	83,50	1,64	31,05	4	93,00	11,00	111,20	52,50	1,90	97,00	263,00	140,00	61,00	27,92	273,88	100,00	70,00	90,00	302,00	2,00	0		
28	2	1	1	3	5	2	2	87,20	1,67	31,30	4	93,00	15,60	94,70	43,70	2,30	91,00	185,00	137,00	31,00	27,40	126,60	120,00	70,00	86,67	174,00	4,42	1		
29	1	2	1	4	5	3	2	64,30	1,45	30,43	4	84,00	11,60	130,80	48,80	2,00	98,00	285,00	127,00	47,00	25,38	132,32	122,00	80,00	94,00	258,00	2,70	1		
30	2	1	1	3	7	3	2	91,50	1,71	31,22	4	107,00	13,60	87,80	50,20	2,00	89,00	173,00	138,00	50,00	27,60	95,40	115,00	70,00	88,33	123,00	2,76	0		
31	1	1	1	4	1	2	2	110,80	1,70	38,34	5	118,00	24,90	215,60	27,70	2,60	94,00	237,00	130,20	38,00	26,00	172,00	120,00	80,00	93,33	188,00	3,33	1		
32	1	1	2	3	8	3	1	113,90	1,71	41,00	6	125,00	13,40	163,50	34,50	2,90	93,00	179,00	122,00	43,00	24,40	105,60	120,00	80,00	93,33	180,00	2,43	1		
33	2	2	1	3	5	1	2	103,70	1,79	32,73	4	103,80	14,80	136,80	47,80	2,10	81,00	285,00	148,00	45,00	23,60	130,40	105,00	70,00	116,7	220,00	3,29	0		
34	2	2	5	3	8	2	1	100,20	1,70	41,63	6	146,00	10,50	199,70	3,20	10,80	107,00	185,00	122,00	43,00	22,40	119,60	120,00	70,00	90,00	142,00	2,60	0		
35	1	4	1	4	3	3	2	113,00	1,82	43,06	6	118,00	19,30	243,70	31,20	2,70	89,00	200,00	149,70	31,00	23,34	133,06	110,00	70,00	93,33	193,00	4,85	1		
36	2	2	1	2	8	1	1	105,30	1,81	32,08	4	108,00	15,40	160,20	44,30	2,30	92,00	153,00	130,00	30,00	26,00	97,00	125,00	80,00	95,00	123,00	4,33	1		
37	2	2	1	2	8	2	1	99,40	1,71	33,39	4	102,00	16,00	111,90	42,00	2,40	96,00	145,00	124,00	36,00	24,80	84,20	120,00	80,00	93,33	189,00	3,44	1		
38	1	2	2	3	8	2	2	91,20	1,66	33,30	4	105,00	7,60	123,90	31,20	1,10	82,00	205,00	145,00	44,00	23,00	132,00	110,00	70,00	83,33	161,00	3,90	1		
39	2	5	1	4	5	1	2	91,50	1,68	32,25	4	115,00	14,50	238,50	28,80	3,50	88,00	247,00	134,00	40,00	26,80	180,20	140,00	90,00	103,33	207,00	3,85	0		
40	2	3	3	8	4	3	2	107,00	1,85	39,90	5	126,00	25,50	170,30	26,30	3,80	107,00	176,00	168,00	45,00	23,60	109,40	103,00	74,00	106,67	190,00	2,40	0		
41	1	2	5	4	4	2	1	75,40	1,48	34,42	4	101,00	10,50	94,30	36,40	2,70	97,00	172,00	119,00	27,00	23,80	121,20	128,00	74,00	92,00	145,00	4,41	1		
42	2	2	1	4	4	2	2	85,20	1,65	31,29	4	101,00	11,90	124,30	55,90	1,80	96,00	160,00	120,00	38,00	24,00	97,00	110,00	60,00	116,67	121,00	3,88	0		
43	2	2	1	3	5	3	2	101,20	1,62	38,56	5	115,00	15,50	167,80	36,50	2,70	96,00	166,00	119,00	35,00	23,80	87,20	120,00	80,00	93,33	111,00	2,96	0		
44	2	5	1	4	8	3	1	107,00	1,78	33,77	4	118,00	20,10	242,70	36,30	2,80	78,00	215,00	148,00	43,00	23,20	142,80	130,00	90,00	103,33	172,00	3,40	0		
45	2	2	1	3	5	3	1	73,40	1,61	30,63	4	105,00	22,80	133,60	23,90	3,30	96,00	182,00	168,00	35,00	23,60	113,40	111,00	70,00	83,67	142,00	3,83	0		
46	1	1	1	4	3	2	2	82,30	1,44	38,53	5	110,00	20,20	16,40	28,90	3,50	85,00	159,00	131,00	62,00	14,60	82,40	110,00	70,00	83,33	37,00	1,18	0		
47	2	2	1	4	5	1	1	100,90	1,83	30,33	4	110,00	13,80	118,60	43,60	2,00	89,00	168,00	127,00	35,00	25,40	107,60	110,00	70,00	83,33	133,00	3,83	1		
48	1	4	1	5	3	1	2	83,30	1,59	32,95	4	105,00	64,90	246,20	10,60	3,40	128,00													

6.3 Libro de códigos

Para la identificación de los datos relevantes, como se propone en el objetivo específico dos, se utilizó la consulta a un experto en temas de salud. Aquí se resaltan las variables seleccionadas, y se da el conocimiento de la exclusión de la variable insulina la cual como se nombró en el planteamiento del problema, es la variable que tiene un valor monetario difícil de acceder para muchos pacientes.

