



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**FACULTAD DE MEDICINA**  
**DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN**

**POSGRADO EN MEDICINA INTERNA**

CREACIÓN DE UN MODELO DE CLASIFICACIÓN BAYESIANO PARA IDENTIFICAR LA  
CONVERSIÓN DE PREDIABETES A DIABETES TIPO 2 EN PACIENTES DE LA  
CONSULTA EXTERNA DE MEDICINA INTERNA DEL HOSPITAL CENTRAL SUR DE ALTA  
ESPECIALIDAD PEMEX.

**TESIS**

QUE PARA OBTENER EL GRADO ACADÉMICO DE:

**MEDICO INTERNISTA**

PRESENTA:

**ESTEBAN SEPÚLVEDA OLIVA**

**TUTOR(A) O TUTORES PRINCIPALES:**

DR. MARTÍN CORONADO MALAGÓN

DR. JOSÉ ROBERTO MENDOZA MARTÍNEZ.

**COMITÉ TUTOR:**

DR. MARTÍN CORONADO MALAGÓN

DR. JOSÉ ROBERTO MENDOZA MARTÍNEZ.



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## **Agradecimientos:**

A mis papás; Olga Sofía Oliva Ramírez y Luis Abelardo Sepúlveda de la O: sin ustedes este camino ni siquiera se hubiera empezado, gracias por su paciencia y apoyo a lo largo ya de muchos años, en la vida nunca alcanzaré a compensar todo lo que han hecho por mí, los amo.

A mi hermano Germán Sepúlveda Oliva: mi loco, me inspiras, y aunque casi no nos vemos, quisiera tener ese valor y carisma que tu tienes, te extraño y me haces mucha falta; en donde estes y los que estés haciendo, vive libre hermanito.

A Elena, Angelina y Carlos, mis otros hermanos, los adoro a los tres, así como sin mis papás esto no hubiera empezado, sin ustedes esto no se habría completado, han sido mi gran apoyo estos años y aunque nacimos en casas separadas, los considero mi familia, han sido años difíciles pero gracias a ustedes las risas y el cariño nunca faltaron.

A mis maestros y guías, los doctores Salvador Mendoza, Cesar Sifuentes, Jose Luis Amaya, Andrés Calderón, Adriana Alarcón, Gabriel Klimek, Ivette Cruz, Andrés Lupián, Raul López, Miguel Angel Falcón. Gracias por todas sus enseñanzas, no solo académicas, también personales, si bien con muchas fallas, hoy soy un mejor médico y persona gracias a su dedicación, confianza, paciencia y sabiduría.

Al Dr. Martín Coronado Malagón, gracias por su liderazgo a lo largo de estos años, por su paciencia y manera tan única y calmada de enseñarnos y sobre todo por su fortaleza, empatía y entereza al plantar la cara por nosotros cuando todo fue oscuro y complicado, ha sido un honor tenerlo como tutor, maestro y jefe de servicio.

A Robert, si algo se de estadística es gracias a ti, y si terminamos este proyecto también es gracias a ti, creo que hablo por los cuatro cuando digo que no hay palabras para agradecerte lo que has hecho por nosotros.

A Magda, mi maestra y amiga, gracias por ser esa mano firme que tanto necesitaba para encaminarme, gracias por mostrarme que el verdadero camino es el estudio de las artes oscuras de la hematología, gracias por estar cuando la vida era difícil y dolorosa, gracias por enseñarme tanto y lograr que me vuelva a enamorar del arte y ciencia de la medicina, antes de conocerte estaba por renunciar, ahora tengo claro lo que quiero hacer de mi vida, se que dices que no debería, pero si quiero ser como alguien, es alguien eres tu. Gracias.

A Naomi, a quien sin buscar, llegaste a mi vida. Haces que cada día tenga un color hermoso, un brillo intenso y un aroma delicioso. Gracias por ser mi compañera de locuras, estudios, viajes, comidas, sueños y alegrías. De todo lo que se vislumbra a futuro, la aventura de la vida a tu lado es la que más me ilusiona. Aunque me digas -tú más- tu eres la más brillante

## **Dedicatoria:**

A mi pulguita, por quien hago todo esto. Todo lo que soy y todo lo que tengo, es tuyo, siempre.

## **RESUMEN.**

### **Introducción.**

La prediabetes, un estado de alteración del metabolismo de la glucosa es una enfermedad con elevada prevalencia en México y el mundo, la cual confiere un riesgo incrementado no solo de conversión a diabetes, sino de desenlaces cardiovasculares. La creación de un índice que tome en cuenta medidas más allá de los niveles de la glucosa, contemplando al paciente que vive con prediabetes de una manera holística, tomando la glucosa como una manifestación más de una enfermedad multifactorial, con múltiples y variables manifestaciones clínicas y bioquímicas, podría ayudarnos a mantener una mejor vigilancia y establecer estrategias de seguimiento que permitan al paciente revertir el proceso inflamatorio que de no ser frenado, llevará al paciente a complicaciones varias que afecten no solo su expectativa de vida, también su calidad de vida, así como la de su entorno familiar-social.

### **Metodología.**

Se realizó un estudio de tipo observacional/ descriptivo de características transversales por temporalidad, retrospectivo por la obtención de los datos, utilizando variables clínicas y bioquímicas del expediente clínico electrónico del Hospital Central Sur de Alta Especialidad de PEMEX; además de un estudio de pruebas diagnósticas.

Para la parte descriptiva del estudio se utilizó el software Stata V.17.0 , realizando prueba t de Student para comparación de medias, prueba U de Mann Whitney, en función de la distribución que presentan las variables, o prueba  $\chi^2$  de independencia para evaluar asociación entre variables categóricas.

Para la creación del modelo se utilizó la técnica de aprendizaje supervisado: Naive Bayes Classifier (Clasificador Bayesiano Ingenuo) utilizando la paquetería “naive bayes” del software R.

### **Resultados.**

Se obtuvo una muestra de un total de 116 pacientes con una edad promedio de 59 años y una desviación estándar de 14 años. En función de la presencia de la conversión de Prediabetes a Diabetes Mellitus se crearon dos grupos, el 40.52% (47) presentaron la Conversión de Prediabetes a Diabetes Mellitus y, complementariamente el 59.48% (69) no presentaron la Conversión. Como principales hallazgos, encontramos una diferencia estadísticamente significativa en la función de distribución para las variables IMC y Perímetro abdominal (PA), encontrando una mediana mayor en el grupo de Conversión comparado con el grupo Sin Conversión, para ambas variables.

Se aprecia una asociación estadísticamente significativa entre el Tabaquismo y la Conversión de prediabetes a DM, encontrando una mayor prevalencia 55.07% (38) en el grupo Sin Conversión comparado con el grupo Con Conversión, siendo de 34.78%

La única variable bioquímica que presentó una diferencia estadísticamente significativa en la función de distribución de los Triglicéridos entre los grupos de Conversión. Además se encontró una asociación estadísticamente significativa fueron el consumo de estatinas y de diuréticos. Con dichas variables se diseñó un clasificador Naive Bayesiano el cual obtuvo una sensibilidad de 0.58 (0.40- 0.74, especificidad 0.96 (0.87- 0.99), VPP 0.91 (0.7- 0.98), VPN 0.8 (0.68- 0.88) con una precisión de 0.8265 [0.7369 – 0.8956]

### **Discusión.**

Se logró el desarrollo exitoso de un modelo predictivo utilizando un modelo clasificador naive Bayesiano, obteniendo una alta especificidad, valor predictivo positivo y precisión para predecir conversión de prediabetes en diabetes tipo 2

### **Conclusión.**

Nuestro modelo predictivo es una herramienta útil para el médico y el paciente para establecer un riesgo determinado para la progresión de la enfermedad y así poder tener un seguimiento más estrecho para alcanzar las metas establecidas y evitar la progresión a diabetes con las complicaciones que conlleva.

*No nos toca a nosotros elegir qué tiempo vivir,  
Solo podemos elegir qué hacer con el tiempo que se nos ha dado.*

**J.R.R Tolkien. El Señor de los Anillos, La comunidad del anillo.**

## INDICE

	1
<b>INDICE</b>	<b>3</b>
<b>TÍTULO.</b>	<b>4</b>
<b>DEFINICIÓN DEL PROBLEMA.</b>	<b>6</b>
<b>MARCO TEORICO Y ANTECEDENTES</b>	<b>8</b>
1.- PREDIABETES: DEFINICIÓN Y ESTADÍSTICA EN MÉXICO Y EL MUNDO.	8
2.- FISIOPATOLOGÍA DE LA DIABETES TIPO 2: ¿UNA ENFERMEDAD METABÓLICA, INFLAMATORIA O CARDIOVASCULAR?	10
3.- CONVERSIÓN DE PREDIABETES A DIABETES TIPO 2.	12
4.- EL SÍNDROME METABÓLICO.	12
5.- MODELOS DE CONVERSIÓN.	14
6.- APRENDIZAJE AUTOMATIZADO	15
7.- MODELO NAIVE BAYESIANO.	16
8.- LA IMPORTANCIA DE UN NUEVO MODELO.	17
<b>JUSTIFICACIÓN.</b>	<b>18</b>
<b>PLANTEAMIENTO DEL PROBLEMA</b>	<b>19</b>
<b>PREGUNTA DE INVESTIGACIÓN.</b>	<b>20</b>
<b>HIPÓTESIS.</b>	<b>20</b>
HIPÓTESIS NULA.	20
HIPÓTESIS ALTERNA.	20
<b>OBJETIVOS</b>	<b>21</b>
OBJETIVO GENERAL.	21
OBJETIVOS ESPECÍFICOS.	21
<b>TIPO DE ESTUDIO.</b>	<b>21</b>
<b>METODOLOGÍA.</b>	<b>22</b>
UNIVERSO.	22
UNIDADES DE OBSERVACIÓN.	22
TIPO DE MUESTREO.	22
TAMAÑO DE MUESTRA.	22
CRITERIOS DE INCLUSIÓN:	23
CRITERIOS DE EXCLUSIÓN:	23
<b>VARIABLES DE ESTUDIO</b>	<b>25</b>
VARIABLE(S) INDEPENDIENTE(S)	25
VARIABLE(S) DEPENDIENTE(S)	29

<b>RECOLECCIÓN DE DATOS.</b>	<b>29</b>
<b>RECURSOS Y LOGÍSTICA (CALENDARIO Y COSTOS)</b>	<b>31</b>
CRONOGRAMA.	31
RECURSOS.	31
<b>ANÁLISIS ESTADÍSTICO.</b>	<b>32</b>
<b>CONSIDERACIONES ÉTICAS</b>	<b>35</b>
<b>RESULTADOS</b>	<b>37</b>
RENDIMIENTO	44
<b>DISCUSIÓN</b>	<b>45</b>
<b>CONCLUSIONES.</b>	<b>49</b>
<b>BIBLIOGRAFÍA</b>	<b>50</b>



## TÍTULO.

CREACIÓN DE UN MODELO DE CLASIFICACIÓN BAYESIANO PARA IDENTIFICAR LA CONVERSIÓN DE PREDIABETES A DIABETES TIPO 2 EN PACIENTES DE LA CONSULTA EXTERNA DE MEDICINA INTERNA DEL HOSPITAL CENTRAL SUR DE ALTA ESPECIALIDAD PEMEX.

