



Estimación de parámetros de desempeño de algoritmos para diagnóstico de Dengue

Karen Daniela López Villegas
Julieth Natalia Salazar Vargas

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia

2022

Estimación de parámetros de desempeño de algoritmos para diagnóstico de Dengue

Karen Daniela López Villegas
Julieth Natalia Salazar Vargas

Trabajo de grado presentado como requisito parcial para optar al título de:
Estadístico(a)

Director:
PhD. José Rafael Tovar Cuevas
Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2022

Dedicatoria o un lema

A mi familia, quienes me han apoyado para llegar a estas instancias de mi estudio y han estado presente en cada obstáculo superado.

Julieth Natalia Salazar Vargas

A Dios por regalarme el privilegio de la vida, a mi familia por su motivación constante, apoyo y amor, y a cada una de las personas que de una u otra forma han influenciado en mi vida.

Karen Daniela López Villegas

Agradecimientos

Primeramente agradecemos a Dios, por regalarnos salud y vida para sacar adelante todo lo que nos hemos propuesto. Gracias a nuestras familias por todo el amor, apoyo y comprensión que nos han brindado a lo largo de nuestras vidas.

A la universidad, y en especial a la escuela de estadística que contribuyó de manera importante para forjarnos como excelentes profesionales y personas.

A nuestros compañeros de carrera, con quienes estrechamos lazos de amistad durante el proceso académico y a quienes les deseamos muchos éxitos, salud y bendiciones en su vida.

A nuestro director de trabajo de grado, PhD José Rafael Tovar por impartirnos las directrices necesarias para el desarrollo de este trabajo de grado.

Resumen

La detección temprana y oportuna de dengue en los pacientes resulta fundamental para una acción eficaz frente al inicio del tratamiento, lo que podría contribuir a la reducción de la mortalidad por esta enfermedad. Este trabajo tiene como objetivo proponer y evaluar la capacidad clasificadora de un grupo de algoritmos para diagnóstico clínico contruidos a partir de signos y síntomas después de agregar variables asociadas al resultado de un hemograma. El archivo de datos que se utilizó para el desarrollo de este trabajo corresponde a un estudio de cohorte para validación de pruebas de clasificación o diagnóstico de Dengue en el marco del proyecto de investigación P4-RedAedes, el cual fue realizado entre agosto del 2017 y julio del 2018 en las ciudades de Cali, Piedacuesta y Yopal. La construcción de estos algoritmos se hizo con base en métodos bayesianos y se comparó su rendimiento mediante la estimación de los parámetros de desempeño desde los enfoques clásicos, bayesiano y considerando remuestreo, donde finalmente se encontró que agregar las variables del hemograma a los algoritmos que inicialmente solo consideraban signos y síntomas contribuyó considerablemente a mejorar su capacidad clasificatoria. Además, se encontró que al hacer la aplicación de las pruebas en paralelo y tener un único resultado clasificador las sensibilidades obtenidas fueron superiores al 70 %.

Palabras clave: dengue/diagnóstico, algoritmos, teorema de Bayes, estimación, dependencia condicional

Contenido

Resumen	vi
Lista de Figuras	x
Lista de Tablas	1
1. Introducción	3
2. Planteamiento del problema, justificación y objetivos del proyecto	5
2.1. Planteamiento del Problema	5
2.1.1. Pregunta de Investigación	6
2.2. Justificación	6
2.3. Objetivos	7
2.3.1. Objetivo General	7
2.3.2. Objetivos Específicos	7
3. Antecedentes y Marco teórico	8
3.1. Antecedentes	8
3.2. Marco Teórico Clínico	12
3.2.1. Dengue	12
3.2.2. Diagnóstico diferencial del Dengue	12
3.2.3. Hemograma	13
3.2.4. Algoritmos de Clasificación	14
3.2.5. Gold Standard para Dengue	15
3.2.6. Aplicación de pruebas diagnosticas en paralelo	15
3.3. Marco Teórico Estadístico	15
3.3.1. Validación de una prueba diagnostica	15
3.3.2. Medidas de Desempeño	16
3.3.3. Falta de Identificabilidad en el Modelo	18
3.3.4. Métodos de remuestreo (Bootstrap)	19
3.3.5. Inferencia Bayesiana	19
3.3.6. Receiver operating characteristic curve (Curva ROC)	20
4. Metodología	23
4.1. Datos:	23

4.2.	Identificación de variables asociadas al Dengue	23
4.2.1.	Signos y síntomas	23
4.2.2.	Hemograma	25
4.3.	Definición de nuevos Prototipos Clasificadores	26
4.4.	Modelo estadístico	28
4.4.1.	Experimento Estadístico	28
4.4.2.	Eventos	28
4.4.3.	Variables aleatorias	28
4.4.4.	Calculo de probabilidades predictivas de Bayes	28
4.4.5.	Puntos de Corte para la probabilidad $P(D = 1 B_j)$	30
4.4.6.	Cálculo de los valores predictivos	32
4.5.	Diseño de Cohorte (un test diagnostico)	32
4.5.1.	Especificación del Modelo	32
4.5.2.	Estimación de los parámetros de desempeño	34
4.6.	Dos pruebas para diagnóstico bajo diseño de cohorte	38
4.6.1.	Especificación del Modelo	38
4.6.2.	Escenario 1: Estimación de parámetros de desempeño asumiendo independencia condicional entre las dos pruebas	42
4.6.3.	Escenario 2: Estimación de parámetros de desempeño asumiendo dependencia condicional entre las dos pruebas para diagnostico en evaluación	47
4.7.	Estimación de los parámetros de desempeño después de obtener un único procedimiento clasificador	51
4.7.1.	Especificación del Modelo	51
4.7.2.	Estimación	52
5.	Resultados	53
5.1.	Resultados Prototipo I	53
5.2.	Resultados Prototipo II	54
5.3.	Resultados Prototipo III	55
5.4.	Resultados Prototipo IV	56
5.5.	Resultados Prototipo V	56
5.6.	Resultados Prototipo VI	57
5.7.	Comparación entre los prototipos clasificadores	58
5.8.	Estimación de parámetros de desempeño asumiendo independencia condicional entre las dos pruebas	60
5.8.1.	Estimación parámetros de desempeño	60
5.9.	Estimación de parámetros de desempeño asumiendo dependencia condicional entre las dos pruebas	61
5.9.1.	Estimación Bayesiana	61

5.10. Estimación de los parámetros de desempeño después de obtener un único procedimiento clasificador	63
5.10.1. Estimación de parámetros de desempeño	63
6. Conclusiones y recomendaciones	65
A. Anexos:	68
Bibliografía	69

Lista de Figuras

5-1. Valores Predictivos	58
5-2. Valores Predictivos para cada Algoritmo de Clasificación bajo diferentes prevelencias	59

Lista de Tablas

3-1. Tabla de clasificación modelo de cohorte con un test diagnóstico	16
3-2. Arreglo de datos en un diseño de cohorte para validación de dos pruebas para diagnóstico	17
4-1. Signos y síntomas identificados	24
4-2. Categorías de signos y síntomas conformadas	25
4-3. Tabla de clasificación modelo de cohorte con un test diagnóstico en términos de variables aleatorias	32
4-4. Valores elicitados e hiperparámetros	38
4-5. Arreglo de datos en un diseño de cohorte para validación de dos pruebas para diagnóstico.	39
5-1. Medidas de desempeño estimadas para el prototipo I	54
5-2. Medidas de desempeño estimadas para el prototipo II	54
5-3. Medidas de desempeño estimadas para el prototipo III	55
5-4. Medidas de desempeño estimadas para el prototipo IV	56
5-5. Medidas de desempeño estimadas para el prototipo V	57
5-6. Medidas de desempeño estimadas para el prototipo VI	58
5-7. Valores de prevalencia donde se maximizan los Valores Predictivos para cada Algoritmo	60
5-8. Medidas de desempeño estimadas asumiendo independencia entre los resultados de las dos pruebas	61
5-9. Estimación de los parámetros de dependencia entre los algoritmos de clasificación (covarianzas)	62
5-10. Estimación bayesiana - Medidas de desempeño estimadas asumiendo dependencia entre los resultados de las dos pruebas	62
5-11. Medidas de desempeño estimadas asumiendo un único proceso clasificador .	64
A-1. Valores para elicitación	68
A-2. Puntos de corte encontrados por Méndez (2017) para cada prototipo	68

Declaración

Nos permitimos afirmar que hemos realizado el presente Trabajo de Grado de manera autónoma y con la única ayuda de los medios permitidos y no diferentes a los mencionados en el propio trabajo. Todos los pasajes que se han tomado de manera textual o figurativa de textos publicados y no publicados, los hemos reconocido. Ninguna parte del presente trabajo se ha empleado en ningún otro tipo de Tesis o Trabajo de Grado.

Igualmente declaramos que los datos utilizados en este trabajo están protegidos por las correspondientes cláusulas de confidencialidad.

Santiago de Cali, 18.03.2022

(Karen Daniela López Villegas)

(Julieth Natalia Salazar Vargas)

1. Introducción

El dengue es una enfermedad endémica en varios países del mundo. A nivel nacional, desde 1978 hasta el 2010, se ha registrado oficialmente un acumulado de 1'020.637 casos de dengue, lo que representa un promedio anual de 30.928 Padilla et al. (2012). Por lo anterior, el diagnóstico del dengue de forma temprana (a partir de los signos y síntomas de alarma) permiten que el pronóstico por parte del personal médico sea oportuno para el inicio del tratamiento.