Los datos utilizados para trabajar en el clasificador después de la limpieza y selección constan de 56 variables, con 912 pacientes encuestados. De los tipos de afectaciones que se están trabajando, para hacer la prueba y el manejo del clasificador creador en la herramienta de Matlab, se seleccionó el tipo uno (delgado sano) y el tipo seis (obeso enfermo) los cuales se trabajaran en la máquina de soporte vectorial haciendo como primera acción el entrenamiento, para así después realizar la clasificación.

The image shows a screenshot of an Excel spreadsheet with the following characteristics:

- Title Bar:** G907
- Columns:** Labeled A through AA, with a final column for formulas.
- Rows:** Labeled 1 through 52.
- Content:** A grid of numerical data. The first few columns (A-E) contain small integers (1 or 2). Columns F through AA contain numerical values, many of which are identical across rows, suggesting a classification or scoring system.
- Taskbar:** Shows the Windows Start button, a search bar with the text "Escribe aquí para buscar", and several application icons including File Explorer, Edge, Chrome, and Word. The system clock shows 17:51 on 09/12/2017.

Figura 6. Base de Datos Seleccionada para el Clasificador

Cuadro 1. Sexo: Sexo del encuestado.

		Valor
Atributos estándar	Posición	2
	Etiqueta	Sexo del encuestado
	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	1	Femenino
	2	Masculino

Cuadro 2. Estado Civil: Estado civil del encuestado

		Valor
Atributos estándar	Posición	3
	Etiqueta	Estado Civil
	Tipo	Numérico
	Formato	F8
	Medición	Nominal
	Rol	Entrada
Valores válidos	1	Soltero
	2	Casado

3	Viudo
4	Divorciado
5	Concubinato

Cuadro 3. Raza: Raza del encuestado

		Valor
Atributos estándar	Posición	4
	Etiqueta	Raza
	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	1	Mezclado
	2	Blanco Hispánico
	3	Afro-Venezolano
	4	Oriental
	5	Wayuu
	6	Yuckpa
	7	Bari
	8	Arabigo

Cuadro 4. EstSocioec: Estatus socioeconómico por Graffar

		Valor
Atributos estándar	Posición	5
	Etiqueta	Estatus Socioeconómico
	Tipo	Numérico
	Formato	F1
	Medición	Ordinal
	Rol	Entrada
Valores válidos	1	Estrato I: Clase alta
	2	Estrato II: Clase Media-Alta
	3	Estrato III: Clase Media
	4	Estrato IV: Clase Obrera
	5	Estrato V: Pobreza Extrema

Cuadro 5. EstEduc: grado educativo mayor

		Valor
Atributos estándar	Posición	6
	Etiqueta	Estatus Educativo
	Tipo	Numérico
	Formato	F1
	Medición	Ordinal

	Rol	Entrada
Valores válidos	1	Analfabeta
	2	Primaria Incompleta
	3	Primaria Completa
	4	Secundaria Incompleta
	5	Secundaria Completa
	6	Universitaria Incompleta
	7	Universitaria Completa
	8	Educación Técnica

Cuadro 6. HabTab: Hábito tabáquico

		Valor
Atributos estándar	Posición	7
	Etiqueta	Hábito Tabáquico
	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	1	Si
	2	No
	3	Fumó en el pasado

Cuadro 7. ConAlcohol: Consumo de alcohol

		Valor
Atributos estándar	Posición	8
	Etiqueta	Consumo de alcohol
	Tipo	Numérico
	Formato	F1
	Medición	Ordinal
	Rol	Entrada
Valores válidos	1	Si
	2	No

Cuadro 8. Peso: Peso en Kg

		Valor
Atributos estándar	Posición	10
	Etiqueta	Peso
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 9. Talla: Talla en metros

		Valor
Atributos estándar	Posición	11
	Etiqueta	Talla
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 10. IMC: Índice de masa corporal

		Valor
Atributos estándar	Posición	12
	Etiqueta	IMC
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 11. IMCclasifOMS: Clasificación ponderal según el IMC según la organización Mundial de la Salud (ordinal)

		Valor
Atributos estándar	Posición	13
	Etiqueta	Clasificación del IMC según la Organización Mundial de la Salud
	Tipo	Numérico
	Formato	F8
	Medición	Ordinal
	Rol	Entrada
Valores válidos	1	Bajo Peso
	2	Normopeso
	3	Sobrepeso
	4	Obesidad 1
	5	Obesidad 2
	6	Obesidad 3

Cuadro 12. CABD: Circunferencia abdominal en centímetros

		Valor
Atributos estándar	Posición	14
	Etiqueta	Circunferencia abdominal
	Tipo	Numérico

	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 13. GBasal: Glicemia en ayuno

		Valor
Atributos estándar	Posición	20
	Etiqueta	Glicemia basal
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 14. CTotal: Colesterol total

		Valor
Atributos estándar	Posición	21
	Etiqueta	Colesterol total
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 15. TAG: Triacilglicéridos

		Valor
Atributos estándar	Posición	22
	Etiqueta	Triacilgliceridos
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 16. HDL: Colesterol de alta densidad o colesterol “bueno”

		Valor
Atributos estándar	Posición	23
	Etiqueta	Concentración de HDL
	Tipo	Numérico
	Formato	F9.2
	Medición	Escala
	Rol	Entrada

Cuadro 17. VLDL: Colesterol de muy baja densidad (TAG/5)

		Valor
Atributos estándar	Posición	24
	Etiqueta	VLDL
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 18. LDL: Colesterol de baja densidad o colesterol malo

		Valor
Atributos estándar	Posición	25
	Etiqueta	LDL
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 19. PAS: Presión arterial sistólica

		Valor
Atributos estándar	Posición	26
	Etiqueta	Presion Arterial Sistolica
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 20. PAD: Presión arterial diastólica

		Valor
Atributos estándar	Posición	27
	Etiqueta	Presion Arterial Diastolica
	Tipo	Numérico
	Formato	F5.2
	Medición	Escala
	Rol	Entrada

Cuadro 21. PAM: Presión arterial media

		Valor
Atributos estándar	Posición	28
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 22. ColNoHDL: Colesterol NO HDL

		Valor
Atributos estándar	Posición	29
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

El colesterol no HDL es un nuevo factor de riesgo para enfermedad cardiovascular (Infarto de miocardio, por ejemplo). Es la suma del COLESTEROL de LDL y el Colesterol de VLDL ó más fácil, el colesterol total menos el colesterol de HDL.