## DEFINICIÓN DEL PROBLEMA.

La prediabetes, un estado de alteración del metabolismo de la glucosa es una enfermedad con elevada prevalencia en México y el mundo, la cual confiere un riesgo incrementado no solo de conversión a diabetes, sino de desenlaces cardiovasculares.

Junto con la hipertensión arterial, el peso, perímetro abdominal, alteración en niveles de lípidos, engloba el llamado “SÍNDROME METABÓLICO”, teniendo como principal mecanismo fisiopatológico la resistencia periférica a la insulina.

Pese a los importantes avances en cuanto a descubrimiento de mecanismos fisiopatológicos, identificación de blancos terapéuticos y estudio de la historia natural de la enfermedad, cambiando el paradigma de la atención del paciente centrada en la glucosa, abriendo el panorama a la realización de que no es una enfermedad únicamente metabólica, sino que es al mismo tiempo una enfermedad cardiovascular e inflamatoria, cambiando actualmente los algoritmos de manejo y seguimiento en pacientes diabéticos, sin embargo, en la pre- diabetes, no existe un consenso en cuanto al riesgo de progresión, qué pacientes deberán tener un seguimiento más estricto; dejando esta decisión en el juicio clínico y en los niveles de glucosa, extendiendo el modelo glucocéntrico existente en antaño. Una posible solución es la implementación de índices de progresión e índices pronóstico, empero, existen pocos índices, y en menos aún aplicables a población mexicana, que puedan predecir la progresión de prediabetes a diabetes, así como su riesgo de desenlaces cardiovasculares y circulatorios.

La creación de un índice que tome en cuenta medidas más allá de los niveles de la glucosa, contemplando al paciente que vive con prediabetes de una manera holística, tomando la glucosa como una manifestación más de una enfermedad multifactorial, con múltiples y variables manifestaciones clínicas y bioquímicas, podría ayudarnos a mantener una mejor vigilancia y establecer estrategias de seguimiento que permitan al paciente revertir el proceso inflamatorio que de no ser frenado, llevará al paciente a complicaciones varias que afectan no solo su

expectativa de vida, también su calidad de vida, así como la de su entorno familiar-social.

Esta necesidad ya fue vislumbrada a lo largo del globo, teniendo ejemplos a lo largo de los distintos continentes, sin embargo, así como existe una gran variedad geográfica, climática, social, ambiental, económica, etc. entre las diversas poblaciones del mundo, los individuos de cada región presentan características únicas que hacen difícil la aplicación de un modelo universal, esta problemática afecta a todas las áreas de la medicina, en México por ejemplo, en el área de la reumatología, se creó el MEX-SLEDAI para la medición de la actividad lúpica

Un ejemplo de una herramienta diseñada con este objetivo es la desarrollada por población japonesa, donde se estudiaron los factores implicados en la progresión de este fenómeno metabólico, sin embargo, como ya se mencionó previamente, no fue diseñada para una población con el fenotipo característico de la población mexicana.

En el año 2021, en México, se diseñó modelo predictor que utilizó como método de análisis discriminatorio por regresión logística, dicho modelo obtuvo un área bajo la curva de 0.816 (0.729- 0.903), sin embargo, tanto el japonés como el mexicano, se han visto superados por modelos de machine learning.

Por lo anterior todo lo anterior planteamos la necesidad de realizar un modelo validado para población mexicana, que a su vez utilice un método de machine learning para su desarrollo para la obtención de un mejor rendimiento diagnóstico representado en el área bajo la curva de la curva ROC, optando por el modelo naive de Bayes para este objetivo.

Con el contexto antes expuesto, el objetivo del estudio es Desarrollar un modelo de clasificación bayesiano para identificar la conversión de prediabetes a diabetes tipo 2 con base en variables demográficas, clínicas y bioquímicas en pacientes de la consulta externa de medicina interna del Hospital Central Sur de Alta Especialidad PEMEX.

## MARCO TEÓRICO Y ANTECEDENTES

### 1.- Prediabetes: Definición y estadística en México y el Mundo.

La prediabetes es un estado tóxico de homeostasis alterada de la glucosa, caracterizado por niveles de glucosa por encima de los considerados “normales”, sin llegar a cumplir con el criterio requerido para el diagnóstico de diabetes; que confiere un riesgo elevado para el desarrollo de diabetes, así como complicaciones cardiovasculares.<sup>1-2.</sup>

Revisando las últimas guías de la asociación americana de la diabetes estas cifras alteradas de glucosa se definen operacionalmente como una glucosa plasmática en ayuno con niveles entre 100- 125 miligramos/ decilitro; una hemoglobina glicosilada entre 5.7- 6.4% o una glucosa plasmática a las dos horas de la prueba de tolerancia oral de 75 gramos de glucosa de 140 a 199 miligramos/ decilitro.<sup>3</sup> La prediabetes no debe ser vista como una entidad clínica aislada, sino como un factor de riesgo para el desarrollo de diabetes y enfermedades cardiovasculares.<sup>4</sup>

Esta entidad, al igual que la diabetes, debe ser estudiada en los siguientes grupos poblacionales por el riesgo incrementado que tienen de desarrollarlo: pacientes con obesidad y sobrepeso que cuenten con uno o más de los siguientes factores de riesgo: familiar de primer grado con diagnóstico de diabetes, etnia/ raza de alto riesgo, historia de enfermedad cardiovascular, diagnóstico de hipertensión o uso de antihipertensivos, niveles de colesterol de alta densidad menores a 35 miligramos/ decilitro y/o triglicéridos mayores a 250 miligramos/ decilitro, diagnóstico de síndrome de ovario poliquístico, sedentarismo, acantosis nigricans y obesidad severa. Para prediabéticos específicamente, este tamizaje deberá ser realizado de manera anual para evaluar conversión.<sup>4</sup>

La prevalencia de esta enfermedad en México, según ENSANUT 2018 es de 14.1%, y del 12.4% al 28.7% en Estados Unidos de América. <sup>5-6</sup>. Específicamente en nuestro centro de investigación, la prevalencia se eleva hasta un 33.1%, el doble que la prevalencia nacional y un 10% más que la estadística en Estados Unidos. <sup>7</sup>

La tasa de conversión anual de prediabetes a diabetes tipo dos varía entre el 3% al 11% al año. <sup>8</sup>

La prediabetes confiere un riesgo relativo para el desarrollo de diabetes de 7.7 (IC 95%= 2.1- 27.9) con una incidencia acumulada de conversión dentro del primer año de 4.5% confirmando de este modo el estatus de factor de riesgo para desarrollo de diabetes tipo 2; <sup>7</sup> Del mismo modo, confiere un riesgo incrementado complicaciones cardiovasculares pese a no presentar una alteración metabólica dentro de los rangos de diabetes tipo 2. Un ejemplo de esto es el incremento de riesgo de cardiopatía isquémica con riesgo relativo estudiado entre 2.13 a 3.12 para enfermedad arterial coronaria, 1.7 (IC 1.24- 2.6) para eventos cardiovasculares en general y 1.8 (ic 1.13- 2.86) para mortalidad por cualquier causa <sup>9-10</sup>. Además, también muestra un incremento en la prevalencia complicaciones microvasculares: renales con una prevalencia estimada de microalbuminuria en esta población de 15.5%, oftalmológicas, gastrointestinales y neurológicas con aproximadamente 1- 25% de las personas con prediabetes presentando evidencia de neuropatía. <sup>11</sup>

Como se mencionó previamente, la prediabetes en sí es un factor de riesgo para la conversión a diabetes tipo 2, por lo que a partir de este punto, nos enfocaremos en la Diabetes tipo 2, sus mecanismos fisiopatológicos y complicaciones cardiovasculares.

## 2.- Fisiopatología de la diabetes tipo 2: ¿Una enfermedad metabólica, inflamatoria o cardiovascular?

La diabetes tipo 2 se caracteriza por una alteración en la regulación en el metabolismo de carbohidratos, lípidos y proteínas, resultado de una alteración en la secreción de insulina, resistencia a la misma, o ambos. De todos los tipos de diabetes, la tipo 2 representa a más del 90% de los casos. <sup>8</sup>

Los mecanismos fisiopatológicos detrás de la diabetes tipo 2 han ido incrementando en una sucesión de descubrimientos, desde el modelo simplista de la deficiencia de insulina secundaria a disfunción de las células beta pancreáticas, y la resistencia a la misma, pasando por el octeto ominoso, actualmente mencionándose hasta 18 mecanismos fisiopatológicos “ los 18 endulzantes”, incluyendo: 1) Incremento en la lipólisis, 2) Inflamación, 3) Deficiencia en el efecto incretina, 4) Sobreproducción de glucosa hepática, 5) Efecto contrarregulador de las catecolaminas, teorizándose la posibilidad de que los pacientes con diabetes tipo 2 tengan un pico matutino en el tono dopaminérgico, llevando a un incremento en la actividad simpática, incrementando a su vez la resistencia a la insulina en la mañana, 6) Deficiencia de vitamina D, 7) Alteraciones en el sistema Renina- angiotensina- aldosterona, 8) Deficiencia de testosterona, donde en el estudio NHANES III se encontró que los hombres con los niveles de testosterona en el tercil inferior presentaron 4 veces más probabilidad de desarrollo de diabetes tipo 2 en comparación con aquellos en el tercil superior, 9) Disminución en niveles de melatonina nocturna, estudiado en personal de enfermería guardias rotativas en comparación con guardias fijas, presentando las enfermeras de guardias fijas un incremento en la incidencia de diabetes tipo 2 con el paso de los años, 20% a los 3- 9 años, 40 % entre 10 y 19 años y 60% con más de 20 años en este esquema de guardias 10) Gluconeogénesis renal, 11) Incremento en la expresión de SGLT 1 intestinal, 12) Alteración en la microbiota del tubo digestivo, entre otros. <sup>12-15</sup>

Todos estos avances en el estudio de la fisiopatología de la diabetes han cambiado el paradigma de la diabetes de una enfermedad metabólica centrada en la glucosa, en una enfermedad inflamatoria. La hipótesis para el desarrollo de la diabetes como una enfermedad inflamatoria parte del nivel excesivo de nutrientes, incluyendo la glucosa y los ácidos grasos libres, causan estrés oxidativo en los islotes pancreáticos y en los órganos sensibles a la insulina como el hígado, músculo y tejido adiposo, llevando a una producción local de citocinas y quimiocinas, teniendo como consecuencia el reclutamiento de células inmunes, contribuyendo a la inflamación del tejido pancreático (islotes pancreáticos) y a la resistencia periférica a la insulina. <sup>16</sup>