Méndez (2017) desarrolló una propuesta metodológica para construir reglas de clasificación para dengue desde el enfoque bayesiano, los cuales corresponden al principal insumo en el desarrollo del presente trabajo, pues el objetivo principal radica en actualizar y mejorar la capacidad de dichas reglas de clasificación mediante la agregación de nuevas variables relacionadas con el resultado de un hemograma que no fueron consideradas en la construcción inicial de los algoritmos. Además, siguiendo la misma metodología por Méndez (2017) se propusieron dos nuevos algoritmos que solo consideraban las variables asociadas al hemograma.

En este trabajo se utilizará la base de datos del proyecto P4-RedAedes, la cual consta de 1039 registros de sujetos febriles entre 0 y 80 años de edad, y que se encontraban en zonas endémicas de Colombia (Cali, Piedecuesta y Yopal) entre agosto 2017 y julio 2018. La base contiene variables clínicas relacionadas con dengue, resultado del hemograma y un algoritmo de pruebas diagnósticas de laboratorio para dengue como prueba de referencia (RT-PCR, NS1 ELISA, IgM), según el diagnóstico confirmatorio se identificaron 25 dengue confirmado, no dengue 307, dengue probable 514 e indeterminado 193. El archivo de datos final, con el cual se desarrollo este trabajo, solo considero a pacientes con dengue positivo o negativo, y que además tuvieran información completa sobre todas las variables consideradas en este estudio, estos son 324 individuos que se dividen entre 24 dengue positivo y 300 dengue negativo.

Al tener una muestra (menor a 30) de sujetos con dengue y con el propósito de encontrar estimaciones más precisas de los parámetros de desempeño para evaluar el rendimiento de cada algoritmo clasificador, se propuso ampliar el archivo de datos con que se contaba inicialmente, ya que bajo un proceso de simulación se podría evaluar la convergencia de estos resultados y así, llegar a conclusiones más certeras. La estimación de los parámetros de

desempeño se realizó bajo los supuestos de dependencia e independencia condicional de las pruebas desde el enfoque clásico y bayesiano, por ultimo, con el objetivo de ser más precisos en la clasificación del dengue se propuso obtener un único procedimiento clasificador a partir de la aplicación simultanea de los prototipos creados por Méndez (2017) y los algoritmos creados con base en las variables del hemograma planteados en este trabajo.

Se encontró que la metodología de simulación propuesta en este trabajo es una excelente herramienta en los casos donde no se cuenta con un archivo de datos diferente al utilizado en la construcción de los algoritmos y por tanto se busca ampliar en conjunto de datos existente para analizar la convergencia en las estimaciones, además que las variables asociadas al hemograma mejoraron la capacidad clasificatoria de los algoritmos que había propuesto Méndez (2017) inicialmente. Por último, al usar un único procedimiento clasificador después de haber hecho la aplicación simultanea de los algoritmos se obtuvo una muy buena clasificación de los pacientes que están verdaderamente enfermos (sensibilidades altas).

2. Planteamiento del problema, justificación y objetivos del proyecto

2.1. Planteamiento del Problema

El dengue es una infección vírica transmitida por la picadura de mosquitos hembras hematófago (se alimentan de sangre) principalmente de la especie *Aedes aegypti*, conocido también como zancudo patas blancas. Es una enfermedad de rápida propagación, y se presenta en los climas tropicales y subtropicales de todo el planeta, sobre todo en las zonas urbanas y semiurbanas World Health Organization (2020). Según la Organización Mundial de la Salud (OMS) el Dengue, tiene un alto impacto epidemiológico, social y económico, por lo que, es considerada como un problema creciente en la salud pública a nivel mundial.

Uno de los principales problemas del dengue es el diagnóstico erróneo, puesto que por sus síntomas puede llegarse a confundir con otras enfermedades como lo son el virus Zika y Chikungunya, adicionalmente, una gran mayoría de los casos son asintomáticos, por lo que el número real de casos está insuficientemente notificado y muchos están mal clasificados Villamil Gómez et al. (2016). Las pruebas de laboratorio existentes tienen un desempeño variable en la práctica clínica, pues su sensibilidad fluctúa entre 45 y 98 %, y su especificidad, entre 4 y 98 %, lo cual se debe, en parte, a la diversidad de contextos en los que se utilizan Caicedo et al. (2019), además que resultan ser costosas, por lo que se dificulta su disponibilidad en centros de atención, y su manejo requieren de un personal calificado, por esta razón, se han desarrollado pruebas rápidas que permiten el diagnóstico de esta enfermedad .

Méndez (2017) desarrolló algoritmos para el diagnóstico de dengue, estos algoritmos, fueron contruidos con base en métodos Bayesianos, a través de los cuales lograron incluir información relacionada con los signos y síntomas que reporta el paciente en el momento de la consulta clínica, además, algunos consideran variables como el conteo de plaquetas y leucocitos. Como continuación a ese estudio se pretende agregar a esos algoritmos variables asociadas al resultado del hemograma que el médico ordena en la primera consulta con el paciente, de esta manera se plantearan nuevas reglas de clasificación y se podrá determinar si el agregar estas variables aumenta la capacidad clasificatoria de los algoritmos, para ésto se tendrán que validar ese nuevo conjunto de algoritmos, estimando sus parámetros de

desempeño (sensibilidad y especificidad).

La detección temprana y oportuna del dengue en los pacientes resulta fundamental para una acción eficaz frente al inicio del tratamiento, lo que podría contribuir a la reducción de la mortalidad por esta enfermedad. Así, las pruebas diagnósticas desarrolladas a partir de metodologías estadísticas que aporten a la exactitud del diagnóstico, frente a un escenario donde se cuentan con pocos registros confirmados de personas enfermas resulta una herramienta que amplía las posibilidades de investigación frente a este problema clínico.

Hay periodos de tiempo donde se cuentan con pocos casos, lo cual implica un reto estimar la sensibilidad y especificidad de los algoritmos, en este caso se cuenta con un conjunto de datos donde hay 24 casos confirmados de dengue en más de 300 registros, es importante aclarar que estos registros fueron recogidos de manera rigurosa y ordenada por una médica especialista en dengue durante un proyecto de investigación. Por esta razón, resulta indispensable diseñar una metodología o estrategia estadística que permita abordar el problema de la muestra pequeña con el fin de obtener la mayor información posible, así, se utilizarán técnicas estadísticas desde el enfoque bayesiano tomando información previa que se tiene de estudios anteriores.

2.1.1. Pregunta de Investigación

- **Clínica:** ¿Cuál algoritmo (combinación de características clínicas) para el diagnóstico del dengue presenta un mejor desempeño en la clasificación de sujetos como caso de dengue?
- **Estadística:** ¿Las reglas clasificadoras utilizando probabilidades predictivas de Bayes para procesar la información de los pacientes tienen un buen desempeño como procedimiento de diagnóstico clínico?

2.2. Justificación

El dengue es una enfermedad endémica en varios países del mundo. A nivel nacional, desde 1978 hasta el 2010, se ha registrado oficialmente un acumulado de 1'020.637 casos de dengue, lo que representa un promedio anual de 30.928 Padilla et al. (2012). Para el año 2018, el Instituto Nacional de Salud reportó una incidencia de 172,9 casos por 100.000 habitantes Instituto Nacional de Salud (2019). Datos del Ministerio de Salud indican que, para lo corrido del año 2021, con corte a 11 de julio, se han registrado 18.262 casos de dengue, de estos 9.593 (52,5 %) corresponden a casos sin signos de alarma, 8.397 (46 %) de dengue con signos de alarma y 272 (1,5 %) corresponden a formas graves Minsalud (2021).

Por lo anterior, el diagnóstico del dengue de forma temprana (a partir de los signos y síntomas de alarma) permiten que el pronóstico por parte del personal médico sea oportuno para el inicio del tratamiento, por esta razón, la detección de esta enfermedad ha sido ampliamente discutida en investigaciones de salud, dado que la correcta clasificación de los individuos puede darse como una tarea difícil, ya que, enfermedades como el Zika y Chikunguya podrían arrojar cuadros clínicos similares, desencadenando un manejo inadecuado de los casos de dengue y contribuyendo así, el aumento de la mortalidad por éste.

Proponer alternativas estadísticas que permitan fortalecer herramientas metodológicas para la toma de decisiones en el ámbito clínico para la confirmación del diagnóstico de dengue ofrece una mirada integral de la contribución que realiza cada una de las variables consideradas, así brindar una alternativa para la toma de decisión confiable en el ámbito médico brindando certeza al inferir resultados, en función de la evidencia disponible acerca de los individuos con presencia de dengue.

La metodología propuesta en este trabajo de grado apunta a mitigar los efectos que pueden causar el tener un tamaño de muestra reducido, lo cual es común cuando se estudia la exposición de una persona a una enfermedad, pues contar con muestras grandes es económica y humanamente difícil. Además, se ejemplifica como evaluar las pruebas de diagnóstico bajo supuestos de dependencia e independencia condicional, de esta manera, contribuir mediante aportes estadísticos, para la validación de algoritmos clínicos del diagnóstico del dengue se encuentra relacionado con la precisión de clasificación, pues las pruebas existentes tienen un desempeño variable según el contexto en que se realizan Zhang et al. (2015).

2.3. Objetivos

2.3.1. Objetivo General

Proponer y evaluar el desempeño de un grupo de algoritmos construidos a partir de signos y síntomas después de agregar variables asociadas al resultado de un hemograma.