Cuadro 23. ÍndiceTAGHDL: División entre los TAG/HDL

		Valor
Atributos estándar	Posición	30
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 24. HDLbajasIDF: HDL bajas

		Valor
Atributos estándar	Posición	31
	Etiqueta	HDLbajasIDF
	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	0	No
	1	Si

HDL bajas, es una dicotomización de las HDL, tomando un punto de corte de 40 mg/dL para hombres y 50 mg/Dl para mujeres. Menos de estos valores se considera bajo, lo cual es

un factor de riesgo para enfermedad cardiovascular e indicio de una persona metabólicamente enferma.

Cuadro 25. CABDeltaIDF: Circunferencia abdominal Alta

		Valor
Atributos estándar	Posición	32
	Etiqueta	CABDeltaIDF
	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	0	No
	1	Si

Dicotomización de la variable circunferencia abdominal, tomando como punto de corte 80 cm en las mujeres y 90cm en los hombres. Un valor mayor a esto, se considera una circunferencia elevada, lo cual confiere riesgo para enfermedades cardiovasculares y metabólicas como la diabetes.

Cuadro 26. TAGaltaIDF: Triacilglicéridos altos

		Valor
Atributos estándar	Posición	33
	Etiqueta	TAGaltaIDF
	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	0	No
	1	Si

Dicotomización de la variable TAG (trialciglicéridos). Triacilglicéridos iguales o mayores a 150 mg/dL se considera indicio de enfermedad metabólica. Personas que consumen cantidades elevadas de carbohidratos (arroz, papa, yuca, pan, azúcar, gaseosas), tienden a tener elevados los triacilgliéridos.

Cuadro 27. GLICEMIAaltaIDF: Glicemia elevada

		Valor
Atributos estándar	Posición	34
	Etiqueta	GLICEMIAaltaI DF

	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	0	No
	1	Si

Dicotomizarían del valor de glucosa en sangre en ayuno. Normal: Menos de 100 mg/dL. Alta: Mayor o igual a 100 mg/dL. Personas con glicemia elevada en ayuno tienen mayor riesgo de enfermedad cardiovascular. Una glicemia en ayuno mayor o igual a 126 es indicativo de diabetes. Una glicemia en ayuno entre 100 y 125 se denomina pre-diabetes.

Cuadro 28. PAaltaIDF Presión arterial elevada

		Valor
Atributos estándar	Posición	35
	Etiqueta	PAaltaIDF
	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	0	No
	1	Si

Dicotomización de las variables PAS y PAD que nos permite saber si la persona tiene la Presión arterial alta (Hipertensión)

Cuadro 29. GLICEMIAaltaATPIII2003: Glicemia alta según criterios del ATPIII (USA)

		Valor
Atributos estándar	Posición	36
	Etiqueta	GLICEMIAaltaATPIII2003
	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	0	No
	1	Si

Dicotomización de la glicemia según ATPIII. Esta variable es idéntica a la dicotomización anterior por la IDF, pues tiene el mismo punto de corte de 100 mg/dL.

Cuadro 30. CABDeltaATPIII2003: Circunferencia abdominal alta

		Valor
Atributos estándar	Posición	37
	Etiqueta	CABDeltaATPIII2003
	Tipo	Numérico

	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	0	No
	1	Si

Dicotomización de la circunferencia abdominal usando los puntos de corte del ATPIII USA. 102 cm en hombres y 88 cm en las mujeres

Cuadro 31. CABDeltaATPIII2005: Circunferencia abdominal alta ATPIII 2005

		Valor
Atributos estándar	Posición	38
	Etiqueta	CABDeltaATPIII2005
	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	0	No
	1	Si

Cuadro 32. SMIDF2005: Diagnóstico de SM según la IDF 2005

		Valor
Atributos estándar	Posición	42
	Etiqueta	SMIDF2005
	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	0	No
	1	Si

Cuadro 33. SMATPIII2003: Diagnóstico de SM según ATPIII 2003

		Valor
Atributos estándar	Posición	43
	Etiqueta	SMATPIII2003
	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	0	No
	1	Si

Cuadro 34. SMATPIII2005: Diagnóstico de SM según ATPIII 2005

		Valor
Atributos estándar	Posición	44
	Etiqueta	SMATPIII2005
	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	0	No
	1	Si

Cuadro 35. RazaLast: Categorías de raza

		Valor
Atributos estándar	Posición	45
	Etiqueta	Raza Categorías
	Tipo	Numérico
	Formato	F8.2
	Medición	Nominal
	Rol	Entrada
Valores válidos	1,00	Mezclado
	2,00	Blanco Hispánico
	3,00	Afro-venezolano

4,00	Indígena
	Americano
5,00	Otros

Cuadro 36. METsTotalesAct

		Valor
Atributos estándar	Posición	46
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 37. PatronActFisica

		Valor
Atributos estándar	Posición	47
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8
	Medición	Ordinal

	Rol	Entrada
Valores válidos	1	Leve
	2	Moderada
	3	Alta

Cuadro 38. CAaltaNUEVOSTUNTOS: Circunferencia abdominal alta puntos de corte maracaibo

		Valor
Atributos estándar	Posición	48
	Etiqueta	CAaltaNUEVOSTUNTOS
	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	0	No
	1	Si

Circunferencia abdominal dicotomizada según los puntos de corte encontrados para la ciudad de Maracaibo

Cuadro 39. IR

		Valor
Atributos estándar	Posición	49
	Etiqueta	IR
	Tipo	Numérico
	Formato	F1
	Medición	Nominal
	Rol	Entrada
Valores válidos	0	Sin IR
	1	IR

Presencia o no de IR. Punto de corte 2 para HOMA.