### 3.- Conversión de prediabetes a diabetes tipo 2.

Existen muchas variables que se han visto asociadas en la progresión de prediabetes a diabetes tipo 2 entre los que se han descrito factores antropométricos como índice de masa corporal y perímetro de cintura, factores clínicos como la presión arterial, y factores bioquímicos incluyendo alteraciones en pruebas de funcionamiento hepático, glicemia en ayuno, glucosa en la primera visita, hemoglobina glicosilada y dislipidemia <sup>17-18</sup>

Otro factor que se ha visto relacionado con el desarrollo y complicaciones de la diabetes tipo 2 es la resistencia a la insulina, estimada por índices como el HOMA- IR, sin embargo, siendo poco aplicables por la poca disponibilidad de obtener los niveles séricos de insulina, por el motivo anterior, han sido desarrollados modelos no insulínicos que miden esta variable. Tal es el caso del índice METS- IR desarrollado en el Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán por O Y Bello-Chavolla, et. al como predictor de diabetes tipo 2 incidente, demostrando un área bajo de curva de 0.845 (IC 95% 0.783- 0.899) <sup>19</sup>

#### 4.- El síndrome metabólico.

Cómo se vió previamente, entre los criterios para tamizaje de diabetes/ prediabetes y los factores de riesgo para conversión de prediabetes en diabetes tipo 2 se encuentra el llamado síndrome metabólico.

Originalmente conocido como el síndrome X, propuesto por Reaven en 1988, teorizando la posibilidad de que la resistencia a la insulina no está involucrada únicamente en la etiología de la diabetes, sino también en la enfermedad cardiovascular; se describió como 1) Resistencia a la insulina, 2) disglucemia, 3) dislipidemia, 4) hipertensión; siendo valorado por seis variables: circunferencia de cintura, niveles de glucosa en ayuno, niveles de triglicéridos, niveles de colesterol total, niveles de colesterol HDL y niveles de presión arterial.

Es una enfermedad con una prevalencia elevada, la cual es de 24% en hombres y 22% en mujeres. Y aproximadamente el 85% de los pacientes con diabetes tipo 2 también presentan síndrome metabólico <sup>20</sup>

Por lo anterior descrito, por su fisiopatología compartida con la diabetes tipo 2, el siguiente paso lógico para el desarrollo de un modelo predictivo es evaluar la diferencia en la función de distribución con el desarrollo de diabetes. Tipo 2 para su posterior inclusión en el modelo.



## 5.- Modelos de conversión.

Existen diferentes tipos de modelos en la literatura, utilizando distintas variables en distintos grupos poblacionales y con diferentes métodos para la creación de los modelos predictivos.

Yokota, et. al en 2017 realizó un estudio longitudinal retrospectivo con 2105 adultos con prediabetes con una observación a 4.7 años; Yokota, et. al. Realizó una regresión logística multivariada con puntuación de riesgo para la realización de su modelo; encontrando como variable estadísticamente significativas la historia familiar positiva para idabetes, sexo hombre, presión arterial sistólica elevada, niveles de glucosa plasmática, alanina aminotransferasa y hemoglobina glicosilada; obteniendo un AUC de 0.8 (IC 95% 0.7- 87) con una especificidad de 0.61 a una sensibilidad de 0.8. <sup>21</sup>

Con el afán de realizar un modelo similar en población mexicana, Shimizu, et. al en 2021 realizó un estudio longitudinal retrospectivo, utilizando al igual que Yokota un regresión logística multivariada, en este caso encontrando como variables estadísticamente significativas la concentración de glucosa medida en la primera determinación, el índice de masa corporal y la diferencia positiva de glucosa al año de seguimiento, obteniendo al final de su estudio un modelo con un AUC (IC 95%) de 0.816 (0.729- 0.903) con una especificidad de 75.9% para una sensibilidad de 80.5%. <sup>22</sup>

Entre los modelos existentes además de los ya mencionados; Yokota, et. al en población japonesa y Seiichi, et. al en población mexicana; en 2018 fue desarrollado por Nichols, et. al un modelo basado en hemoglobina glicosilada e índice de masa corporal estratificado en cortes establecidos por el autor, sin embargo, en este estudio únicamente se reportaron riesgos relativos, sin establecer la sensibilidad y especificidad del modelo. <sup>23</sup> Otro modelo es el desarrollado en 2019 por Cahn et. al, utilizando técnicas de machine learning supervisado con un árbol de decisiones, logró un área bajo la curva ROC de 0.865 (IC 0.86- 0.869) utilizando como variables

significativas para la progresión: la edad, glucosa en ayuno, hemoglobina glicosilada, índice de masa corporal, triglicéridos, alanina aminotransferasa, leucocitos, uso de estatinas, lipoproteína de alta densidad y uso de aspirina. <sup>24</sup>

## 6.- Aprendizaje Automatizado

Abandonando la idea del modelo probabilístico de generación de datos (asociado con la habilidad de calcular valores  $p$ , intervalos de confianza, etc), el aprendizaje automatizado apuesta por estructuras algorítmicas genéricas (no lineales) entrenadas para realizar ciertas funciones con el objetivo de minimizar una pérdida de función general que no sea ligada a la probabilidad de las observaciones. El ideal de los algoritmos de aprendizaje automatizado es tomar una estructura algorítmica y posteriormente ajustar sus parámetros a los datos, adaptando al mismo tiempo su complejidad, optimizando el intercambio sesgo contra varianza para que el modelo generalice a nueva información.

En general estos modelos son clasificados como supervisados y no supervisados. En los modelos de aprendizaje supervisado, son presentados al algoritmo ejemplos de correcta ejecución de las funciones, y el modelo es entrenado para minimizar las diferencias entre sus acciones y las acciones “correctas”. Por otro lado, en el aprendizaje no supervisado se refiere a tareas de las cuales no se dan ejemplos, y los algoritmos optimizan una pérdida de función general. <sup>25</sup>

## 7.- Modelo Naive Bayesiano.

Los modelos clasificadores Naive Bayesianos son algoritmos de aprendizaje automatizado supervisado, usados para tareas de clasificación, el cual busca la distribución de las variables de una clase o categoría, estos modelos no aprenden qué características son las más importantes para diferenciar entre clases.

Como su nombre lo indica, está basado en el teorema de Bayes, el cual menciona que las probabilidades condicionales, representan la probabilidad de un evento dado que otro evento ocurrió, y se distingue por el uso de eventos secuenciales, donde la información adicional adquirida a posteriori impacta la probabilidad inicial.

Los clasificadores Naïve Bayesianos trabajan sobre un par de asunciones clave (de ahí el nombre naive). Asuma que los predictores son condicionalmente independientes de cualquier otra variable del modelo, y asume que todas las variables contribuyen de igual manera al desenlace. Pese a la irreal asunción de independencia, este algoritmo de clasificación tiene un buen desempeño con tamaños de muestra pequeños.<sup>26</sup>

Dada la baja complejidad del modelo naive bayesiano y su buen desempeño con tamaños de muestra pequeños, consideramos por ende que para el desarrollo de un modelo predictor el clasificador naive bayesiano es el método ideal para nuestros objetivos.

## 8.- La importancia de un nuevo modelo.

La importancia del desarrollo de índices para poblaciones específicas radica en la necesidad de retrasar la historia natural de esta enfermedad para prevenir la aparición temprana de complicaciones, que una vez establecidas no son reversibles.

El desarrollo de un modelo aplicable a características de la población mexicana será de especial utilidad para los médicos de atención de primer nivel, para establecer a partir de una estratificación de riesgo, estrategias para evitar la progresión de la enfermedad.

Los modelos existentes en la actualidad no tienen validación en población mexicana, por tanto, es imperativo el desarrollo y validación de un modelo aplicable a las características de esta población.

## JUSTIFICACIÓN.

Existe una gran tasa de conversión anual de prediabetes a diabetes, con todas las complicaciones que ello implica, y más que las recomendaciones de seguimiento anual establecidas por los órganos creadores de guías como la Asociación Americana de la Diabetes, no existe una herramienta universal para predecir esta conversión y así poder tener un monitoreo más estrecho de los pacientes.

Existen diversos modelos creados con esta meta, sin embargo, estos han sido desarrollados en diversas poblaciones, por lo que es de especial importancia y será de gran utilidad para los médicos de primer nivel en México el poder contar con una herramienta de estas características para el seguimiento de sus pacientes con diagnóstico de prediabetes.

Para tal propósito será esencial identificar qué variables en esta población tienen una diferencia estadísticamente significativa en la función de distribución para el desarrollo de diabetes tipo 2.

Con el auge del machine learning, el siguiente paso lógico es su implementación en diversas áreas de la medicina, por lo cual, en esta ocasión, nos proponemos crear un modelo con un buen rendimiento ( $AUC > 0.8$ ), utilizando variables clínicas y bioquímicas en pacientes de población mexicana basado en el modelo Naive de Bayes.

## PLANTEAMIENTO DEL PROBLEMA

La prediabetes es un estado de alteración en el metabolismo de la glucosa, la cual puede progresar a diabetes, tener regresión a normoglucemia o permanecer como prediabetes. <sup>27</sup>

Pese a los importantes avances en cuanto a descubrimiento de mecanismos fisiopatológicos, identificación de blancos terapéuticos y estudio de la historia natural de la enfermedad, existen pocos índices, y en menos aún aplicables a población mexicana, que puedan predecir la progresión de prediabetes a diabetes, y que puedan permitirnos tener un enfoque más agresivo para detener y de ser posible revertir la historia natural de la enfermedad.

Un ejemplo de una herramienta diseñada con este objetivo es la desarrollada por Yokota, et. al. en 2017 quien en un estudio retrospectivo con 2105 adultos con prediabetes en población japonesa estudió los factores implicados en la progresión de este fenómeno metabólico, creando un modelo predictivo que tomó en cuenta historia familiar positiva, sexo masculino, presión sistólica incrementada, prueba de tolerancia a glucosa oral, hemoglobina glicosilada y alanina aminotransferasa, obteniendo una predicción de diabetes tipo 2 incidente con un área bajo la curva de 0.8 (IC 95% 0.70- 0.87) <sup>21</sup>, sin embargo, como ya se mencionó previamente, no fue diseñada para una población con el fenotipo característico de la población mexicana, considerando no significativas en población japonesa medidas antropométricas que si influyen en población latina.

En el año 2021 fue diseñado por Seiichi, et. al. un modelo predictor, quien encontró como variables relacionadas a la progresión de la enfermedad, la concentración de glucosa inicial, índice de masa corporal y la diferencia positiva de glucosa al año de seguimiento, obteniendo un área bajo la curva de 0.816 (0.729- 0.903) <sup>22</sup> alcanzando un rendimiento similar al de su contraparte japonesa.