2.3.2. Objetivos Específicos

- 1 Estimar los parámetros de desempeño de los nuevos algoritmos construidos mediante métodos del paradigma clásico y bayesiano.
- 2 Estimar los parámetros de desempeño de los algoritmos en un modelo estadístico que asuma dependencia entre dos pruebas.
- 3 Evaluar el desempeño de los 4 algoritmos de Méndez (2017) y el resultado de las pruebas de laboratorio asumiendo que son aplicados en pares con diseño en paralelo.

3. Antecedentes y Marco teórico

3.1. Antecedentes

Pan-ngum et al. (2013) llevaron a cabo un estudio para estimar la precisión de tres pruebas de diagnóstico rápido de dengue (prueba de casete Panbio NS1, IgM e IgG) utilizando modelos bayesianos de clase latente (LCM). Los datos del estudio corresponden a una cohorte de 549 pacientes sospechosos de infección por Dengue registrados en el Hospital de Enseñanza del Norte de Colombo, Ragama, Sri Lanka (Tailandia). Los autores optaron por aplicar un ensayo de referencia (combinación de ELISA que captura tanto información de IgM como IgG) que se asumió como perfecto debido a que no se cuenta con un “Gold Standar”, y mediante LCM se obtuvieron los parámetros de sensibilidad y especificidad, donde posteriormente se contrastaron con los resultados de las pruebas rápidas.

La técnica usada por los autores corresponde a los modelos bayesianos de clase latente, pues les permitía estimar la prevalencia (Probabilidad de que un paciente con sospecha de dengue esté realmente enfermo) y la precisión de cada prueba. Para esto, debían usar el modelo de mejor ajuste, el cual se seleccionó con respecto a las correlaciones, donde se consideraron tanto pruebas existentes como evidencia externa. Se utilizaron los criterios AIC y DIC para evaluar la bondad de ajuste de los modelos y compararlos entre sí, de esta forma obtener el modelo con el cual se determinaría la precisión de las pruebas y sus combinaciones.

Es importante mencionar, que en el desarrollo del estudio se encontraron algunos limitantes, puesto que al usar LCM bayesianos básicos para estimar la sensibilidad y especificidad de las pruebas se deja por fuera información referente a la duración de síntomas y el momento que se tomó la muestra; efectos que pudieron evaluarse utilizando LCM bayesianos avanzados Bernatsky et al. (2011).

Finalmente, el estudio concluyó que, una vez demostrado que el ensayo de referencia representa un “Gold Standar” defectuoso contra el cual comparar pruebas de diagnóstico rápido de dengue, los LCM bayesianos son útiles, ya que son imparciales frente a tener o no un prueba de referencia perfecta.

Tovar Cuevas (2015) presenta el proceso de estimación de los parámetros de desempeño diagnóstico (sensibilidad y especificidad) de la colangiopancreatografía magnética (CRM) temprana en pacientes con pancreatitis aguda de origen Biliar (PAB) para la detección de coledocolitiasis usando como prueba de referencia la Colangiopancreatografía endoscópica retrógrada (CPRE). El autor propone realizar la estimación de los parámetros mediante métodos bayesianos (tanto para caso discreto como continuo) y frecuentistas, con el fin de comparar cual de las dos metodologías es más precisa en un caso como el que se está tratando. Para el desarrollo del estudio, se tomaron 154 registros de historias clínicas que correspondían a individuos con PAB leve y fueron expuestos al CRM, los cuales fueron trabajados anteriormente por Mogollón et al. (2014), sin embargo, se encontró que no todos los sujetos contaban con los mismos resultados clínicos, por ende, se agruparon obteniendo un total de 3 grupos, cada uno con alrededor de 80 registros, y que finalmente representarían el tamaño de muestra para el análisis de datos.

Por último, y según los resultados obtenidos, las estimaciones bayesianas para los parámetros de desempeño obtuvieron un grado de precisión más alto, pues la información que se utiliza no solo proviene de los datos sino de información externa como lo puede ser estudios previos o un experto en el tema, por ende sus resultados podrían ser más confiables, esto frente a las estimaciones obtenidas por métodos clásicos.

Tovar & Achcar (2011) Plantean el problema de evaluar dos o más test de diagnóstico cuando estos son aplicados a un mismo individuo, ya que se considera que los resultados presentan algún tipo de dependencia que puede ser ignorado en la etapa de análisis debido a la presentación binaria de los datos.

Por ello, en este estudio se comparó el comportamiento de dos parámetros de dependencia presentes en funciones de cópula (Farlie Gumbel Morgerstern (FGM) y copula Gumbel) con el de la covarianza binaria y dos índices creados por Böhning & Patilea (2008) para medir dependencia entre pruebas diagnósticas de respuesta dicótoma. El objetivo principal es evaluar la relación existente entre los cinco parámetros de dependencia, donde tres de ellos (covarianza e índices de Böhning) se construyen para estudiar la dependencia entre variables binarias y los otros dos (parámetros de cópula) se desarrollan para modelar la dependencia entre variables continuas.

Una de las menciones importantes que se realiza en el estudio es que cuando los resultados de los test están en una escala continua, no necesariamente la dependencia entre estos se mide para dos variables continuas, si no que al fijar un punto de corte esta dependencia se puede estimar para variables binarias.

En presencia de sesgo de verificación los índices creados por Böhning & Patilea (2008) muestran un buen desempeño. Las dos funciones de cópula usadas modelan dependencias débiles no lineales y los índices de Böhning & Patilea (2008) no pudieron modelarlas, sin embargo, hay muchas otras familias de cópulas cuya relación con los índices de Böhning & Patilea (2008) podría estudiarse.

Méndez (2017) propuso una metodología (abordada desde el enfoque bayesiano) para construir reglas de clasificación para el diagnóstico de Dengue, donde desarrolló cuatro algoritmos. Cada prototipo de algoritmo clasificador cuenta con cualidades propias, las cuales provienen de la experiencia aportada por los expertos en la selección de los signos y síntomas que debían hacer parte de cada uno de ellos, así lo afirma el autor. La base de datos utilizada en el estudio correspondía a 987 registros de individuos, donde se capturo la información de pacientes provenientes de Palmira, Cali, Neiva y Bucaramanga.

Para la validación de los prototipos el autor estimó los parámetros de desempeño (sensibilidad y especificidad), los cuales fueron estimados empleando dos métodos distintos. El primero consiste en utilizar la misma base de datos, sin embargo, advierte que esto podría desencadenar una sobreestimación, por lo tanto considera pertinente un segundo método (Bootstrap) donde se extraen sub-muestras que componen la base datos original, posterior a esto se construyeron curvas ROC no paramétricas, donde el cálculo del área bajo la curva indica la precisión diagnóstica de cada algoritmo.

Una de las limitaciones que tuvo el estudio, hace referencia a la falta de información externa por parte de un experto en el área, ya que la estimación de los parámetros de desempeño podrían haber arrojado un mejor resultado. Por último, se resalta que el uso de metodologías bayesianas como herramienta para el desarrollo de algoritmos diagnóstico, permite que estos se puedan enriquecer con más información técnica (pruebas adicionales) o empírica (conocimiento a priori).

Caicedo-Borrero et al. (2020) Desarrollaron y validaron prospectivamente algoritmos clínicos de diagnóstico de dengue en Colombia. El estudio se realizó en dos fases, la primera consistió en desarrollar los algoritmos con una base de datos de 1.130 casos de dengue y 918 casos de no dengue, teniendo en cuenta para su construcción la opinión de expertos sobre el tema y la revisión de literatura con el fin de elegir los signos y síntomas a más relevantes, adicionalmente se incluyeron variables del hemograma como: recuento de leucocitos y plaquetas. Posteriormente, en la segunda fase, se validó prospectivamente a aquellos algoritmos que tuvieron sensibilidad $\geq 70\%$, para esto, se realizó un ensayo cuasi experimental de un solo grupo.

El análisis propuesto por los autores consiste en un diseño bayesiano adaptativo, donde se incluyeron análisis intermedios (en total 5) con el fin de evaluar la sensibilidad de cada algoritmo clínico, de esta manera, si el resultado era ¡70 % el algoritmo debía ser ajustado por medio de un cambio de variables y/o la modificación de los valores de corte, donde idealmente estos últimos debían alcanzar una sensibilidad del 95 % y una especificidad entre el 70 % y 95 %. Es preciso definir qué, la validación prospectiva se hizo en tres zonas endémicas de dengue en Colombia y el tamaño de muestra estimado para los análisis intermedios se calculó a partir de **modeling Bayesian effective sample size scenarios**.

Finalmente, los resultados obtenidos del estudio indicaron que: primero, la inclusión de parámetros del hemograma mejoró la especificidad de los algoritmos sin afectar la sensibilidad; segundo, al considerar sólo los casos de dengue y no dengue se identificó un algoritmo con una sensibilidad del 65.4 % y una especificidad de 40.1 % y, por último, todos los algoritmos tenían razones de verosimilitud cercanas a uno, por lo tanto, no eran útiles para confirmar dengue en zonas endémicas pues era probable encontrar una mala clasificación debido a las infecciones secundarias difíciles de discriminar en estas áreas.

Quiñones & Suarez (2019) realizaron un estudio de simulación para evaluar la validez de criterio de los algoritmos propuestos por Méndez (2017), para esto se construyeron distintos archivos de datos, los cuales se generaron a partir de la simulación de variables aleatorias binarias correlacionadas. Esta propuesta de estudio se planteó bajo diferentes escenarios de prevalencia de dengue en campo y diferentes prevalencias en la presencia de conjunto de signos y síntomas en la población de interés.