Cuadro 40. Combinaciones TODAS Combinaciones de los criterios de síndrome metabólico

		Valor
Atributos estándar	Posición	52
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Nominal
	Rol	Entrada
Valores válidos	,00	Sanos

1,00	C
2,00	P
3,00	G
4,00	H
5,00	T
6,00	CP
7,00	CG
8,00	CH
9,00	CT
10,00	PG
11,00	PH
12,00	PT
13,00	GH
14,00	GT
15,00	HT
16,00	CPG
17,00	CPH
18,00	CPT
19,00	CGH
20,00	CGT
21,00	CHT
22,00	PGH
23,00	PGT
24,00	GHT
25,00	PHT
26,00	CPGH

27,00	CPGT
28,00	CPHT
29,00	PGHT
30,00	CGHT
31,00	Todos

Cuadro 41. PAJNC7nueva: Clasificación de personas según su presión arterial

		Valor
Atributos estándar	Posición	53
	Etiqueta	PAJNC7nueva
	Tipo	Numérico
	Formato	F8.2
	Medición	Ordinal
	Rol	Entrada
Valores válidos	1,00	Normotensos
	2,00	Prehipertensos
	3,00	Hipertensos

Cuadro 42. EstatusGlicemico: Estado de la glicemia en ayuno

		Valor
Atributos estándar	Posición	54
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Nominal
	Rol	Entrada
Valores válidos	1,00	No
	2,00	APDM
	3,00	Glicemia 126 o mas
	4,00	GAA

Cuadro 43. SMIDF2009: Diagnóstico de SM según criterios del 2009 (los más nuevos)

		Valor
Atributos estándar	Posición	56
	Etiqueta	SMIDF2009
	Tipo	Numérico
	Formato	F1
	Medición	Nominal

	Rol	Entrada
Valores válidos	0	No
	1	Si

Cuadro 44. TallaCM: Talla en cm

		Valor
Atributos estándar	Posición	57
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 45. ÍndiceCinturaAltura: índice cintura/altura

		Valor
Atributos estándar	Posición	58
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 46. ÍndiceCinturaAlturaLOG: Transformación logarítmica del anterior

		Valor
Atributos estándar	Posición	59
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 47. Alteraciones Lipídicas: Frecuencia de alteraciones en los lípidos sanguíneos

		Valor
Atributos estándar	Posición	61
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8
	Medición	Nominal
	Rol	Entrada
Valores válidos	0	Eulipemicos
	1	LDLaislada
	2	HDLaislada
	3	TAGaislados
	4	TAGyHDL
	5	TAGyLDL
	6	LDLyHDL
	7	LDL,HDLyTAG

Cuadro 48. AFtrabajoMETs

		Valor
Atributos estándar	Posición	63
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Ordinal
	Rol	Entrada
Valores válidos	,00	Inactivos
	1,00	Muy baja
	2,00	Baja
	3,00	Moderada
	4,00	Alta
	5,00	Muy alta

Cuadro 49. AFtransporteMETs: Actividad física en METs esfera transporte

		Valor
Atributos estándar	Posición	64
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2

	Medición	Ordinal
	Rol	Entrada
Valores válidos	,00	Inactivos
	1,00	Muy baja
	2,00	Baja
	3,00	Moderada
	4,00	Alta
	5,00	Muy alta

Cuadro 50. AFhogarMETs: Actividad física en METs esfera hogar

		Valor
Atributos estándar	Posición	65
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Ordinal
	Rol	Entrada
Valores válidos	,00	Inactivos
	1,00	Muy baja
	2,00	Baja
	3,00	Moderada
	4,00	Alta
	5,00	Muy alta

Cuadro 51. AFocioMETs: Actividad física en METs esfera ocio

		Valor
Atributos estándar	Posición	66
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Ordinal
	Rol	Entrada
Valores válidos	,00	Inactivos
	1,00	Muy baja
	2,00	Baja
	3,00	Moderada
	4,00	Alta
	5,00	Muy alta

Cuadro 52. VAI: Índice visceral adiposo (mide el riesgo de diabetes)

		Valor
Atributos estándar	Posición	67
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 53. LogVAI

		Valor
Atributos estándar	Posición	68
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 54. MRFA: Agregación de múltiples factores de riesgo (Menos de 2 y 2 o más)

		Valor
Atributos estándar	Posición	73
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Nominal
	Rol	Entrada
Valores válidos	,00	<2
	1,00	2 o mas

Cuadro 55. TGI

		Valor
Atributos estándar	Posición	76
	Etiqueta	<ninguno>
	Tipo	Numérico
	Formato	F8.2
	Medición	Escala
	Rol	Entrada

Cuadro 56. TGIDico

		Valor
Atributos estándar	Posición	77
	Etiqueta	TGIDico
	Tipo	Numérico
	Formato	F8.2
	Medición	Nominal
	Rol	Entrada
Valores válidos	,00	<4,5
	1,00	4,5 o mas

Cuadro 57. Fenotipoclusters: Fenotipos metabólicos por análisis de clúster

		Valor
Atributos estándar	Posición	71
	Etiqueta	Fenotipos
	Tipo	Numérico
	Formato	F8
	Medición	Nominal
	Rol	Entrada
Valores válidos	1	Delgado Sano
	2	Delgado Enfermo

3	Sobrepeso Sano
4	Sobrepeso Enfermo
5	Obeso Sano
6	Obeso Enfermo

6.4 Verificación del funcionamiento de la GUI implementada computacionalmente

En esta fase se denota lo que se plantea en el objetivo específico tres, el cual describe una interfaz gráfica de usuario GUI, para la clasificación.