Sin embargo, el modelo mencionado previamente no cuenta con validación para su aplicación, por ende, el objetivo de este estudio es el perfeccionamiento y validación del modelo previamente elaborado enfocado a población mexicana.

## PREGUNTA DE INVESTIGACIÓN.

¿El nuevo modelo de clasificación bayesiana logrará una eficiencia superior a 0.8 en el área bajo la curva ROC en predicción de progresión de prediabetes a diabetes tipo 2 incidente en población mexicana del Hospital Central Sur de Alta Especialidad?

## HIPÓTESIS.

### Hipótesis nula.

El nuevo modelo a desarrollar NO logrará una eficiencia superior a 0.8 en área bajo la curva de ROC en predicción de progresión de prediabetes a diabetes tipo 2 incidente en población mexicana.

### Hipótesis alterna.

El nuevo modelo a desarrollar logrará una eficiencia superior a 0.8 en área bajo la curva de ROC en predicción de progresión de prediabetes a diabetes tipo 2 incidente en población mexicana.

## OBJETIVOS

### Objetivo general.

Desarrollar un modelo de clasificación bayesiano para identificar la conversión de prediabetes a diabetes tipo 2 con base en variables demográficas, clínicas y bioquímicas en pacientes de la consulta externa de medicina interna del Hospital Central Sur de Alta Especialidad PEMEX.

### Objetivos específicos.

Obtener variables bioquímicas iniciales del expediente clínico electrónico de pacientes de la consulta externa de medicina interna del Hospital Central Sur de Alta Especialidad PEMEX con diagnóstico de prediabetes.

Describir y Comparar las variables demográficas, clínicas y bioquímicas entre los grupos de conversión y no conversión para identificar cuales presentan una diferencia o asociación estadísticamente significativa.

Construcción del modelo de Clasificación Bayesiano únicamente con las variables de la primera medición que mostraron ser estadísticamente significativas entre los grupos de conversión.

Validación del modelo de Clasificación Bayesiano con el 20% de la muestra obtenida.

## TIPO DE ESTUDIO.

Se realizó un estudio de tipo observacional/ descriptivo de características transversales por temporalidad, retrospectivo por la obtención de los datos, utilizando variables clínicas y bioquímicas del expediente clínico electrónico del Hospital Central Sur de Alta Especialidad de PEMEX; además de un estudio de pruebas diagnósticas utilizando la técnica de aprendizaje automatizado clasificador naive bayesiano.



## METODOLOGÍA.

### Universo.

Pacientes adscritos al Hospital Central Sur de Alta Especialidad de Petróleos Mexicanos.

### Unidades de observación.

Hombres y mujeres de 18- 75 años que hayan acudido al laboratorio clínico del Hospital Central Sur de Alta Especialidad de Petróleos Mexicanos entre 2010 y 2015 con resultado de glucosa en ayuno, curva de tolerancia a la glucosa oral o hemoglobina glicosilada que cumpla con los criterios establecidos por la Asociación Americana de la Diabetes 2022 para el diagnóstico de prediabetes, que además hayan tenido cita en el servicio de nutrición con obtención de medidas antropométricas y cuenten con determinaciones de pruebas de función hepática, concentración de lípidos (colesterol, triglicéridos y colesterol HDL, ácido úrico y creatinina utilizando ensayos colorimétricos, calculando los niveles de colesterol LDL con la ecuación de Friedwald cuando los triglicéridos sean menores a 250 miligramos/decilitro.

### Tipo de muestreo.

No aleatorizado, a conveniencia.

### Tamaño de muestra.

Para la estimación del tamaño de muestra nos basamos en la prevalencia de la Pre-diabetes que fue del 14.1% en la ENSANUT 2016. Se estableció una significancia estadística para todas las pruebas aplicadas menor o igual a 0.05 y un poder estadístico del 0.80, además un AUC de al menos 0.80 contra una hipótesis nula de AUC de 0.60 y una relación 1:1 de los casos. Se presenta la siguiente imagen del software MEDCALC, donde fue estimado el tamaño de muestra con los parámetros mencionados:

Sample size: area under ROC curve					
<b>Options</b>					
Type I error (Alpha, Significance)	0.05				
Type II error (Beta, 1-Power)	0.20				
<b>Data</b>					
Area under ROC curve	0.8				
Null Hypothesis value	0.6				
Ratio of sample sizes in negative / positive groups	1				
<b>Result</b>					
Number of positive cases required:	28				
Number of negative cases required:	28				
Total sample size (both groups together)	56				
<b>Table</b>					
		Type I Error - Alpha			
		0.20	0.10	0.05	0.01
Type II Error - Beta	0.20	15 + 15	22 + 22	28 + 28	42 + 42
	0.10	21 + 21	28 + 28	35 + 35	51 + 51
	0.05	27 + 27	35 + 35	42 + 42	60 + 60
	0.01	39 + 39	48 + 48	57 + 57	77 + 77

Monday, May 15, 2023 06:55  
 MedCalc® Statistical Software version 22.001 (MedCalc Software Ltd, Ostend, Belgium; <https://www.medcalc.org>; 2023)

Para lograr el principal objetivo del presente trabajo de investigación nos plantearemos una meta de 56 pacientes, con 28 de ellos que presenten el evento de conversión de pre-diabetes a diabetes mellitus. (18)

**Criterios de inclusión:**

Pacientes con diagnóstico de prediabetes según criterios de la ADA 2022 entre 18 y 75 años.

Índice de masa corporal entre 18- 34.9 kilogramos/ metro cuadrado.

**Criterios de exclusión:**

Diagnóstico de diabetes tipo 2, uso de insulina o hipoglicemiantes orales.

Enfermedades endocrinológicas además de prediabetes. (páncreas, tiroides, suprarrenales, hipófisis)

Embarazo.

Historia de enfermedades cardiovasculares.

Consumo de esteroides.

## VARIABLES DE ESTUDIO

### Variable(s) independiente(s)

- Edad: expresada en años a partir de fecha de nacimiento como variable cuantitativa discreta.
- Sexo: expresada como el sexo asignado al nacer como variable cualitativa categórica dicotómica.
- Glucosa en ayuno: Cifra de glucosa obtenida con al menos 8 horas de ayuno reportada en el expediente clínico electrónico como variable cuantitativa discreta.
- Hemoglobina glicosilada: Cifra del porcentaje de glicosilación de la cadena alfa 1 de la hemoglobina obtenida de sangre periférica por medio de cromatografía líquida de alto desempeño como variable cuantitativa continua.
- Glucosa posterior a 2 horas de prueba de tolerancia oral a 75 gramos de glucosa expresada en miligramos/ decilitro como variable cuantitativa discreta.
- Peso: Cifra en kilogramos obtenida de báscula calibrada reportada en el expediente clínico electrónico como variable cuantitativa discreta.
- Talla: obtenida en la consulta externa con estadímetro de piso, reportada en metros en el expediente clínico electrónico como variable cuantitativa continua.

- Índice de masa corporal: Calculada como el peso en kilogramos entre la altura en metros al cuadrado, reportada en el expediente clínico electrónico como variable cuantitativa continua.
- Tensión arterial sistólica: Obtenida por el servicio de enfermería con esfigmomanómetro manual o electrónico, reportado en el expediente clínico electrónico como variable cuantitativa discreta.
- Tensión arterial diastólica: Obtenida por el servicio de enfermería con esfigmomanómetro manual o electrónico, reportado en el expediente clínico electrónico como variable cuantitativa discreta.
- Perímetro Abdominal: Obtenida con cinta métrica con técnica estandarizada, expresada en centímetros en el expediente clínico electrónico en la consulta de nutrición como variable cuantitativa discreta.
- Colesterol HDL: Niveles de colesterol de alta densidad obtenida de muestra sanguínea posterior a 8- 12 horas de ayuno por medio de ensayo colorimétrico, expresada como variable cuantitativa discreta.
- Colesterol LDL: Calculada con la ecuación de Friedewald expresada en el expediente clínico electrónico como variable continua discreta.
- Triglicéridos: Determinación sérica de niveles de triglicéridos usando ensayo colorimétrico, expresado en el expediente clínico electrónico como variable continua discreta.

- Creatinina: Niveles de creatinina sérica obtenida de sangre periférica por medio de ensayo colorimétrico, expresada como miligramos/ decilitro en el expediente clínico electrónico como variable cuantitativa continua.
- Tasa de filtrado glomerular: Tasa de filtrado glomerular calculado a partir de la creatinina sérica con la fórmula CKD- EPI 2021.
- Transaminasa glutámico oxalacética: Determinación de transaminasa glutámico oxalacética registrada en el expediente clínico electrónico como unidades/ litro siendo variable cuantitativa discreta.
- Transaminasa glutámico pirúvica: Determinación de transaminasa glutámico pirúvica registrada en el expediente clínico electrónico como unidades/ litro siendo variable cuantitativa discreta.
- Deshidrogenasa láctica: Determinación de deshidrogenasa láctica en muestra de sangre periférica reportada en el expediente clínico electrónico como variable cuantitativa discreta.
- Fosfatasa alcalina: Determinación sérica de la fosfatasa alcalina en muestra de sangre periférica reportada en el expediente electrónico y representada como variable cuantitativa discreta.
- Ácido úrico: Determinación sérica de ácido úrico en muestra de sangre periférica reportada en el expediente electrónico y representada como variable cuantitativa continua.
- Antecedente heredofamiliar de primera línea de Diabetes tipo 2: representado como variable categórica nominal dicotómica.

- Historia de diabetes gestacional: diagnóstico establecido con los criterios de la asociación americana de la diabetes, representado como variable categórica nominal dicotómica.
- Actividad física: tiempo semanal dedicado a la actividad física, expresado de acuerdo al sistema horario como variable cuantitativa continua.
- Etanol: cantidad de alcohol consumida expresada en gramos a la semana, representado como variable cuantitativa continua.
- Tabaco: expresada con el índice tabáquico, representada como variable cuantitativa continua.
- Comorbilidades: Presencia de cualquier otro diagnóstico excluyendo prediabetes, expresado como variable categórica nominal multicotómica.
- Medicamentos: consumo de fármacos de manera crónica para tratamiento de una enfermedad, representado de acuerdo al mecanismo de acción del fármaco como variable cualitativa nominal multicotómica.

## Variable(s) dependiente(s)

- Conversión de prediabetes a diabetes: cambio de diagnóstico basado en los criterios establecidos por la última revisión de la asociación americana de la diabetes, representado como variable categórica nominal dicotómica.
- Reversión de prediabetes a normoglucesmia: cambio de diagnóstico por criterios establecidos por la última revisión de la asociación americana de la diabetes, representado como variable categórica nominal dicotómica.