Para la validación de los algoritmos, utilizaron remuestreo Bootstrap y calcularon los parámetros de desempeño (sensibilidad y especificidad) de cada uno en los distintos escenarios de simulación propuestos, seguidamente se realizaron tablas de clasificación teniendo en cuenta los puntos de corte establecidos por Méndez (2017).

Finalmente, según los resultados obtenidos por las autoras, se encontró que los prototipos III y IV presentaron un mejor desempeño en la clasificación de los pacientes como caso dengue, además, se concluyó que construir archivos de datos mediante simulación fue una herramienta importante para la validación de los prototipos clasificadores, puesto que permitió analizar el comportamiento de cada uno, así mismo como la convergencia de sus medidas de desempeño y sus probabilidades predictivas en distintos panoramas de ocurrencia de la enfermedad.

3.2. Marco Teórico Clínico

3.2.1. Dengue

El virus del Dengue (DENV) es un arbovirus del género *Flavivirus*, familia *Flaviviridae*, que tiene 4 serotipos (DENV1, DENV2, DENV3 y DENV4). Los cuatro serotipos circulan periódicamente en áreas endémicas e hiperendémicas y, sin distinción alguna, todos causan la enfermedad conocida como dengue Velandia & Castellanos (2011). El DENV se transmite por la picadura de mosquitos hembras hematófago (se alimentan de sangre) principalmente de la especie *Aedes aegypti*. El mosquito se infecta al picar a una persona que tenga el virus, el cual se reproduce y se localiza en las glándulas salivares, para luego transmitir el virus al picar a otra persona sana Guzmán et al. (2006).

Los pacientes con dengue consultan a instituciones de salud motivadas por síntomas de fiebre alta, dolor de cabeza, dolor retrobulbar, artralgia, escalofríos y, en algunos casos, erupción cutánea. Estos signos y síntomas definen una probable presentación clínica de dengue Castellanos & Coronel-Ruiz (2014). El rápido diagnóstico del Dengue resulta muy importante puesto que la detección temprana, puede evitar que el paciente llegue a una etapa letal de la enfermedad.

3.2.2. Diagnóstico diferencial del Dengue

Al unir los síntomas y signos del paciente para posteriormente construir un síndrome e indagar sobre todas las enfermedades que este pueda presentar, es posible realizar un diagnóstico diferencial, ya que este es un procedimiento mediante el cual el médico excluye la posibilidad de cualquier otra enfermedad, basado en las diferencias del caso del paciente y todos los otros casos posibles Díaz Novás et al. (2006).

El diagnóstico del dengue en los primeros días de evolución es complejo, ya que presenta un cuadro febril agudo con síntomas inespecíficos que pueden confundirse con otras infecciones. Dependiendo de las características epidemiológicas del entorno, es conveniente considerar infecciones por virus como la influenza, chikunguya, hepatitis, fiebre amarilla y otras fiebres hemorrágicas. Además, deben tenerse en cuenta enfermedades como la rubéola, gastroenteritis, fiebre tifoidea y leptospirosis. Por tanto, para el diagnóstico diferencial del dengue es muy importante tener en cuenta la secuencia con la que aparecen los signos y síntomas de la enfermedad Calvo et al. (2016).

3.2.3. Hemograma

El hemograma es una de las pruebas diagnósticas más utilizadas en la práctica médica habitual. Los actuales analizadores automáticos permiten determinar con un grado elevado de fiabilidad, rapidez y un bajo coste los principales parámetros hematológicos en sangre periférica, aportando una valiosa información acerca de las tres series hemáticas glóbulos rojos, blancos y plaquetas Huerta Aragonés et al. (2018). En el diagnóstico del dengue los resultados más utilizados son el recuento de leucocitos y el recuento de plaquetas.

Las siguientes definiciones de algunos parámetros del hemograma fueron tomadas de AACCC (2020)

- **Plaquetas:** Las plaquetas también llamadas trombocitos, son pequeños fragmentos de células que circulan en la sangre y son esenciales para la coagulación normal de la sangre. Cuando hay una lesión y comienza el sangrado, las plaquetas ayudan a detener el sangrado al adherirse al sitio de la lesión y agruparse para formar un tapón temporal. Si una persona tiene una enfermedad o afección que causa plaquetas bajas (trombocitopenia), puede tener un mayor riesgo de sangrado excesivo y hematomas.
- **Hematocrito:** Un hematocrito es una prueba que mide la proporción de su sangre que se compone de glóbulos rojos. La sangre se compone de glóbulos rojos, glóbulos blancos y plaquetas suspendidas en una porción líquida llamada plasma. El hematocrito es una relación entre el volumen de glóbulos rojos y el volumen de todos estos componentes juntos, denominada sangre completa. El valor se expresa como porcentaje o fracción.
- **Hemoglobina:** La hemoglobina es la proteína que contiene hierro que se encuentra en todos los glóbulos rojos y que le da a las células su característico color rojo. La hemoglobina permite que los glóbulos rojos se unan al oxígeno en los pulmones y lo lleven a los tejidos y órganos de todo el cuerpo. También ayuda a transportar una pequeña porción de dióxido de carbono, un producto del metabolismo celular, desde los tejidos y órganos hasta los pulmones, donde se exhala.
- **Linfocitos:** Un linfocito es un tipo de glóbulo blanco que es parte del sistema inmune. Hay dos tipos principales de linfocitos: las células B y las células T. Las células B elaboran los anticuerpos para luchar contra bacterias, virus y toxinas invasoras. Las células T destruyen las propias células del cuerpo que han sido infectadas por virus o que se han vuelto cancerosas.
- **Neutrófilos:** Los neutrófilos son un tipo de glóbulo blanco y ayudan al cuerpo a combatir infecciones. Se puede usar un recuento absoluto de neutrófilos para determinar si hay infección, inflamación, leucemia u otras afecciones. Entre menor sea el recuento absoluto de neutrófilos de una persona, mayor es su riesgo de infección.

- **Monocito:** Tipo de célula inmunitaria que se elabora en la médula ósea y viaja por la sangre hasta los tejidos del cuerpo donde se convierte en un macrófago o una célula dendrítica. Los macrófagos rodean y destruyen los microorganismos, además se tragan el material extraño, eliminan las células muertas y estimulan las respuestas inmunitarias. Durante un proceso inflamatorio, las células dendríticas estimulan las respuestas inmunitarias al llevar los antígenos en su superficie y presentarlos a otras células del sistema inmunitario. Un monocito es un tipo de glóbulo blanco y un tipo de fagocito.

3.2.4. Algoritmos de Clasificación

Méndez (2017) propuso una metodología abordada desde el enfoque bayesiano para construir reglas decisión para el diagnóstico de dengue. Para la construcción de los algoritmos se obtuvieron un total de 25 signos y síntomas que fueron determinados a partir de una revisión bibliográfica y consulta a expertos académicos y médicos clínicos en dengue. Posteriormente, se crearon categorías (ocho en total) de signos y síntomas, agrupando aquellos cuya connotación fuese similar. Además, con el uso de una prueba Chi-cuadrado, se seleccionaron los 12 signos y síntomas mas relacionados con la enfermedad del dengue.

Tanto las categorías de signos y síntomas creadas como los signos y síntomas seleccionados a partir de la prueba Chi-cuadrado, fueron el insumo para la creación de los prototipos clasificadores.

- 1 **Prototipo I:** Consiste en la aplicación sucesiva del teorema de Bayes usando la fórmula con probabilidades puntuales para eventos para el cálculo de probabilidades predictivas. Aquí, el interés radica en cuantificar la probabilidad de tener dengue dado que el individuo presentó una de las combinaciones de categorías de signos y síntomas.
- 2 **Prototipo II:** Se construyó siguiendo la misma metodología para el cálculo de las probabilidades del prototipo I, sin embargo difieren en la selección de signos y síntomas, pues en este caso se tuvo como criterio la prueba chi-cuadrado para hacer una selección más estricta de los mismos, obteniendo así un total de 12 signos y síntomas.
- 3 **Prototipo III y IV:** Estos algoritmos clasificadores se desarrollaron bajo la continuidad en el Teorema de Bayes para calcular la probabilidad de presentar dengue dado que se observó una de las combinaciones de categorías de signos y síntomas. Esta nueva propuesta hace uso de la categorización realizada para el prototipo I como insumos de entrada para el prototipo III, y los síntomas seleccionados por la prueba Chi-cuadrado en el prototipo II como entradas para el prototipo IV, estos dos prototipos pueden interactuar de una forma más directa con la experiencia de los expertos mediante la elicitación de una distribución a priori Méndez (2017).

3.2.5. Gold Standard para Dengue

Según Salech et al. (2008) un Gold Standard o patrón de oro es una técnica diagnóstica que define la presencia de la enfermedad con la máxima certeza conocida.

La detección del virus del dengue mediante el uso de pruebas de laboratorio es considerado un proceso eficiente aunque costoso para algunos centros de salud. Según las directrices de la Organización Mundial de la Salud existen tres pruebas que son consideradas como Gold Standard para el diagnóstico de dengue: Aislamiento rápido del virus, Detección de nucleótidos y pruebas serológicas para seroconversión de IgM (inmunoglobulina M) o IgG (inmunoglobulina G) World Health Organization et al. (2009).