Mediante esta fase se verifica el funcionamiento de la interfaz gráfica desarrollada, cargando las bases de datos seleccionada, tomada de la encuesta nacional de salud y nutrición NANHES para efectuar el proceso de verificación de funcionamiento adecuado. De este modo, se carga en la GUI una matriz de datos con dos tipos a clasificar.

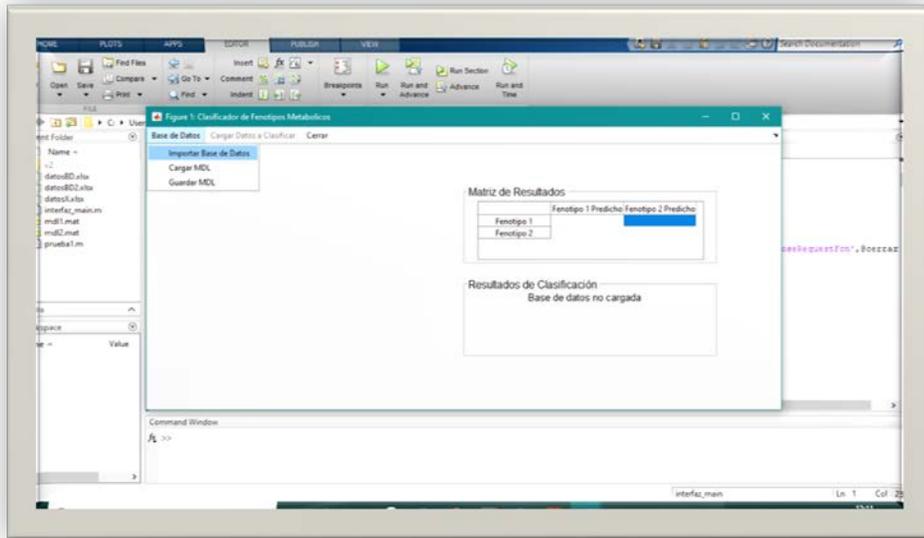


Figura 7. GUI Cargar Base de Datos.

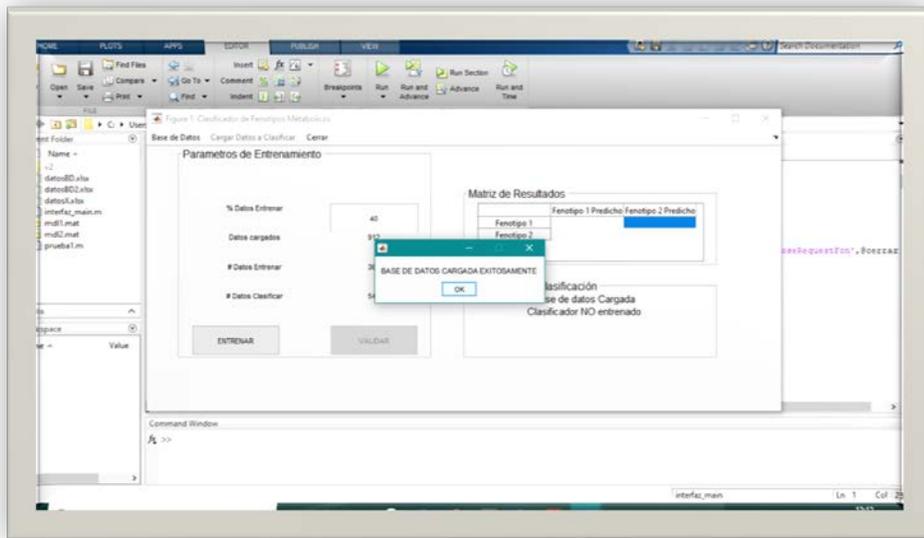


Figura 8. GUI Base de Datos Cargada Exitosamente.

Siguiendo el modelo de máquinas de soporte vectorial se hace el entrenamiento del modelo, sugiriendo un porcentaje de datos con los que se deba entrenar, con el fin de que la máquina de soporte vectorial entienda los datos que se deben clasificar siguiendo el concepto de aprendizaje supervisado.