## RECOLECCIÓN DE DATOS.

Se obtendrá información en el expediente clínico electrónico de pacientes que hayan acudido a la consulta externa del Hospital Central Sur de Alta Especialidad de PEMEX con diagnóstico de prediabetes de acuerdo con los criterios establecidos previamente en el protocolo entre el año 2010 y 2015, se tomará como el día 0 el día con resultado de laboratorio alterados en el historial de análisis clínicos.

Se revisarán resultados de laboratorios previos al diagnóstico para encontrar alteraciones previas. A partir del día 0 se revisará el expediente clínico electrónico para realizar una historia médica y familiar completa incluyendo a aquellos que cumplan los criterios de inclusión y descartando a los que presenten criterios de exclusión o eliminación.

Se incluirán aquellos pacientes que tengan actualización de las variables de estudio al año de la primera medición tanto clínicas como bioquímicas.



Se registrarán en la base de datos las variables de los primeros dos años de seguimiento. Se registrarán las variables categóricas como porcentaje, las variables cuantitativas con distribución normal se expresarán como el valor puntual con su desviación estándar, y las variables cuantitativas con distribución no normal serán expresadas como el valor puntual con su rango intercuartílico.

El seguimiento se dará de manera anual de acuerdo con las recomendaciones de la asociación americana de la diabetes para pacientes prediabéticos. Se considerará como desenlace el diagnóstico de Diabetes tipo 2 a partir de dos pruebas diagnósticas distintas que cumplan los criterios establecidos por la asociación americana de la diabetes. Los pacientes que pierdan seguimiento serán considerados como eventos de diagnóstico de diabetes tipo 2.

## RECURSOS Y LOGÍSTICA (CALENDARIO Y COSTOS)

### Cronograma.

#	ACTIVIDAD	Mes Calendario Programado AÑO 2023											
		1	2	3	4	5	6	7	8	9	10	11	12
	Recolección de datos.	X	X										
	Desarrollo de modelo.			X									
	Valoración del desempeño del modelo.				X								
	Validación del modelo.					X							
	Análisis estadístico.						X						

### Recursos.

CONCEPTO	NÚMERO	COSTO (paciente, muestra, etc.)	UNITARIO caso, encuesta,	SUBTOTAL
<b>Recursos Materiales</b>				
<i>Computadora personal.</i>	1	N/A		0 MXN
<i>Computadora con acceso a expediente clínico electrónico</i>	1	N/A		0 MXN

## ANÁLISIS ESTADÍSTICO.

OBJETIVO (S)	VARIABLE (S)	ANÁLISIS ESTADÍSTICO
<p>Describir y comparar las variables demográficas, clínicas y bioquímicas entre los grupos de conversión y no conversión para identificar cuales presentan una diferencia o asociación estadísticamente significativa.</p>	<ul style="list-style-type: none"> <li>● Edad.</li> <li>● Sexo.</li> <li>● Glucosa en ayuno.</li> <li>● Hemoglobina glicosilada:</li> <li>● Peso.</li> <li>● Talla.</li> <li>● Índice de masa corporal.</li> <li>● Tensión arterial sistólica.</li> <li>● Tensión arterial diastólica.</li> <li>● Perímetro Abdominal</li> </ul>	<p>Prueba t de Student para comparación de medias o prueba U de Mann Whitney, en función de la distribución que presenten las variables.</p> <p>Prueba <math>\chi^2</math> de independencia para evaluar asociación entre variables.</p>

	<ul style="list-style-type: none"><li>● Colesterol HDL.</li><li>● Colesterol LDL.</li><li>● Triglicéridos .</li><li>● Creatinina.</li><li>● Tasa de filtrado glomerular.</li><li>● Transaminasa glutámico oxalacética.</li><li>● Transaminasa glutámico pirúvica.</li><li>● Deshidrogenasa láctica.</li><li>● Fosfatasa alcalina.</li><li>● Ácido úrico.</li><li>● Antecedente hereditario de primera línea de Diabetes tipo 2.</li></ul>	
--	---	--

	<ul style="list-style-type: none"> <li>• Historia de diabetes gestacional</li> <li>• Actividad física.</li> <li>• Etanol.</li> <li>• Tabaco.</li> <li>• Comorbilidades.</li> <li>• Medicamentos.</li> </ul>	
Construcción del modelo de Clasificación Bayesiano únicamente con las variables de la primera medición que mostraron ser estadísticamente significativas entre los grupos de conversión	Sólo las variables de la primera medición que mostraron ser estadísticamente significativa entre los grupos de conversión	Técnica de aprendizaje supervisado: Naive Bayes Classifier (Clasificador Bayesiano Ingenuo) utilizando la paquetería "naive bayes" del software R.
Validación del modelo de Clasificación Bayesiano con el 20% de la muestra obtenida.	Sólo las variables de la primera medición que mostraron ser estadísticamente significativa entre los	Curva ROC y estimación de los parámetros de rendimiento: Sensibilidad Especificidad VPP, VPN y AUC.

	grupos de conversión	
--	-------------------------	--

## CONSIDERACIONES ÉTICAS

Los investigadores confirmamos que la revisión de los antecedentes científicos del proyecto justificaron su realización, nos comprometimos a mantener un estándar científico elevado para obtener información útil y salvaguardar la confidencialidad de los datos de los participantes del estudio, se puso el bienestar y la seguridad de los pacientes sujetos de investigación por encima de cualquier otro objetivo; nos regimos de acuerdo con lo aceptado nacional e internacionalmente según lo establecido por la Ley General de Salud, Las pautas éticas internacionales para la investigación y experimentación biomédica en seres humanos de la OMS y la Declaración de Helsinki.

En apego a las normas éticas de la declaración de Helsinki y al artículo 17 del Reglamento de la Ley General de Salud en Materia de Investigación para la Salud, la participación de los pacientes en este estudio conlleva un tipo de riesgo: I (riesgo mínimo.) ya que implica técnicas y métodos de investigación documental que son retrospectivos y no se realiza ninguna intervención directa o indirecta con el paciente, tampoco se realizan modificaciones intencionadas o no intencionadas que pudieran provocar alguna alteración en el estado fisiológico, psicológico, hemodinámico o social de los individuos implicados durante la realización de esta investigación.

Basado en lo previamente expuesto, la carta de consentimiento informado no debe ser un requisito para solicitar la autorización para la realización del protocolo de investigación (“Título segundo, Capítulo I, Artículo 17, Sección I, investigación sin riesgo”)

De acuerdo con el Artículo 12 “De la información implicada en investigaciones de la NOM-012, SSA03-2012. Que establece los criterios para la ejecución de proyectos de investigación para la salud en seres humanos, el cual cita que la información relacionada con cualquier investigación deberá ser clasificada como confidencial, se emplearon para evitar la dispersión de los datos y el mal uso de los mismos las siguientes medidas.. 1) Los datos recabado del expediente clínico electrónico de Petróleos Mexicanos fueron utilizados y almacenados solamente por el residente responsable de la tesis y el asesor o asesores de tesis. 2) Los datos obtenidos del expediente clínico serán registrados con un número de ficha y no con el nombre del paciente. 3) Todos los datos serán manejados en una USB encriptada en formato ZIP que sólo las personas involucradas en este protocolo conocerán las cuales enlisto a continuación.

- Esteban Sepúlveda Oliva.
- Martín Coronado Malagón.
- José Roberto Mendoza Martínez.

4) Los datos serán utilizados con fines descriptivos para este estudio. 5) al finalizar el estudio, los datos obtenidos serán resguardados en formato ZIP en la USB designada. 5) Los datos serán almacenados con el fin de realizar futuras investigaciones, las cuales únicamente serán empleadas por los investigadores de esta tesis. 6) Los datos sólo podrán ser utilizados previa autorización por escrito por los investigadores titulares (Martín Coronado Malagón, José Roberto Mendoza Martínez), con copia a dirección médica y área de enseñanza del Hospital Central Sur de Alta Especialidad.

## RESULTADOS

Se obtuvo una muestra de un total de 116 pacientes con una edad promedio de 59 años y una desviación estándar de 14 años. En función de la presencia de la conversión de Prediabetes a Diabetes Mellitus se crearon dos grupos, el 40.52% (47) presentaron la Conversión de Prediabetes a Diabetes Mellitus y, complementariamente el 59.48% (69) no presentaron la Conversión ya que, mantuvieron niveles de Glucosa venosa central en ayuno menores al criterio diagnóstico. A continuación, se presenta la Tabla 1 y en ella se comparan las variables de tipo demográficas, antropométricas, Antecedentes Personal Patológicos y Familiares, por tipo de grupo.

**Tabla 1. Principales características entre grupos de Conversión.**

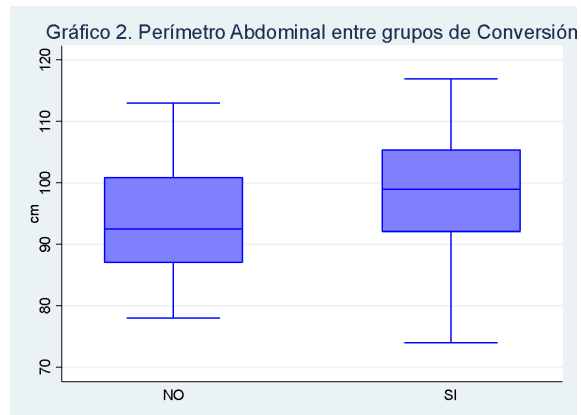
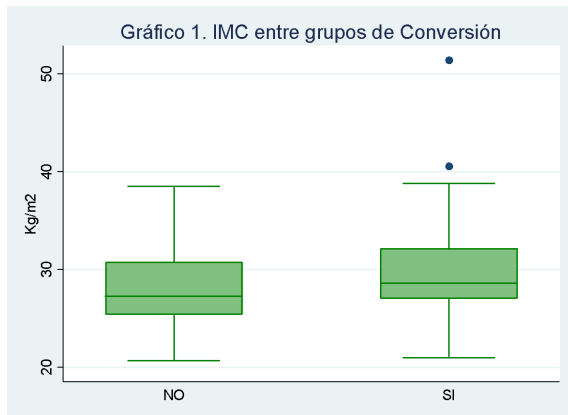
	Global/Categoría 100% (116)		Sin Conversión 59.48% (69)	Con Conversión 40.52% (47)	Valor p
<b>Edad</b>	59.2±14.1		58.9±13.5	59.7±15.1	0.768 <sup>z</sup>
<b>Sexo</b>	<b>Hombre</b>	37.07% (43)	39.13% (27)	34.04% (16)	0.578*
	<b>Mujer</b>	62.93% (73)	60.87% (42)	65.96% (73)	
<b>IMC</b>	28.0 (26.0 – 31.9)		27.2 (25.3 – 30.7)	28.6 (27.0 – 32.1)	<b>0.017<sup>a</sup></b>
<b>PA</b>	95.5 (88.5 – 103)		92.5 (87 – 101)	99 (92 – 105.5)	<b>0.014<sup>a</sup></b>
<b>TAS</b>	120 (120 – 130)		120 (120 – 130)	120 (115 – 130)	0.706 <sup>a</sup>
<b>TAD</b>	80 (70 – 80)		80 (70 – 80)	80 (70 – 80)	0.583 <sup>a</sup>
<b>AHF DM</b>	<b>Sí</b>	36.61% (41)	32.84% (22)	42.22% (22)	0.312*
<b>Ejercicio</b>	<b>Sí</b>	25.0% (29)	30.43% (21)	17.02% (8)	0.101*
<b>Etanol</b>	<b>Sí</b>	40.0% (46)	46.38% (32)	30.43% (14)	0.087*
<b>Tabaquismo</b>	<b>Sí</b>	46.96% (54)	55.07% (38)	34.78% (16)	<b>0.033*</b>
<b>Dislipidemia</b>	<b>Sí</b>	44.83% (52)	49.28% (34)	38.3% (18)	0.243*
<b>HAS</b>	<b>Sí</b>	52.59% (61)	50.72% (35)	55.32% (26)	0.627*
<b>Sobrepeso</b>	<b>Sí</b>	17.24% (20)	13.04% (9)	23.40% (11)	0.147*
<b>ERC</b>	<b>Sí</b>	2.59% (3)	2.90% (2)	2.13% (1)	1.000 <sup>Δ</sup>
<b>Gota</b>	<b>Sí</b>	6.03% (7)	5.80% (4)	6.38% (3)	1.000 <sup>Δ</sup>