Además, pruebas de laboratorio como: la reacción en cadena por polimerasa de transcriptasa reversa (RT-PCR), la detección del antígeno NS1 por ELISA (Enzyme-Linked ImmunoSorbent Assay) y la detección por medio de cantidad de antígenos IgM e IgG, duran alrededor de 24 horas en dar resultados y son consideradas como Gold Standard Zhang et al. (2014).

3.2.6. Aplicación de pruebas diagnosticas en paralelo

Consiste en realizar simultáneamente dos o más pruebas en un individuo, y se realizan con el fin de aumentar la sensibilidad diagnostica Donis (2012), además aumenta el valor predictivo negativo tras realizarse las dos pruebas y esto lleva a disminuir la especificidad pero se aumenta la sensibilidad, permitiendo detectar con mayor probabilidad los enfermos Escrig-Sos et al. (2006).

3.3. Marco Teórico Estadístico

3.3.1. Validación de una prueba diagnostica

La validez de una prueba diagnóstica se refiere a que esta mide con exactitud lo que desea cuantificar Donis (2012). Entre los diferentes tipos de validez se encuentra la validez de criterio cuya función es evaluar el grado de precisión diagnóstica de la prueba (los algoritmos) en diferentes situaciones durante su desarrollo. Los datos pueden ser datos retrospectivos y, por lo tanto, la validación externa es posible para las pruebas que necesitan un seguimiento prolongado para reunir suficientes eventos de resultado Altman & Bland (1994).

Lynch (1982) trabajó sobre tres tópicos (*perspectiva, robustez y realismo*) relacionados con la validación externa, donde se busca deducir si el valor de los resultados y conclusiones obtenidas, a partir de una muestra, permite una generalización estadística en una población de mayor interés y un entorno diferente. En el diagnóstico clínico, Knottnerus et al. (2002) recomienda que la validez externa se realice preferiblemente con estudios repetidos en otras poblaciones similares que tengan información un gold standard.

3.3.2. Medidas de Desempeño

Diseño de Cohorte (un test diagnóstico)

De acuerdo con Pepe et al. (2003) el diseño de cohorte para un test diagnostico considera la evaluación del desempeño de un test diagnóstico dado un Gold Standard:

Sea D la variable aleatoria que denota el estado del paciente:

$$D = \begin{cases} 1 & ; \text{ El individuo está enfermo} \\ 0 & ; \text{ El individuo está sano} \end{cases} \quad (3-1)$$

Sea Y la variable aleatoria que denota el resultado de la prueba diagnóstica:

$$Y = \begin{cases} 1 & ; \text{ Resultado positivo} \\ 0 & ; \text{ Resultado negativo} \end{cases} \quad (3-2)$$

Los resultados de la prueba pueden ser clasificados como:

	$D = 0$	$D = 1$
$Y = 0$	a	b
$Y = 1$	c	d

Tabla 3-1.: Tabla de clasificación modelo de cohorte con un test diagnóstico

Dos pruebas para diagnóstico bajo diseño de cohorte

Este diseño considera la evaluación del desempeño de dos tests diagnósticos dado un gold standard:

Sea D la variable aleatoria que denota el resultado del Gold Standard:

$$D = \begin{cases} 1 & ; \text{ El individuo está enfermo} \\ 0 & ; \text{ El individuo está sano} \end{cases} \quad (3-3)$$

Sea X_i con $i = 1, 2$ la variable aleatoria que denota el resultado de la prueba diagnóstica:

$$X_i = \begin{cases} 1 & ; \text{Resultado positivo} \\ 0 & ; \text{Resultado negativo} \end{cases} \quad (3-4)$$

	D=1		D=0	
	$X_2 = 1$	$X_2 = 0$	$X_2 = 1$	$X_2 = 0$
$X_1 = 1$	r_1	r_2	r_5	r_6
$X_1 = 0$	r_3	r_4	r_7	r_8

Tabla 3-2.: Arreglo de datos en un diseño de cohorte para validación de dos pruebas para diagnóstico

1 **Sensibilidad y Especificidad:** Altman & Bland (1994) afirman que son un enfoque para cuantificar la capacidad de diagnóstico de la prueba. Sin embargo, en la práctica clínica, el resultado de la prueba es todo lo que se conoce, por lo que se desea saber qué tan buena es para predecir anormalidades.

- **Sensibilidad:** Proporción de verdaderos positivos correctamente identificados por el test del total de individuos enfermos según el estándar de referencia Altman & Bland (1994).
- **Especificidad:** Proporción de verdaderos negativos que fueron correctamente identificados por el test, del total de individuos sanos según el estándar de referencia Altman & Bland (1994).

2 **Valores Predictivos:** Los valores predictivos positivos (VPP) y negativos (VPN) buscan responder la inquietud sobre la probabilidad de que la persona efectivamente esté enferma ante un test positivo, es decir, corresponden a las probabilidades de que un nuevo individuo en la población al que no se le ha aplicado un Gold Standard esté realmente enfermo o no, cuando se cuenta solo con la información del resultado de la prueba bajo evaluación.

- **Valor Predictivo Positivo:** Corresponde a la probabilidad condicional de que el paciente tenga la enfermedad, dado que el test resultó positivo. Expresado de otra manera, es la proporción de pacientes con la prueba diagnóstica positiva que efectivamente tienen la condición Bravo-Grau & Cruz (2015).

$$VPP = \frac{\tau\beta}{\tau\beta + (1 - \tau)(1 - \lambda)} \quad (3-5)$$

- **Valor Predictivo Negativo:** Corresponde a la probabilidad condicional de que el paciente no tenga la enfermedad, dado que la prueba diagnóstica resultó negativa. En otras palabras, es la probabilidad de que el individuo no tenga la condición en estudio luego de que el test es negativo Bravo-Grau & Cruz (2015).

$$VPN = \frac{(1 - \tau)\lambda}{(1 - \tau)\lambda + \tau(1 - \beta)} \quad (3-6)$$

Donde $\tau = \text{Prevalencia}$, $\beta = \text{Sensibilidad}$ y $\lambda = \text{Especificidad}$

Prevalencia

La prevalencia mide la proporción de personas que se encuentran enfermas al momento de evaluar el padecimiento en la población, por lo tanto, no hay tiempo de seguimiento Fajardo Gutiérrez (2017).

3.3.3. Falta de Identificabilidad en el Modelo

Cuando se considera un modelo con dos test diagnósticos aplicados a la misma población el diseño muestral fija el total de cada tabla, por lo que los datos forman dos tablas de 2×2 , cada una con tres grados de libertad, dando como resultado un total de seis grados de libertad (ver Tabla 3-2). Cuando se asume que las pruebas son condicionalmente independientes por el estado de la enfermedad se estiman alrededor de seis parámetros (dos sensibilidades, dos especificidades y dos prevalencias) y comúnmente se cree que si los grados de libertad en los datos cumplen o exceden el número de parámetros en un modelo ajustado entonces el modelo es identificable Dendukuri & Joseph (2001), sin embargo, la suposición de independencia condicional entre las pruebas es poco razonable especialmente cuando la prueba es aplicada a un mismo individuo y apuntan a medir lo mismo, esta suposición podría dar lugar a estimaciones sesgadas de la sensibilidad y la especificidad Vacek (1985), por ello se sugiere incluir al modelo la correlación entre las pruebas sin embargo esto implica incluir dos parámetros más al modelo, por lo que ya se tendrían más parámetros de los que se pueden estimar con los datos disponibles lo cual se traduce a la falta de identificabilidad del modelo. Georgiadis et al. (2003)

Jones et al. (2010) afirma que bajo el enfoque frecuentista, la no identificabilidad sólo puede ser eludida mediante la fijación de los valores de ciertos parámetros, mientras que, para el enfoque Bayesiano, la identificabilidad no es obligatoria si se dispone de buena información previa.

3.3.4. Métodos de remuestreo (Bootstrap)

Solanas & Olivera (1992) afirma que las técnicas de remuestreo se caracterizan por la obtención de submuestras a partir de los datos que constituyen la muestra original, permitiendo evaluar diferentes propiedades de los estimadores. En este sentido, la técnica bootstrap debe ser considerada como un plan de remuestreo, posibilitando estudiar el error estadístico, ya sea en cuanto a sesgo, error estándar o tasa de error en una predicción.

Según Fox (2002), el bootstrap no paramétrico permite estimar la distribución empírica de un estadístico sin hacer suposiciones sobre la forma de la población. Adicionalmente, especifica los pasos para realizar el remuestreo, de tal forma que:

- 1 El investigador selecciona un estadístico de interés $\hat{\theta}$
- 2 Obtiene una muestra correspondiente a la realización de n variables aleatorias independientes e idénticamente distribuidas, y esta representará la muestra original mediante $S = \{x_1, x_2, \dots, x_n\}$.
- 3 Se extrae una muestra de tamaño n de entre los elementos de S , muestreando con reemplazo, obteniendo así $S_1^* = \{x_{11}^*, x_{12}^*, \dots, x_{1n}^*\}$ un subconjunto de datos que representan la muestra bootstrap.
- 4 Se repite el proceso anterior B veces, de este modo, la b^{th} muestra bootstrap se denota como $S_b^* = \{x_{b1}^*, x_{b2}^*, \dots, x_{bn}^*\}$

3.3.5. Inferencia Bayesiana

Gelman et al. (2013) menciona que la inferencia estadística se refiere a sacar conclusiones, a partir de datos numéricos, sobre cantidades que no se observan. Además, la característica esencial de los métodos bayesianos es su uso explícito de la probabilidad para cuantificar la incertidumbre en las inferencias basadas en el análisis de datos estadísticos.