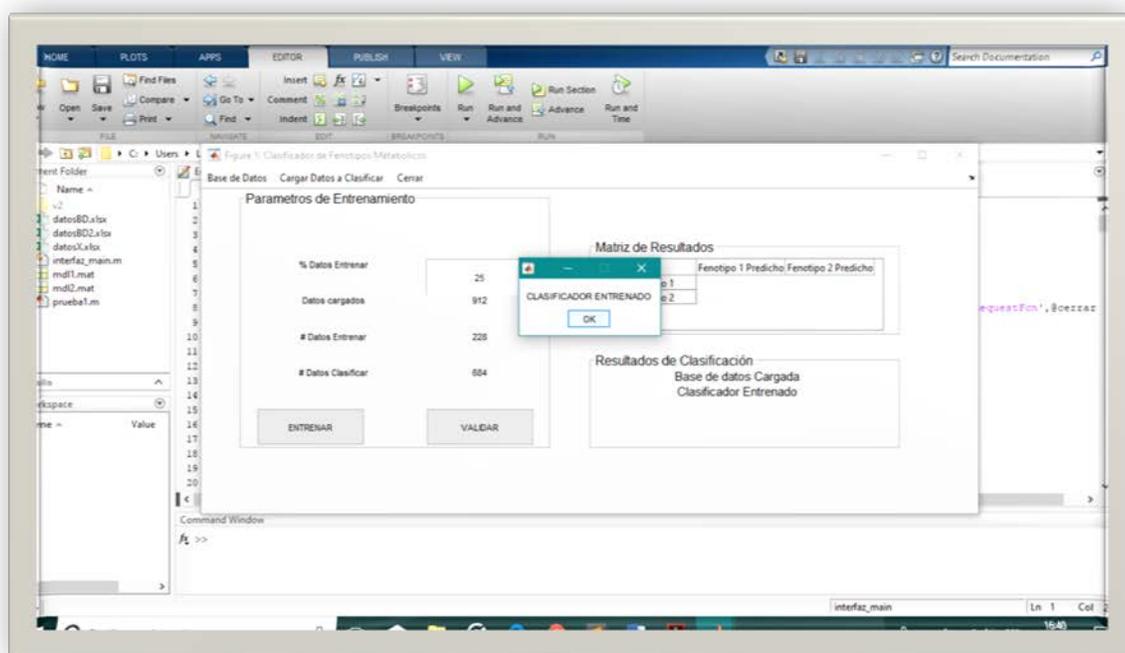


Figura 9. GUI Clasificador Entrenado.

Después de entrenar la máquina se hace la clasificación adecuada de los datos, para esto nuestra interfaz gráfica propone una opción de botón para realizar esta acción.

Realizada esta acción, en la interfaz gráfica de usuario se muestra la tabla de resultados que indica el número de datos clasificados exitosamente, y el número de datos que no se clasificaron de manera correcta, también nos muestra el porcentaje de sensibilidad y especificidad de nuestro modelo.

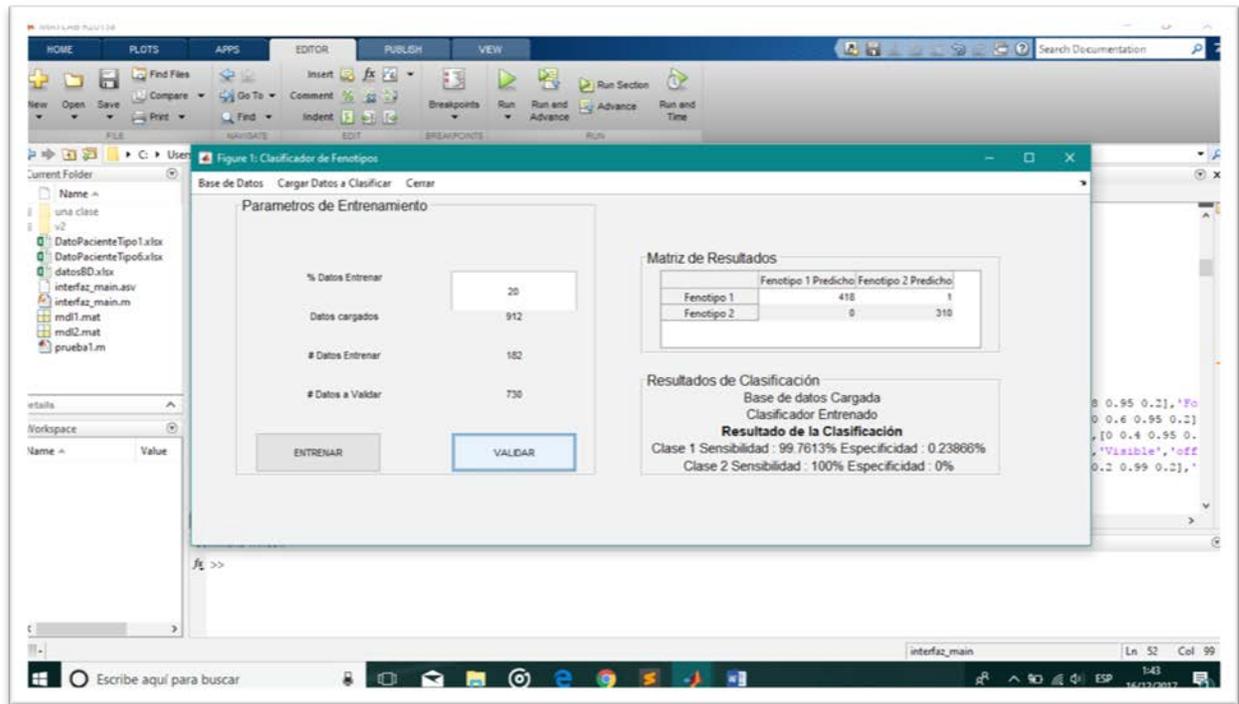


Figura 10. GUI Validación del Clasificador.

Una vez teniendo el modelo de máquinas de soporte vectorial entrenado y validado, se puede realizar la clasificación de un paciente cargando los datos del mismo contenido en un archivo de Excel.