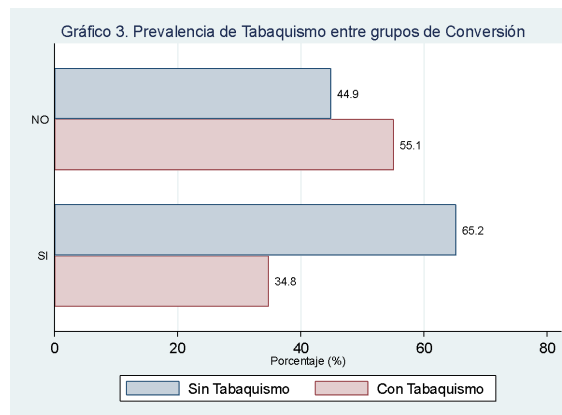
<b>PARK</b>	<b>Sí</b>	1.72% (2)	1.45% (1)	2.13% (1)	0.783 <sup>Δ</sup>
<b>CA</b>	<b>Sí</b>	2.59% (3)	2.90% (2)	2.13% (1)	1.000 <sup>Δ</sup>
<b>EPIL</b>	<b>Sí</b>	1.72% (2)	2.90 % (2)	0%	0.514 <sup>Δ</sup>
<b>ICC</b>	<b>Sí</b>	0.86% (1)	1.45% (1)	0%	1.000 <sup>Δ</sup>
<b>SICA</b>	<b>Sí</b>	1.72% (1)	2.90% (2)	0%	0.514 <sup>Δ</sup>
<b>ERGE</b>	<b>Sí</b>	7.76% (9)	7.25% (5)	8.51% (4)	1.000 <sup>Δ</sup>
<b>CARISQ</b>	<b>Sí</b>	4.31% (5)	5.80% (4)	2.13% (1)	0.647 <sup>Δ</sup>
<b>DEPRE</b>	<b>Sí</b>	3.45% (4)	5.80 % (4)	0%	0.146 <sup>Δ</sup>
<b>EAP</b>	<b>Sí</b>	1.72% (2)	0%	4.26% (2)	0.162 <sup>Δ</sup>
<b>HEP</b>	<b>Sí</b>	1.72% (2)	2.90% (2)	0%	0.514 <sup>Δ</sup>
<b>OSTEOP</b>	<b>Sí</b>	4.31% (5)	4.35% (3)	4.26% (2)	1.000 <sup>Δ</sup>
<b>DINSIPIDA</b>	<b>Sí</b>	0.86% (1)	1.45% (1)	0%	1.000 <sup>Δ</sup>
<b>FA</b>	<b>Sí</b>	1.72% (2)	1.45% (1)	2.13% (1)	1.000 <sup>Δ</sup>

*Σ Prueba de t de Student para muestras independientes para varianzas desiguales. α Prueba U de Mann-Whitney para muestras independientes. \*Prueba de Independencia  $\chi^2$ . Δ Prueba exacta de Fischer*

Como principales hallazgos, encontramos una diferencia estadísticamente significativa en la función de distribución para las variables IMC y Perímetro abdominal (PA), encontrando una mediana mayor en el grupo de Conversión comparado con el grupo Sin Conversión, para ambas variables. Estas diferencias podemos apreciarlas de manera visual en los Gráficos 1 y 2.



Se aprecia una asociación estadísticamente significativa entre el Tabaquismo y la Conversión de prediabetes a DM, encontrando una mayor prevalencia 55.07% (38) en el grupo Sin Conversión comparado con el grupo Con Conversión, siendo de 34.78% (16). Esta relación podemos apreciar visualmente en el Gráfico 3.



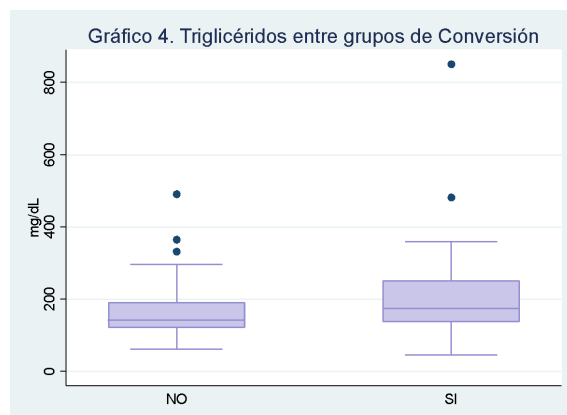
El resto de las variables presentadas en la Tabla 1 no mostraron una diferencia o asociación estadísticamente significativa.

A continuación, se presenta en la Tabla 2 la comparación de diversas variables bioquímicas entre grupos de Conversión. Cabe resaltar que, únicamente se incluyen las cuantificaciones de la primera consulta.

**Tabla 2. Determinaciones bioquímicas entre grupos de Conversión**

	<b>Global 100% (116)</b>	<b>Sin Conversión 59.48% (69)</b>	<b>Con Conversión 40.52% (47)</b>	<b>Valor p</b>
<b>Creatinina</b>	0.89 (0.73 – 0.99)	0.86 (0.73 – 1)	0.9 (0.74 – 0.95)	0.895*
<b>TFG</b>	84 (71 – 94)	84 (71 – 94)	81 (70 – 95.5)	0.888*
<b>Ácido Úrico</b>	5.7 (5 – 6.7)	5.65 (5.1 – 6.3)	5.7 (4.4 – 7.6)	0.796*
<b>Colesterol Total</b>	212 (186 – 249)	204 (189 – 255)	212 (186 – 245)	0.988*
<b>TAG</b>	158 (122 – 233)	142 (120 – 192)	174 (136 – 252)	<b>0.050*</b>
<b>HDL</b>	43 (36 – 54)	44 (36 – 54)	43 (37 – 54)	0.901*
<b>LDL</b>	129 (108 – 152)	131 (109 – 162)	126 (107 – 150)	0.527*
<b>TGO</b>	24 (20 – 31)	23 (20 – 30)	26.5 (23 – 38)	0.272*
<b>TGP</b>	27 (17 – 43)	26 (15 – 36)	31 (17 – 46)	0.305*
<b>DHL</b>	264 (164 – 370)	218 (191 – 360)	334 (145 – 370)	0.814*
<b>FA</b>	75.5 (70 – 77)	77 (70 – 84)	75 (69 – 75.5)	0.198*
<b>HbA1c</b>	5.8 (5.7 – 6.0)	5.8 (5.6 – 6)	5.95 (5.8 – 6.1)	0.510*
<i>* Prueba U de Mann-Whitney para muestras</i>				

Observamos que, la única variable bioquímica que presentó una diferencia estadísticamente significativa en la función de distribución de los Triglicéridos entre los grupos de Conversión, encontrando una mediana mayor en el grupo con Conversión comparado con el grupo Sin Conversión, esta diferencia podemos apreciar visualmente en el Gráfico 4.



Posteriormente, para finalizar la parte descriptiva del análisis se presenta en la Tabla 3 los diferentes medicamentos que reportaron los pacientes estar consumiendo, evaluando su posible asociación con los grupos de Conversión.

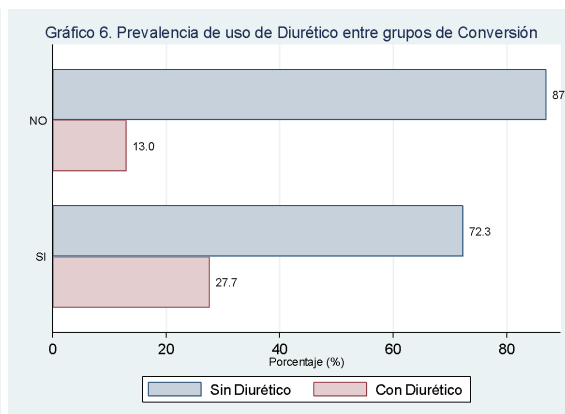
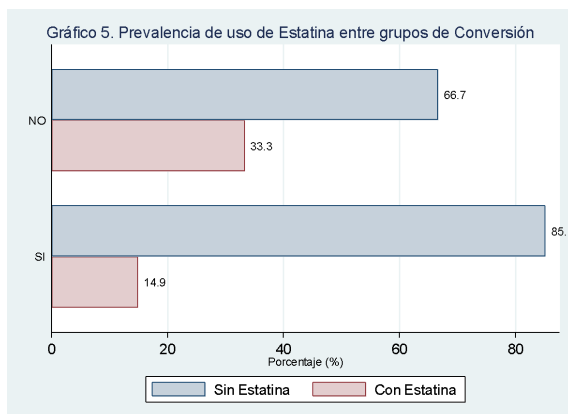
**Tabla 3. Medicamentos entre grupos de Conversión**

	<b>Categoría 100% (116)</b>		<b>Sin Conversión 59.48% (69)</b>	<b>Con Conversión 40.52% (47)</b>	<b>Valor p</b>
<b>METFOR</b>	<b>SÍ</b>	10.34% (12)	11.59% (8)	8.51% (4)	0.592 <sup>Δ</sup>
<b>ESTATINA</b>	<b>SÍ</b>	25.86% (30)	33.33% (23)	14.89% (7)	<b>0.026*</b>
<b>CAANT</b>	<b>SÍ</b>	12.93% (15)	8.70% (6)	19.15% (9)	0.100*
<b>IECAARA</b>	<b>SÍ</b>	48.28% (56)	52.17% (36)	42.55% (20)	0.309*
<b>BBLOC</b>	<b>SÍ</b>	12.93% (15)	15.94% (11)	8.51% (4)	0.242 <sup>Δ</sup>