El paradigma bayesiano se basa en una interpretación de la probabilidad como una medida de incertidumbre racional y condicional, que coincide estrechamente con el sentido de la palabra “probabilidad” en el lenguaje ordinario. La inferencia estadística sobre una cantidad de interés se describe como la modificación de la incertidumbre sobre su valor a la luz de la evidencia, y el teorema de Bayes especifica con precisión cómo se debe hacer esta modificación Bernardo & Smith (2009).

- **Teorema de Bayes:** Suponga que $y' = (y_1, y_2, \dots, y_n)$ es un vector de n observaciones, cuya distribución de probabilidad depende de los valores de k parámetros $\theta' = (\theta_1, \theta_2, \dots, \theta_k)$. Suponga también que θ tiene una distribución de probabilidad $p(\theta)$. Entonces:

$$p(y|\theta)p(\theta) = p(y, \theta) = p(\theta|y)p(y) \quad (3-7)$$

Dados los datos observados y , la distribución condicional de θ es:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (3-8)$$

También se puede escribir:

$$p(y) = E[p(y|\theta)] = c^{-1} = \begin{cases} \int p(y|\theta)p(\theta)d(\theta) & ; \theta \text{ Continuo} \\ \sum p(y|\theta)p(\theta) & ; \theta \text{ Discreto} \end{cases} \quad (3-9)$$

donde la suma o integral toman valores dependiendo del rango de θ y donde la $E[f(\theta)]$ es el valor esperado de $f(\theta)$ con respecto a la distribución de $p(\theta)$. Por lo tanto, se puede escribir (3-8) como:

$$p(\theta|y) = cp(y|\theta)p(\theta) \quad (3-10)$$

La expresión (3-8), o su equivalente (3-10), generalmente se conoce como el Teorema de Bayes. En esta expresión $p(\theta)$ brinda información sobre θ sin el conocimiento de los datos, es conocida como la distribución de θ *a priori*. Del mismo modo, $p(\theta|y)$ da información sobre θ dado el conocimiento de los datos, es llamada distribución *posteriori* de θ dado y . La cantidad c , es una constante “normalizadora” necesaria para asegurar que la distribución posterior $p(\theta|y)$ se integre o sume uno.

Tomado de: Box & Tiao (2011)

3.3.6. Receiver operating characteristic curve (Curva ROC)

El análisis ROC se utiliza en epidemiología clínica para cuantificar la precisión con la que las pruebas (o sistemas) de diagnóstico médico pueden discriminar entre dos estados de pacientes, típicamente denominados “enfermos” y “no enfermo” Hajian-Tilaki (2013).

En las pruebas de diagnóstico con resultado dicotómico (resultados positivos / negativos de la prueba), la forma convencional de evaluación es mediante el calculo de la sensibilidad y especificidad como medidas de precisión Swets (1979), de este modo, las curvas ROC representan sobre un sistema de coordenadas la sensibilidad y especificidad de una prueba, debido a esto es un método que se utiliza frecuentemente para determinar la calidad diagnóstica de la misma.

Martínez Cambor (2007) define la curva ROC de la siguiente manera: Se tiene una medida determinada X realizada sobre una población de positivos (enfermos) X_P , y otra de negativos (sanos) X_N con función de distribución G y F , respectivamente. Suponiendo que $E(X_N) \leq E(X_P)$, para clasificar a los individuos de un grupo u otro se debe fijar un criterio, o, punto de corte, a partir del cual un individuo será considerado positivo. Por tanto, fijado un punto de corte t , la sensibilidad de la prueba vendrá determinada por $1 - G(t)$, la especificidad por $F(t)$ y quedando por tanto determinada la curva ROC por las coordenadas $(1 - F(t), 1 - G(t))$, o, equivalentemente por la función

$$ROC(t) = 1 - G(F^{-1}(1 - t)), \quad t \in [0, 1] \quad (3-11)$$

Martínez Cambor (2007) además, asegura que cuando se desconocen las distribuciones de la variable en las poblaciones de positivos y negativos y deben estimarse a partir de muestras aleatorias. Una de las posibilidades es suponer que las poblaciones siguen algún modelo paramétrico o sustituir las funciones de distribución desconocidas por sus Funciones de Distribución Empíricas (FDE) o por las Funciones de Distribución Empírica Suavizadas (FDES).

Estimación empírica de la Curva ROC

Según Gonçalves et al. (2014) el método no paramétrico más simple para estimar la curva ROC es la estimación empírica, la cual consiste en sustituir en (3-11) las funciones de distribución desconocidas por sus correspondientes Funciones de Distribución Empíricas. Específicamente, la estimación empírica de la curva ROC viene dada por:

$$\widehat{ROC}(t) = 1 - \widehat{G}(\widehat{F}^{-1}(1 - t)) \quad (3-12)$$

Donde \widehat{G} es la función de distribución empírica asociada a la muestra X_P y \widehat{F} es la función de distribución empírica asociada a la muestra X_N .

Área Bajo la Curva ROC (AUC)

Según lo expuesto por Cerda & Cifuentes (2012) una de las formas de resumir las curvas ROC es mediante la estimación del área bajo la curva (AUC, area under the curve), la cual mide la capacidad discriminatoria de un test diagnóstico, siendo más discriminatorio el test con la mayor AUC. Más concretamente, el AUC puede interpretarse como la probabilidad de que un sujeto enfermo elegido aleatoriamente sea clasificado como más propenso a estar enfermo que un sujeto no enfermo elegido aleatoriamente Hanley & McNeil (1982).

Gonçalves et al. (2014) define el AUC como:

$$AUC = \int_0^1 ROC(u) \cdot du \quad (3-13)$$

Cuando la curva ROC se genera por el método empírico, el área puede calcularse mediante la regla trapezoidal, es decir, como la suma de las áreas de todos los rectángulos y trapecios correspondientes a los empates $(1 - G(t), 1 - F(t))$ que se pueden formar bajo la curva de Ullibarri Galparsoro & Fernández (1998).

4. Metodología

4.1. Datos:

En este trabajo se utilizará la base de datos del proyecto P4-RedAedes, la cual consta de 1039 registros de sujetos febriles entre 0 y 80 años de edad, y que se encontraban en zonas endémicas de Colombia (Cali, Piedecuesta y Yopal) entre agosto 2017 y julio 2018. La base contiene variables clínicas relacionadas con dengue, resultado del hemograma y un algoritmo de pruebas diagnósticas de laboratorio para dengue como prueba de referencia (RT-PCR, NS1 ELISA, IgM), según el diagnostico confirmatorio se identificaron 25 dengue confirmado, no dengue 307, dengue probable 514 e indeterminado 193. El archivo de datos final, con el cual se desarrollo este trabajo, solo considero a pacientes con dengue positivo o negativo, y que además tuvieran información completa sobre todas las variables consideradas en este estudio, estos son 324 individuos que se dividen entre 24 dengue positivo y 300 dengue negativo.

Se realizó una exploración de la base de datos donde previamente se leyó y conversó con la investigadora sobre el significado clínico de cada signo y síntoma para comprender los nombres técnicos médicos (como se interpretan en términos generales y clínicos); seguidamente, se estructuró el archivo de datos para realizar el análisis estadístico (limpiar y codificar).

4.2. Identificación de variables asociadas al Dengue

4.2.1. Signos y síntomas

Méndez (2017) realizó la identificación de signos y síntomas más relevantes para el diagnostico del dengue, puesto que son el insumo principal para la construcción de los algoritmos clasificadores junto con los resultados del hemograma, los cuales se presentan en la Tabla 4-1.

Síntoma	Descripción
Gingivorragia (*)	Hemorragia espontánea que se produce en las encías.
Hematemesis (*)	Expulsión de vómito con sangre procedente del tracto digestivo alto.
Cefalea	Dolor de cabeza intenso y persistente que va acompañado de sensación de pesadez.
Escalofrío	Es un temblor involuntario y repentino asociado a una sensación de frío más o menos pronunciada.
Hoporexia	Pérdida o disminución parcial del apetito.
Odinofagia	Dolor de garganta.
Exantema	Es una erupción cutánea, como enrojecimiento, protuberancia, y a veces pústulas, que cubre un área grande del cuerpo.
Prurito	Es un hormigueo o irritación de la piel que provoca el deseo de rascarse en la zona.
Vómito	Expulsar el contenido del estómago a través de la boca.
Hepatomegalia	Aumento del tamaño del hígado, por sobre los límites estimados como normales para cada grupo de edad.
Epistaxis (*)	Hemorragia nasal o sangrado nasal.
Petequías (*)	Son pequeños derrames vasculares cutáneos del tamaño de una cabeza de alfiler.
Rinorrea	Flujo o emisión abundante de líquido por la nariz

Tabla 4-1.: Signos y síntomas identificados

Los signos y síntomas que se encuentran con (*) en la Tabla **4-1** fueron agrupados en una categoría llamada **Manifestaciones Hemorrágicas** con el fin de disminuir el número de combinaciones a analizar. Las categorías con las que se trabajaron y sus respectivos síntomas asociados se encuentran en la tabla **4-2**, estas categorías fueron conformadas con la ayuda de médicos especialistas en dengue. Las categorías con (*) hacen parte del diagnostico diferencial de la enfermedad.