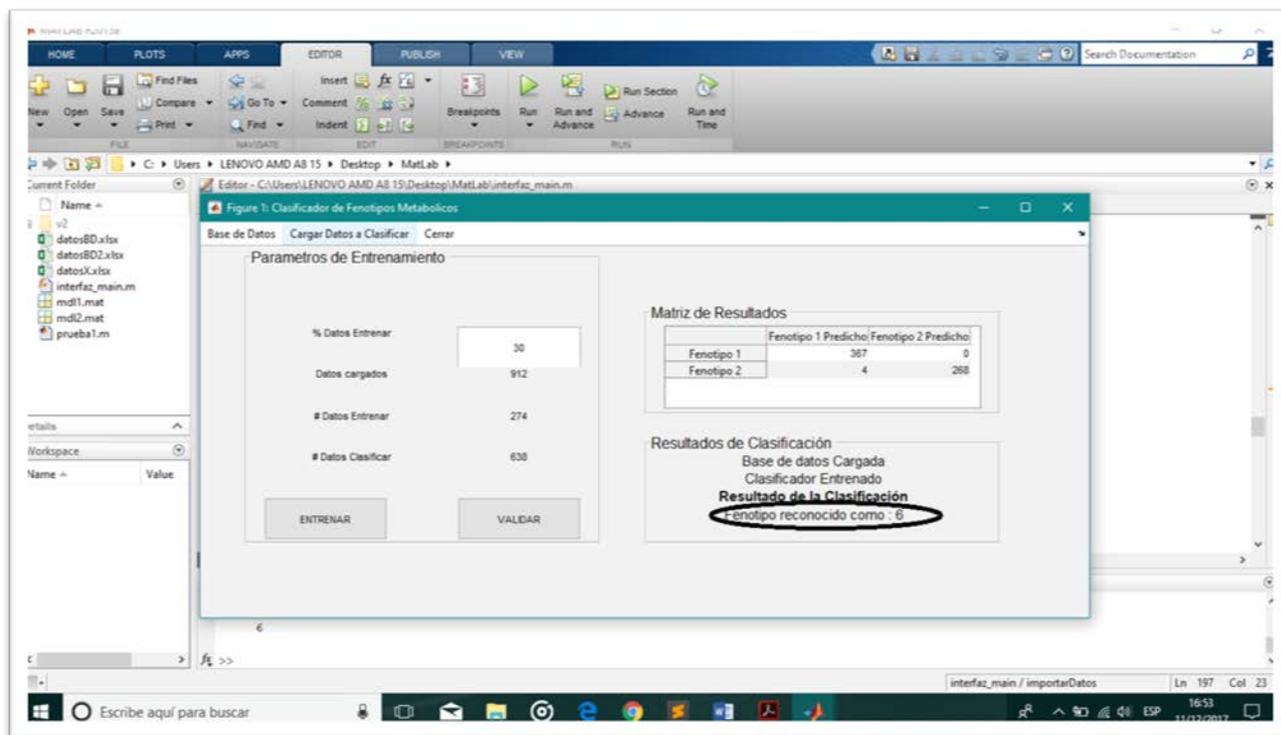


Figura 12. Clasificación del Tipo de Fenotipo Metabólico con un Paciente.

6.5. Fundamentación

Las principales razones del uso de un ciclo de desarrollo iterativo e incremental de tipo SCRUM para la ejecución de este proyecto son:

- Sistema modular. Las características del “Desarrollo de una interfaz gráfica de usuario para la clasificación del tipo de fenotipo metabólico que presenta un paciente, usando la base de datos NHANES” permiten desarrollar una base funcional mínima y sobre ella ir incrementando las funcionalidades o modificando el comportamiento o apariencia de las ya implementadas.
- Entregar un prototipo funcional del sistema que permita describir claramente lo que nos muestra el clasificador.

6.6. Personas y Roles del Proyecto

Cuadro 58. Personas y roles del trabajo.

Persona	Contacto	Rol
José Gerardo Chacón Rangel	j.chacon@unisimonbolivar.edu.co	Scrum Master
Luis Enrique Ramirez Gutiérrez	l_ramirez@unisimon.edu.co	Team Scrum
Médico Especialista		Stakeholders

Cuadro 59. Historias de usuario y criterios de aceptacion.

Desarrollo ágil: Historias de usuario y criterios de aceptación

Identificador (ID) de la historia	Enunciado de la historia				Criterios de aceptación			Criterios de aceptación	
	Rol	Característica / Funcionalidad	Razón / Resultado	Número (#) de escenario	Criterio de aceptación (Título)	Contexto	Resultado / Comportamiento esperado	Estimacion	Dependencia
FM-001	Medico Especialista en el area	Cargar Base de Datos de pacientes con fenotipos metabolicos.	Con la finalidad de poder entrenar la maquina de soporte vectorial	1	Carga el Archivo .XLSX al sistema	Importa y valida cada registro de datos por paciente encontrado en el archivo .XLS	Muestra un mensaje de que la base de datos ha sido importada correctamente y se visualizan los parametros de : Porcentaje de datos a entrenar , Datos Cargados, Número Datos a entrenar, Número Datos a validar	13 Ptos.	
FM-002	Medico Especialista en el area	Entrenar maquina de soporte vectorial.	Establecer un aprendizaje supervisado	1	Parametrice el aprendizaje de la maquina de soporte vectorial	Entrenar Maquina de soporte vectorial	Muestra un mensaje de clasificador entrenado y habilita la opcion de validar los datos.	21 Ptos.	FM-001
FM-003	Medico Especialista en el area	Validar los datos restantes de la base de datos.	Comprobacion de que el clasificador tuvo un correcto aprendizaje supervisado	1	Matriz de resultados	Validar los resultado de registro de datos restantes para la clasificacion de Fenotipos metabolicos de tipo 1 o 6	Muestra la matriz de resultados de acuerdo a los fenotipos metabolicos de tipo 1 y 6. Muestra los resultados de sensibilidad y especificidad de la clasificacion de datos de pacientes con fenotipos metabolicos de tipo 1 y 6	21 Ptos.	FM-001, FM-002
FM-004	Medico Especialista en el area	Clasificar paciente.	Con la finalidad de clasificar el fenotipo metabolico de un paciente.	1	Clasificacion de datos importados en un archivo .XLSX del paciente a diagnosticar	Importar y valida el registro de datos del paciente encontrado en el archivo .XLS	Muestra un mensaje que dice el paciente pertenece a un tipo de fenotipo metabolico	8 Ptos.	FM-001, FM-002, FM-003

Cuadro 60. Sprint 1 Historias de Usuario FM-001

Backlog ID	Tarea	Tipo	Estado	Responsable
FM-001	Cargar base de Datos de pacientes con fenotipos metabólicos.	Implementación	Completo	Luis Enrique Ramírez

Cuadro 61. Sprint 2 Historias de Usuario FM-002

Backlog ID	Tarea	Tipo	Estado	Responsable
FM-002	Entrenar máquina de soporte vectorial.	Diseño, análisis e implementación	Completo	Luis Enrique Ramírez