<b>DIURETICO</b>	<b>SÍ</b>	18.97% (22)	13.04% (9)	27.66% (13)	<b>0.049*</b>
<b>FIBRATOS</b>	<b>SÍ</b>	11.21% (13)	8.70% (6)	14.89% (7)	0.299*
<b>ANTIAGREG</b>	<b>SÍ</b>	3.45% (4)	5.8% (4)	0%	0.146 <sup>Δ</sup>
<b>ACURICO</b>	<b>SÍ</b>	10.34% (12)	8.70% (6)	12.77% (6)	0.480*
<b>ANTIDEPRE</b>	<b>SÍ</b>	6.03% (7)	7.25% (5)	4.26% (2)	0.699 <sup>Δ</sup>
<b>ANSIOLITICO</b>	<b>SÍ</b>	6.90% (8)	7.25% (5)	6.38% (3)	1.000 <sup>Δ</sup>
<b>ANTIPARK</b>	<b>SÍ</b>	1.72% (2)	1.45% (1)	2.13% (1)	1.000 <sup>Δ</sup>
<b>BRONCODIL</b>	<b>SÍ</b>	4.31% (5)	2.90% (2)	6.38% (3)	0.394 <sup>Δ</sup>
<b>IBP</b>	<b>SÍ</b>	11.21% (13)	10.14% (7)	12.77% (6)	0.660*
<b>HORMONAL</b>	<b>SÍ</b>	3.45% (4)	2.90% (2)	4.26% (2)	1.000 <sup>Δ</sup>
<b>ANTICONVUL</b>	<b>SÍ</b>	1.72% (2)	2.90% (2)	0%	0.514 <sup>Δ</sup>
<b>OTROS</b>	<b>SÍ</b>	21.55% (25)	24.64% (17)	17.02% (8)	0.327*

\*Prueba de Independencia  $\chi^2$ . <sup>Δ</sup> Prueba exacta de Fischer

Podemos apreciar que, las variables que presentaron una asociación estadísticamente significativa fueron el consumo de estatinas y de diuréticos. Llama la atención que, para el uso de Estatinas encontramos una prevalencia mayor en el grupo Sin Conversión siendo de 33.33% (23) comparado con el grupo Con Conversión presentando un 14.89% (7). Caso contrario a la administración de Diurético el cual, presenta una mayor prevalencia de consumo en el grupo Con Conversión siendo de 27.66% (13) comparado con el grupo Sin Conversión presentando un 13.04% (9). Estas relaciones podemos apreciarlas de manera visual en los Gráficos 5 y 6.



El resto de las variables no presentaron una asociación estadísticamente significativa.

Finalmente, para darle cumplimiento al objetivo principal de nuestro proyecto de investigación, se eligieron las variables:

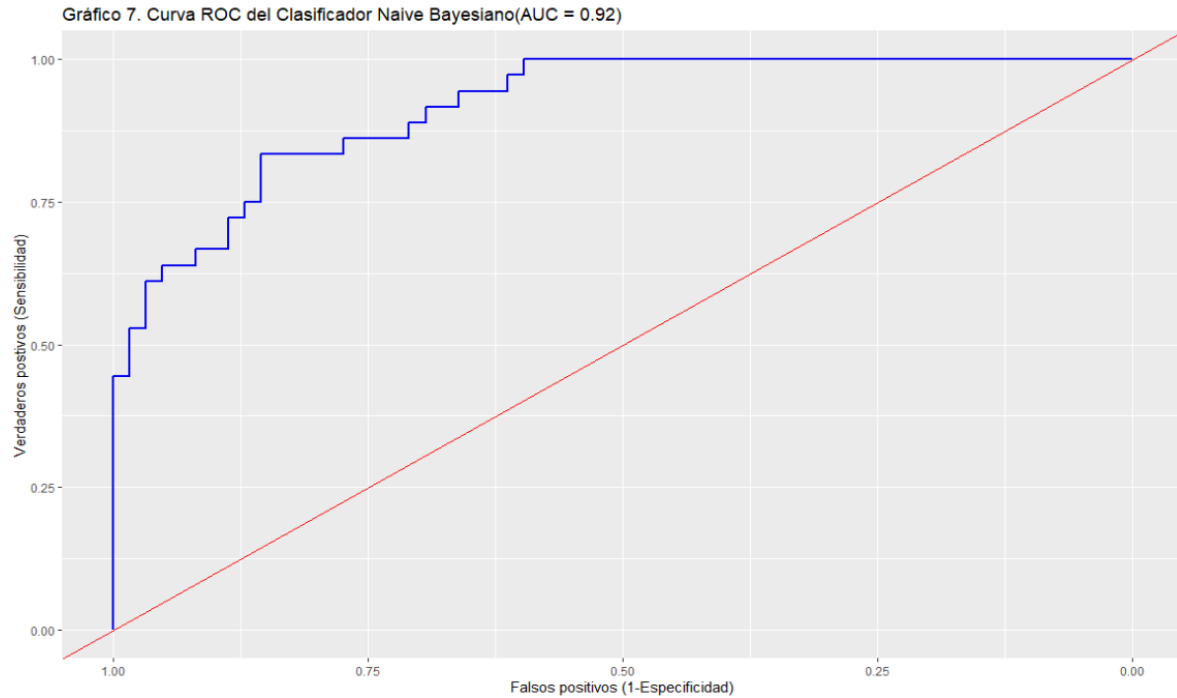
- Demográficas: Edad y Sexo, aunque estas variables no mostraron significancia estadística se decide incluir en el modelo contemplado una posible modificación del efecto sobre las demás variables dentro del modelo.
- Significativas: IMC, PA, Tabaquismo, Triglicéridos (TAG), Estatinas y Diuréticos.
- Efecto indirecto: Creatinina, HDL, LDL, Metformina, aunque estas variables no mostraron una diferencia o asociación estadísticamente significativas, consideramos que se encuentran relacionadas de manera directa o indirecta con el resto que sí mostró significancia estadística: ejemplo Diurético → Creatinina, Estatinas → HDL y LDL, Metformina → Glucosa.

El modelo de Clasificación Naive Bayesiano queda de la siguiente manera:

$$Pr(\text{Edad, Sexo, IMC, Tabaquismo, TAG, Estatina, HDL, LDL, Diurético, Cr, Metformina}) = Pr(\text{Conversión}) * Pr$$

### Rendimiento

Se obtuvo un área bajo la curva (AUC por sus siglas en ingles) de 0.9162 [0.864 – 0.9685], la cual podemos apreciar visualmente en el Gráfico 7.



Se presentan en la Tabla 4 los valores de los parámetros de rendimiento del modelo Clasificador Naive Bayesiano para la Conversión de prediabetes a DM2.

<b>Tabla 4. Parámetros de rendimiento</b>	
<b>Parámetro</b>	<b>Valor [IC-95]</b>
<b>Accuracy (Precisión)</b>	<b>0.8265 [0.7369 – 0.8956]</b>
<b>Sensibilidad</b>	<b>0.5833 [0.4088 – 0.7404]</b>
<b>Especificidad</b>	<b>0.9677 [0.8782 – 0.9943]</b>
<b>Valor Predictivo Positivo (VPP)</b>	<b>0.9130 [0.7048 – 0.9847]</b>
<b>Valor Predictivo Negativo (VPN)</b>	<b>0.8000 [0.6885 – 0.8801]</b>

Observamos que, el modelo presenta un mayor valor de Especificidad comparado con la Sensibilidad, pero presenta un equilibrio para los VPP y VPN siendo mayor el VPP.

Cabe mencionar que, la técnica del clasificador Naive Bayesiano pertenece al grupo de técnicas de aprendizaje de máquina supervisado por lo que, no se presentan valores p.

## DISCUSIÓN

En este estudio realizado en pacientes con diagnóstico de prediabetes, se buscaron las variables estadísticamente significativas para la conversión de prediabetes en diabetes tipo 2, basándonos en variables que fueron significativas en estudios que anteceden al nuestro, así como variables que forman parte del camino causal para el desarrollo de diabetes tipo 2.

En la parte descriptiva del análisis de variables para conversión a Diabetes tipo 2; en las variables clínicas, esperábamos una diferencia estadísticamente significativa en la función de distribución para las variables relacionadas con el síndrome metabólico ya que su fisiopatología obedece al mismo fenómeno de resistencia a la insulina en órganos insulino dependientes, encontrándose esta diferencia únicamente en el IMC y Perímetro abdominal (PA), con una mediana mayor en ambas variables, no encontrándose esta diferencia con la variable de presión arterial sistólica; además esto resalta la importancia de las variables antropométricas en la salud metabólica de los pacientes. Estos hallazgos también coinciden con estudios descriptivos previos en donde se encontraron esta diferencia en variables como el índice de masa corporal.

Resalta la ausencia de diferencia en la función de distribución de variables que histórica y científicamente han mostrado relación con el desarrollo de la diabetes tipo 2 como la carga hereditaria familiar, y el estilo de vida sedentario, así como otros factores recientemente estudiados como el consumo de alcohol y el estatus de tabaquismo.



Pese a no alcanzar significancia estadística se encontró una mayor prevalencia de pacientes con actividad física programada, con 30.43% en el grupo sin conversión a diabetes y 17.02% en el grupo con conversión a diabetes.

Al momento de analizar las variables bioquímicas encontramos al igual que con las variables clínicas una diferencia en la función de distribución en la variable de triglicéridos, siendo la dislipidemia como antes mencionado, parte del síndrome metabólico. En cuanto a los niveles de colesterol total, cabe resaltar que aunque no hubo una diferencia estadísticamente significativa en la función de distribución de la variable de colesterol total, con una mediana mayor en el grupo con conversión 212 (186- 245) en comparación con el grupo sin conversión con una mediana de 204 (189- 255), si se encontró esta diferencia con la variable de uso de estatinas, con una prevalencia mayor en el grupo Sin Conversión siendo de 33.33% (23) comparado con el grupo Con Conversión presentando un 14.89%, constatando, junto con las diferencias estadísticamente significativas encontradas con perímetro abdominal, índice de masa corporal y niveles de triglicéridos el rol fisiopatológico de la resistencia a la insulina que comparten estas variables junto con la prediabetes, englobadas en el síndrome metabólico, que a su vez confieren un riesgo incrementado para el desarrollo de diabetes tipo 2.