Categoría	Sintoma
Dolor General	Artralgias
	Cefalea
	Dolor retrocular
	Mialgias
Signo de Alarma	Diarreas
	Dolor Abdominal
	Higado Palpable o Hepatomelagia
	Vómito
	Edemas
Malestar General	Escalosfríos
	Hiporexia
Erupción Cutánea	Exantema
	Erupción maculopapular
	Eritema
	Rahs
Manifestaciones Hemorrágicas	Hemorragia
	Petequias
Ictericia (*)	Ictericia
Alteración neurológica	Irritabilidad
	Insomnio
	Somnolencia
Síntomas Respiratorio (*)	Odinofagia
	Rinorrea
	Inyección conjuntival
	Tos

Tabla 4-2.: Categorías de signos y síntomas conformadas

4.2.2. Hemograma

Las variables asociadas al hemograma tenidas en cuenta para el estudio son de tipo cuantitativo continuo y corresponden a: conteo de plaquetas, conteo de leucocitos, conteo de monocitos, razón de linfocitos sobre neutrófilos y razón entre hematocrito y hemoglobina, donde fue necesario fijar un valor para cada una con la finalidad de transformarlas a variables dicotómicas, éste valor hace referencia al punto de corte de cada variable, el cual fue definido a partir de la búsqueda en la literatura y con ayuda de una médica epidemióloga que cuenta con una amplia experiencia en el diagnóstico del dengue, de esta manera, es posible clasificar al individuo como 1 (presencia) ó 0 (ausencia) de acuerdo a la regla de cada parámetro del hemograma.

- $(Plaquetas \leq 150000) = 1$ ó $(Plaquetas > 150000) = 0$
- $(Leucocitos \leq 4500) = 1$ ó $(Leucocitos > 4500) = 0$
- $(Monocitos > 1000) = 1$ ó $(Monocitos \leq 1000) = 0$
- $(Linfocitos/Neutrofilos > 0,8) = 1$ ó $(Linfocitos/Neutrofilos \leq 0,8) = 0$
- $(Hematocrito/Hemoglobina \geq 3) = 1$ ó $(Hematocrito/Hemoglobina < 3) = 0$

4.3. Definición de nuevos Prototipos Clasificadores

Para la creación de las nuevas reglas de decisión, cada algoritmo de Méndez (2017) fue actualizado adicionandole la clasificación realizada mediante cada una de las cinco variables del hemograma, es decir, en los algoritmos I y III de Méndez (2017) a cada individuo le correspondía un vector de tamaño ocho pues se identificaba si presentaba o no cada una de las categorías de signos y síntomas consignados en la Tabla 4-2, mientras que en los algoritmos II y IV cada individuo tenía un vector de tamaño 12 pues se consideraban los signos y síntomas de la Tabla 4-1, generando así que, con la nueva actualización quedaran vectores de tamaño 13 y 17 respectivamente, donde en cada componente se almacena la información correspondiente a si el individuo tiene presencia (1) o ausencia (0) del signo, síntoma o variable del hemograma.

En otras palabras, cada vez que un individuo hace consulta, en ese momento presenta un conjunto de síntomas, es decir, si el médico revisa cada uno de éstos, encontrará que el individuo presenta unos y otros no. Por lo anterior, se pueden armar las permutaciones sin repetición de los n síntomas o categorías con los dos posibles estados, presencia o ausencia, en este orden de ideas se pueden armar 2^n arreglos distintos con las permutaciones entre la presencia (1) y la ausencia (0) de los n síntomas.

Entonces, sea V la matriz de dimensión $2^n \times n$ donde n hace referencia a la cantidad de categorías o signos y síntomas según sea el caso del algoritmo, por lo tanto V será la matriz que almacenará cada una de las 2^n permutaciones posibles entre la presencia o ausencia de las n categorías de signos y síntomas. La primera fila de la matriz V estará compuesta por el valor de 1 en cada una de sus celdas indicando que todas las n categorías de signos y síntomas se encontraron presentes en el individuo, mientras que la ultima fila estará conformado en la totalidad de sus celdas por el valor de 0 indicando que no se observó la presencia de alguna de las n categorías en el individuo, a continuación se ejemplifica el ordenamiento de la matriz V donde cada una de las filas de la matriz será nombrada como v_j donde j varía de 1 hasta 2^n .

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & \cdots \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & \cdots \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & \cdots \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \end{bmatrix}_{2^n \times n}$$

Con base en lo expuesto anteriormente, se puede plantear el siguiente evento de interés.

$$B_j = \{ \text{El individuo presentó la combinación } v_j \text{ de categorías de signos y síntomas} \}$$

Adicionalmente, se creó un algoritmo que solo iba a tener en cuenta los resultados del hemograma, esto se hizo siguiendo la misma metodología propuesta por Méndez (2017) para la creación de reglas de decisión.

En todos los prototipos el interés radica en cuantificar la probabilidad de tener dengue dado que el individuo presentó una de las v_j combinaciones de signos o síntomas y parámetros del hemograma.

- **Prototipo I:** Incluye las ocho categorías y los parámetros del hemograma, consiste en la aplicación sucesiva de la fórmula de Bayes con probabilidades puntuales para eventos para el cálculo de probabilidades predictivas.
- **Prototipo II:** Se construyó siguiendo la misma metodología para el cálculo de las probabilidades del prototipo I, sin embargo difieren en la selección de signos y síntomas, ya que este considera los 12 signos y síntomas de la Tabla 4-1.
- **Prototipo III y IV:** Para calcular la probabilidad de presentar dengue dado que se observó una de las combinaciones de categorías de signos y síntomas, estos algoritmos clasificadores se desarrollaron asumiendo la forma de la fórmula de Bayes para variables continuas. Esta nueva propuesta hace uso de la categorización realizada para el prototipo I como insumos de entrada para el prototipo III, y los síntomas seleccionados en el prototipo II como entradas para el prototipo IV, estos dos prototipos pueden interactuar de una forma más directa con la experiencia de los expertos mediante la elicitación de una distribución a priori Méndez (2017).

- **Prototipo V:** Incluye únicamente los parámetros del hemograma y para el calculo de las probabilidades se considera la aplicación de la fórmula de Bayes con probabilidades para eventos.
- **Prototipo VI:** Incluye únicamente los parámetros del hemograma y para el calculo de las probabilidades se asume la forma de la fórmula de Bayes para variables continuas.

4.4. Modelo estadístico

4.4.1. Experimento Estadístico

El paciente ingresa a la consulta médica con sospecha de tener dengue, se le interroga sobre los signos y síntomas asociados al cuadro febril que presenta, simultáneamente se va registrando la información. Adicionalmente, a cada individuo se le tomó una muestra de sangre en laboratorio para obtener los resultados del hemograma y la prueba confirmatoria (Gold Standard) que permite determinar la presencia de la enfermedad del dengue.

4.4.2. Eventos

A continuación se definen los eventos:

A_i = El individuo tiene un resultado positivo por el prototipo i donde $i = 1, \dots, 6$

GS = El individuo tiene un resultado positivo por el Gold Standard

4.4.3. Variables aleatorias

A cada evento A_i se le asigna una variable aleatoria X_i que cuantifique el respectivo resultado:

$$X_i = \begin{cases} 1 & ; \text{Ocurre } A_i \\ 0 & ; \text{No ocurre } A_i \end{cases} \quad D = \begin{cases} 1 & ; \text{Ocurre GS} \\ 0 & ; \text{No ocurre GS} \end{cases}$$

4.4.4. Calculo de probabilidades predictivas de Bayes

A continuación se describe el procedimiento que se usó para calcular la probabilidad de tener dengue dado que el paciente presenta la j –ésima combinación de signos y síntomas para los protitipos I, II ,III, y IV, o parámetros del hemograma para los prototipos V y VI, es decir, se busca cuantificar $P(D = 1|B_j)$.

Caso asumiendo probabilidades puntuales y eventos

Para aplicar el Teorema de Bayes en el caso discreto, debido a que el evento B_j está compuesto por las posibles combinaciones entre presencia o ausencia de las categorías, se asumió que los eventos asociados a la presencia o ausencia de cada categoría de signos y síntomas o solamente síntomas son condicionalmente independientes entre sí. Sea Z_n una variable aleatoria asociada a la presencia-ausencia (0 ó 1) de cada una de las categorías de signos y síntomas o parámetros del hemograma.

Por lo descrito anteriormente, la variable aleatoria Z_n adquiere esta propiedad de independencia. Entonces, la probabilidad $P(B_j|D = 1)$ puede descomponerse en el producto de la probabilidades condicionadas de cada categoría de signos y síntomas.

El supuesto de independencia condicional entre los eventos asociados a la presencia o ausencia de las categorías de signos y síntomas se fundamenta en el hecho de que la presencia de algún síntoma no influye en la presencia de los otros, a manera de ejemplo: el hecho de que en un individuo en el cual se observó que presentaba vómito no implica intrínsecamente se le deba de observar un conteo de plaquetas por debajo de 150000 en el hemograma o cualquiera de los demás signos o síntomas.