Cuadro 62. Sprint 3 Historias de Usuario FM-003

Backlog ID	Tarea	Tipo	Estado	Responsable
FM-003	Validar los datos restantes de la base de datos.	Diseño, análisis e implementación	Completo	Luis Enrique Ramírez

Cuadro 63. Sprint 4 Historias de Usuario FM-004

Backlog ID	Tarea	Tipo	Estado	Responsable
FM-004	Clasificar paciente.	Diseño, análisis e implementación	Completo	Luis Enrique Ramírez

Cuadro 64. Tabla de producto

GRAFICA DEL PRODUCTO				
SPRINT	TOTAL DE PUNTOS DEL PROYECTO	COMPLETADO	ESFUERZO EN PUNTOS ACUMULADOS	TRAYECTORIA ESTIMADA
0	63	0	0	0
1	63	13	13	#N/A
2	63	21	34	#N/A
3	63	21	55	#N/A
4	63	8	63	63

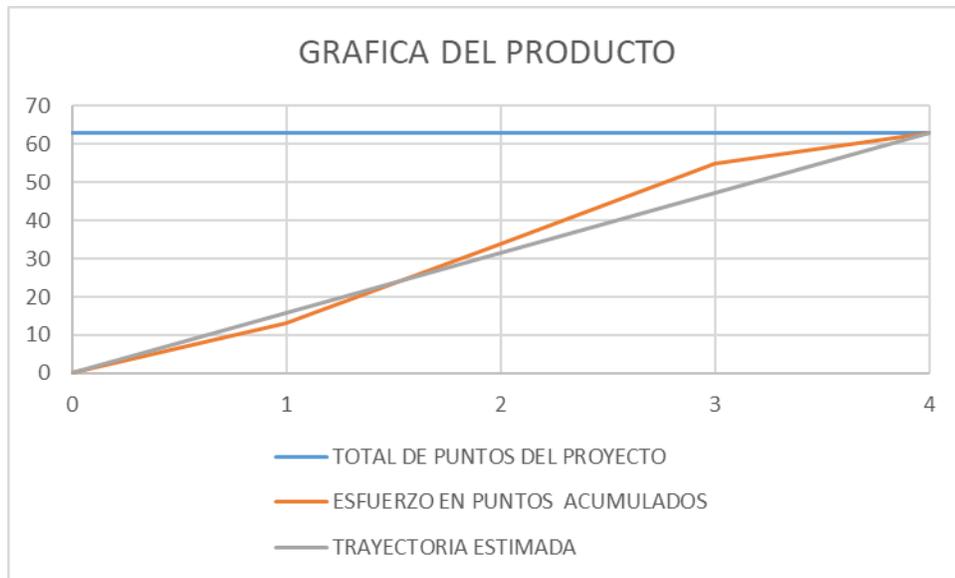


Figura 13. Grafica del producto

7. Conclusiones

El producto final, sistema de clasificación, consiste en un prototipo de una herramienta versátil para la clasificación de los diferentes tipos de fenotipos metabólicos, con el fin de poder realizar esto sin la necesidad de tener el examen médico de la insulina, la cual tiene un costo elevado y es de difícil acceso para algunos pacientes.

Para la elaboración de este sistema, se contó con la base de datos NHANES, quien aportó las variables necesarias para ser utilizadas después de un análisis y una limpieza de las mismas, en el clasificador con un modelo de máquinas de soporte vectorial, mediante el aprendizaje supervisado.

Para la elaboración de la interfaz gráfica se utilizó la herramienta Matlab que propone un entorno de desarrollo integrado con un lenguaje de programación propio, permitiendo así la manipulación por parte del usuario del clasificador y de los datos utilizados para este modelo.

Bibliografía

Alianza Tecnológica Ecuatoriana. (2010). Manual de interfaz gráfica usando Matlab

Disponible en: http://www.matpic.com/esp/matlab/interfaz_grafica.html

Bárbara Echiburúa, Amanda ladrón de Guevara & Cecilia Pereira. (2014). Clasificación de los

fenotipos de síndrome de ovario poliquístico de acuerdo a los criterios de rotterdam:

¿una condición estática o variable?

Corinna Cortes & Vladimir Vapnik. (1995). *Support-Vector Networks*. AT&T Bell Labs.,

Holmdel, NJ 07733, USA.

David Llanos & Alfonso Azuaje. (2013). Redes de Aprendizaje Supervisado y No

Supervisado.

Dr. Ricardo Batista Moliner & Dr. Orlando Landrove Rodríguez. (2000). Sistema de vigilancia

de enfermedades no transmisibles en cuba.

Dra. Milvia Ramírez Rodríguez, Lic. Diana Rodríguez Gabaldá & Dr. Orlando Landrove.

(2001). Sistema de monitoreo y evaluación de la calidad de la atención médica en las

enfermedades no transmisibles.

Eduardo Mauricio león García. (2010). Clasificación de datos de ids basada en ls-svm con pso.

Gustavo a. Betancourt. (2005). Las máquinas de soporte vectorial (svms). scientia et technica año xi, no 27, abril 2005. utp. issn 0122-1701.

Juan Carlos Llibre GuerraI; Milagros A. Guerra HernándezII & Elianne Perera Miniet. (2008). Comportamiento de las enfermedades crónicas no transmisibles en adultos mayores.

Luis Gerardo Meza Cascante. (2013). El paradigma positivista y la concepción dialéctica del conocimiento. Escuela de Matemática, Instituto Tecnológico de Costa Rica.

S. Karger GmbH, Freiburg. (2017). El fenotipo metabólico en la obesidad: masa grasa, distribución de grasa corporal y función del tejido adiposo.

Salabert, E. (2017) Obesos metabólicamente sanos. Web consultas Revista de salud y bienestar.

SH Zeisel, C. Freake & DE Bauman. (2014). El fenotipo nutricional en la era metabólica.