Las diferencias que se encontraron estadísticamente significativas en este estudio coinciden con aquellas encontradas por Shimizu S en su modelo predictivo, incluyéndose las mismas variables para la creación del estudio salvo el cambio en los niveles de la glucosa central, dado que nuestro modelo emplea únicamente una medición de las variables.

En cuanto al modelo clasificador Naive Bayesiano diseñado en nuestro estudio se obtuvo un área bajo la curva (AUC por sus siglas en inglés) de 0.9162 [0.864 – 0.9685],

no solo logrando cumplir con el objetivo del estudio, de obtención de un área bajo la curva de 0.8, superando a su modelo predecesor diseñado por Shimizu, el cual presentó un área bajo la curva de 0.816 y superando a otro modelo de aprendizaje automatizado diseñado por Cahn et. al, el cual obtuvo un área bajo la curva de 0.86 (0.860- 0.869)

Cabe resaltar que nuestro modelo presentó una menor sensibilidad que las obtenidas en estudios previos con 0.58 (0.4088 – 0.7404), en comparación con 0.8 encontrado en el modelo de Shimizu et. al, 0.86 encontrado en el modelo de Yokota et. al y 0.6 encontrado en el modelo de Cahn et. al.

Por otro lado, la especificidad encontrada fue superior a cada uno de los estudios previamente mencionados, obteniendo una especificidad de 0.9677 [0.8782 – 0.9943] con un valor predictivo positivo de 0.9130 [0.7048 – 0.9847], esto comparado con el modelo de Shimizu et. al. Que obtuvo una especificidad de 0.75, con valor predictivo positivo de 0.702, el estudio de Yokota con una especificidad de 0.61 o el estudio de Cahn con especificidad de 0.9.

Finalmente, con una precisión de 0.8265 [0.7369 – 0.8956], obtuvimos un a mejor discriminación entre grupos que el modelo antecesor en el estudio de Shimizu, et. al que obrubo una precisión de 0.78.

Con este estudio pudimos crear un modelo predictivo sencillo de utilizar en primer nivel de atención, y ya que no requerimos dos mediciones subsecuentes, podemos estimar el riesgo de progresión desde la primera visita al centro de atención de salud. Además de ser sencillo en su aplicación, nuestro modelo mostró un mejor rendimiento, con una mejor capacidad de discriminación etre grupos que estudios previos, teniendo una elevada especificidad, con un alto valor predictivo positivo, con el inconveniente de una menor sensibilidad y por ende un menor valor predictivo

negativo, siendo necesario su uso con otros modelos para su utilización como método de tamizaje.

Las fortalezas de este estudio son el desarrollo exitoso de un modelo predictivo utilizando un modelo clasificador naive Bayesiano, obteniendo una mejor especificidad, valor predictivo positivo y precisión para predecir conversión de prediabetes en diabetes tipo 2, y ya que se utilizó este tipo de modelo de aprendizaje automatizado supervisado, no fue necesario el contar con una gran base de datos para el desarrollo de un modelo preciso.

Las limitaciones del estudio son la poca heterogeneidad de la población al ser un estudio realizado en un solo centro, así como el diseño retrospectivo del mismo, pudiendo presentar sesgos al momento de recolectar datos del expediente clínico electrónico. Será necesaria una validación del modelo fuera del centro de estudio; además, por su esencia, aunque la asunción de independencia en el teorema de Bayes suele tener un buen desempeño, esto no siempre ocurre de este modo llevando a una clasificación incorrecta. Por otro lado, el modelo clasificador naive bayesiano asume que todas las variables contribuyen de igual manera al desenlace.

## CONCLUSIONES.

Entre todas las variables de estudio se encontraron como significativas muchos componentes del síndrome metabólico, además de ser variables fácilmente obtenidas en cualquier nivel de atención, además de ser variables con significancia estadística en población mexicana.. Además, se cumplió con creces el objetivo establecido de un área bajo la curva mayor a 0.8 para un modelo predictivo de conversión de prediabetes a diabetes; por tanto, es una herramienta útil para el médico y el paciente para establecer un riesgo determinado para la progresión de la enfermedad y así poder tener un seguimiento más estrecho para alcanzar las metas establecidas y evitar la progresión a diabetes con las complicaciones que conlleva.

## BIBLIOGRAFÍA

1. Rett K, Gottwald-Hostalek U. Understanding prediabetes: definition, prevalence, burden and treatment options for an emerging disease. *Curr Med Res Opin* 2019;35:1529–1534
2. Cefalu WT. “Prediabetes”: are there problems with this label? No, we need heightened awareness of this condition! *Diabetes Care* 2016;39:1472–1477
3. Blonde L, Umpierrez GE, Reddy SS, McGill JB, Berga SL, Bush M, et al. American Association of Clinical Endocrinology Clinical Practice Guideline: Developing a diabetes mellitus comprehensive care plan—2022 update. *Endocrine Practice*. 2022;28(10):923–1049. doi:10.1016/j.eprac.2022.08.002
4. ElSayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, et al. 2. classification and diagnosis of diabetes: <I>standards of care in diabetes—2023</I>. *Diabetes Care*. 2022;46(Supplement\_1). doi:10.2337/dc23-s002
5. Shamah-Levy T, Vielma-Orozco E, Heredia-Hernández O, Romero-Martínez M, Mojica-Cuevas J, Cuevas-Nasu L, Santaella-Castell JA, Rivera-Dommarco J. Encuesta Nacional de Salud y Nutrición 2018-19: Resultados Nacionales. Cuernavaca, México: Instituto Nacional de Salud Pública, 2020.
6. Middelbeek, Roeland J.w; and Martin J. Abrahamsori. “Diabetes, Prediabetes, and Glycemic Control in the United States: Challenges and Opportunities. “*Annals of Internal Medicine*, vol. 160, no. 8, 2014, p. 572, doi: 10.7326/m14-0539.
7. Coronado Malagón, Martín, et al. “Progresión de Prediabetes a Diabetes Mellitus Tipo 2 en Mexicanos. Evaluación En Una Cohorte.” *Gaceta Médica De México*, vol. 145, no. 4, Aug. 2009, pp 269- 272.

8. DeFronzo, R., Ferrannini, E., Groop, L. et al. Type 2 diabetes mellitus. *Nat Rev Dis Primers* 1, 15019 (2015). <https://doi.org/10.1038/hrdp.2015.19>
9. Tai, E. S; et al. "Lowering the Criterion for Impaired Fasting Glucose: Impact on Disease Prevalence and Associated Risk of Diabetes and Ischemic Heart Disease." *Diabetes Care*, vol 27, no. 7. 2004, pp. 1728- 1734; doi: 10.2337/diacare.27.7.1728.
10. Khaw, Kay- Tee, et al. "Association of Hemoglobin A1c with Cardiovascular Disease and Mortality in Adults: The European Prospective Investigation into Cancer in Norfolk." *Annals of Internal Medicine*, vol. 141, no. 6, 2004, p. 413; doi: 10.7326/0003-4819-141-6-200409210-00006.
11. Brannick B, Wynn A, Dagogo-Jack S. Prediabetes as a toxic environment for the initiation of microvascular and macrovascular complications. *Experimental Biology and Medicine*. 2016;241(12):1323-1331.
12. Lawal Y, Bello F, Kaoje Y. Prediabetes Deserves More Attention: A Review. *Clinical Diabetes*. 2020;38(4):328-338.
13. Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. *Nature Reviews Microbiology*. 2020;19(1):55-71.
14. Phatak SR, Saboo B, Dwivedi S, Zinzuwadia P, Panchal D, Ganguli A, et al. Sweetening sixteen: Beyond the ominous octet. *J Diabetol* 2021;12:1-9.
15. Selvin E, Feinleib M, Zhang L, Rohrmann S, Rifai N, Nelson WG, et al. Androgens and diabetes in men: Results from the Third National Health and Nutrition Examination Survey (NHANES III) *Diabetes Care*. 2007;30:234-8

16. Donath, M., Shoelson, S. Type 2 diabetes as an inflammatory disease. *Nat Rev Immunol* **11**, 98–107 (2011). <https://doi.org/10.1038/nri2925>
17. Balkau, B, et al. “Risk Factors for Early Death in Non- Insulin Dependent Diabetes and Men with Known Glucose Tolerance Status.” *Bmj*, vol. 307, no. 6899, 1993, pp. 295- 299; doi:10.1136/bmj.307.6899.295.
18. Paprott, Rebecca, et al. “Determinants of Change in Glycemic Status in Individuals with Prediabetes: Results from a Nationwide Cohort Study in Germany.”. *Journal of Diabetes Research*, vol. 2018, 2018, pp. 1- 12; doi:10.1155/2018/5703652.
19. Bello-Chavolla O, Almeda-Valdes P, Gomez-Velasco D, Viveros-Ruiz T, Cruz-Bautista I, Romo-Romo A et al. METS-IR, a novel score to evaluate insulin sensitivity, is predictive of visceral adiposity and incident type 2 diabetes. *European Journal of Endocrinology*. 2018;178(5):533-544.
20. Fahed G, Aoun L, Bou Zerdan M, Allam S, Bou Zerdan M, Bouferraa Y, et al. Metabolic syndrome: Updates on pathophysiology and management in 2021. *International Journal of Molecular Sciences*. 2022;23(2):786. doi:10.3390/ijms23020786
21. Yokota, N., et al. “Predictive Models for Conversion of Prediabetes to Diabetes.” *Journal of Diabetes and Its Complications*, vol. 31, no. 8, 2017, pp. 1266- 1271., doi:10.1016/j.jdiacomp.2017.01.005.
22. Shimizu, S, Diseño de un modelo pronóstico de la conversión de prediabetes a diabetes en prediabéticos del hospital central sur de alta especialidad, Tesis de posgrado, Ciudad de México, Universidad Nacional Autónoma de México, 2021.

23. Nichols, Gregory. "A Simple Model for Predicting Two- Year Risk of Diabetes Development in Individuals with Prediabetes. " *The Permanente Journal*, vol. 22, no. 17, 2018, doi:10.7812/tpp/17-050.
24. Cahn, Avivit, et al. "Predictino of Progression from Pre- Diabetes to Diabetes: Development and Validation of a Machine Learning Model." *Diabetes/ Metabolism Research and Reviews*, vol. 36, no. 2, 2020, doi:10.1002/dmrr.3252.
25. Pichler M, Hartig F. Machine learning and deep learning—a review for ecologists. *Methods in Ecology and Evolution*. 2023;14(4):994–1016. doi:10.1111/2041-210x.14061
26. Lina L. Dhande and Dr. Prof. Girish K. Patnaik, "Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier", *IJETTCS*, Volume 3, Issue 4 July-August 2014, ISSN 2278-6856.
27. Rooney MR, Rawlings AM, Pankow JS, et al. Risk of Progression to Diabetes Among Older Adults with Prediabetes. *JAMA Intern Med*. 2021;181(4):511–519. doi:10.1001/jamainternmed.2020.8774