Entonces, usando la aplicación de la formula de Bayes se obtiene que:

$$P(D = 1|B_j) = \frac{P(Z_1 = z_1, Z_2 = z_2, \dots, Z_n = z_n|D = 1)P(D = 1)}{P(B_j)} \quad (4-1)$$

Donde $P(D = 1)$ es asumida como el conocimiento previo sobre la probabilidad de observar un individuo con dengue en la población (prevalencia), se tiene entonces que $P(B_j|D = 1)$ es la probabilidad de observar la j – ésima combinación de categorías en un individuo que tiene dengue; $P(B_j)$ es la probabilidad total de encontrar la combinación de signos y síntomas ó parámetros del hemograma que componen la v_j fila de la matriz V , entonces:

$$P(B_j) = P(D = 1) \cdot P(B_j|D = 1) + P(D = 0) \cdot P(B_j|D = 0) \quad (4-2)$$

Por la propiedad de la independencia condicional expuesta por Dawid (1979), $P(B_j|D = 1)$ puede calcularse como el producto de las probabilidades condicionales de cada categoría que conforma la fila v_j , que a su vez conforman al evento B_j , entonces:

$$\begin{aligned} P(B_j|D = 1) &= P(Z_1 = z_1|D = 1) \cdot P(Z_2 = z_2|D = 1) \cdots P(Z_k = z_k|D = 1) \\ &= \prod_{k=1}^n P(Z_k = z_k|D = 1) \quad \forall j = 1, \dots, 2^n \end{aligned} \quad (4-3)$$

Así pues, la ecuación 4-1 se puede escribir como:

$$P(D = 1|B_j) = \frac{\prod_{k=1}^n P(Z_k = z_k|D = 1)P(D = 1)}{P(D = 1) \prod_{k=1}^n P(Z_k = z_k|D = 1) + P(D = 0) \prod_{k=1}^n P(Z_k = z_k|D = 0)} \quad (4-4)$$

Se calculó para cada una de las combinaciones de signos y síntomas o categorías de signos y síntomas según sea el caso del algoritmo la formula de Bayes de la ecuación 4-4. El valor de la prevalencia de la enfermedad del dengue $P(D = 1)$ se fijó en 0,1.

Aproximación usando variables aleatorias

Para el calculo de $P(D = 1|B_j)$ asumiendo la formula de bayes para variables aleatorias continuas se define la variable aleatoria Y_j , la cual identifica la cantidad de individuos enfermos con dengue que presentaron la j –ésima combinación de signos y síntomas ó parámetros del hemograma, además, sea θ_j la proporción de individuos enfermos con dengue que presentaron la j –ésima combinación de signos y síntomas ó parámetros del hemograma. Por lo tanto, $Y_j \sim \text{Bin}(\theta_j, (Y_j + Y_j^c))$.

Como es de interés estimar la cantidad θ_j se asignó una distribución a priori no informativa $\text{Beta}(1, 1)$ a este parámetro, pues no se contaba con información externa sobre esta probabilidad, de esta manera, fue posible aproximar $P(D = 1|B_j)$ generando 1000 valores aleatorios provenientes de la distribución posterior $\text{Beta}(\alpha + Y_j; \beta + Y_j^c)$ y calculando el promedio de estos valores.

De esta manera, la estimación de la probabilidad $P(D = 1|B_j)$ se define como:

$$P(D = 1|B_j) = \theta_j = \frac{\sum_{m=1}^{1000} \gamma_m}{1000} \quad (4-5)$$

donde γ es un valor generado de la distribución posterior y m corresponde al número de simulaciones realizadas.

4.4.5. Puntos de Corte para la probabilidad $P(D = 1|B_j)$

Como se mostró anteriormente, al aplicar cualquiera de estos algoritmos a un individuo lo que se obtiene como resultado es una probabilidad y a lo que se pretende llegar es a tener como resultado final la clasificación del individuo como enfermo o no enfermo de dengue, por lo tanto era necesario encontrar un valor δ óptimo para cada algoritmo que sirviera como punto de corte para las probabilidades y así, usar este valor como criterio de clasificación para la distinción de individuos como enfermos o no enfermos de dengue.

Se construyeron curvas ROC del tipo no paramétrico y fue el índice de Youden el método utilizado para encontrar el valor óptimo de δ para cada uno de los algoritmos de clasificación, pues su criterio se basa en la maximización de la sensibilidad y especificidad.

Con el propósito de hacer que la estimación de la curva ROC fuera más precisa y evaluar como era el comportamiento del índice de Youden en distintas bases de datos, se propuso ampliar el archivo de datos que se tenía inicialmente con base en un algoritmo de simulación planteado en este trabajo de grado. Para ejecutar el algoritmo de simulación se necesita como insumo un archivo de datos que contenga el verdadero estado del paciente (enfermo-sano) y la probabilidad de interés $P(D = 1|B_j)$; en este caso, se utilizó como insumo para el algoritmo de simulación el archivo de datos con el cual se calcularon las probabilidades predictivas (324 registros). A continuación, se describe el paso a paso de esta simulación:

1. Se dividió el conjunto de datos en 2 grupos (enfermos y no enfermos) de acuerdo al resultado del gold standard.
2. Se calculó la media y varianza de las probabilidades predictivas ($P(D = 1|B_j)$) en cada uno de los grupos formados.
3. Asumiendo que las probabilidades para cada uno de los grupos formados (enfermos y no enfermos) siguen una distribución beta, se estimaron los parámetros para cada una de estas distribuciones usando el método de los momentos, esto es, igualar la media y varianza de cada una a los valores calculados en el paso anterior.
4. Se simularon 1.000 valores de la distribución beta con los parámetros estimados para cada uno de los grupos.
5. Se construyó el archivo de datos final el cual contiene 2.000 registros (filas), donde 1.000 corresponden a las simulaciones del grupo de enfermos y las otras 1.000 a las del grupo de no enfermo, además, se tienen dos columnas donde la primera almacena el resultado del diagnóstico de dengue y la segunda la probabilidad predictiva calculada.
6. Para el archivo de datos final, se estimó el punto de corte con el criterio del índice de Youden y las medidas de desempeño (sensibilidad y especificidad).
7. Este procedimiento fue repetido un total de 1.000 veces, obteniendo así 1.000 valores estimados para cada uno de los siguientes parámetros: punto de corte, sensibilidad y especificidad.
8. Finalmente, se utilizó como medida de resumen para cada parámetro el promedio de las estimaciones.

El procedimiento descrito anteriormente se realizó para cada uno de los prototipos clasificadores y adicionalmente, su capacidad clasificatoria se midió a través del área bajo la curva ROC (AUC).

La construcción de archivos de datos de acuerdo al procedimiento planteado previamente resultó ser una herramienta importante para la validación de los algoritmos de clasificación, pues al generar archivos de datos con un mayor volumen de registros (individuos) permite analizar la convergencia de sus medidas de desempeño y probabilidades predictivas.

4.4.6. Cálculo de los valores predictivos

Para realizar el calculo de los valores predictivos, los valores de sensibilidad y especificidad que se usaron fueron la estimación que resulta luego de aplicar los pasos de la sección 4.4.5. En cuanto a la prevalencia, se dieron diferentes valores entre cero y uno, y se calcularon los valores predictivos para cada una, posteriormente se graficaron las diferentes curvas para cada prototipo.

4.5. Diseño de Cohorte (un test diagnostico)

4.5.1. Especificación del Modelo

Dado que las variables X_i y D anteriormente declaradas son binarias y asumiendo independencia condicional entre individuos es posible afirmar que éstas siguen un distribución Bernoulli, es así que: $X_i \sim \text{Bernoulli}(\theta_i)$ y $D \sim \text{Bernoulli}(\theta_D)$, de aquí se obtiene una prevalencia real de (θ_D) la cual corresponde al cociente entre el número de individuos clasificados como dengue positivo según el Gold Standard y el total de individuos, y un conjunto de prevalencias aparentes (θ_i) que corresponde al cociente entre el número de individuos clasificados como dengue positivo según cada uno de los algoritmos de clasificación y el total de pacientes.

Prototipo \ Gold Standard	Positivo (D=1)	Negativo (D=0)	TOTAL
Positivo ($X_i = 1$)	a	b	a + b
Negativo ($X_i = 0$)	c	d	c + d
TOTAL	a + c	b + d	a + b + c + d

Tabla 4-3.: Tabla de clasificación modelo de cohorte con un test diagnóstico en términos de variables aleatorias

Dada la tabla **4-3** se definen las variables aleatorias como:

$$w_1 = \begin{cases} 1; & X_i = 1, D = 1 \\ 0; & \text{En otro caso} \end{cases} \quad (4-6)$$

$$w_2 = \begin{cases} 1; & X_i = 1, D = 0 \\ 0; & \text{En otro caso} \end{cases} \quad (4-7)$$

$$w_3 = \begin{cases} 1; & X_i = 0, D = 1 \\ 0; & \text{En otro caso} \end{cases} \quad (4-8)$$

$$w_4 = \begin{cases} 1; & X_i = 0, D = 0 \\ 0; & \text{En otro caso} \end{cases} \quad (4-9)$$

Donde:

$$W_1 = \sum_{j=1}^n w_{1j} \quad , \quad W_2 = \sum_{j=1}^n w_{2j} \quad , \quad W_3 = \sum_{j=1}^n w_{3j} \quad \text{y} \quad W_4 = \sum_{j=1}^n w_{4j}$$

$$P(W_1 = a, W_2 = b, W_3 = c, W_4 = d | n, \alpha_1, \alpha_2, \alpha_3, \alpha_4) = \frac{n!}{a! \cdot b! \cdot c! \cdot d!} \alpha_1^a \cdot \alpha_2^b \cdot \alpha_3^c \cdot \alpha_4^d \quad (4-10)$$

Sujeto a:

$$\alpha_k \in (0, 1) \quad ; \quad \sum_{k=1}^4 \alpha_k = 1 \quad ; \quad \text{y} \quad a + b + c + d = n$$

Dado que el interés resulta en estimar $\tau = P(D = 1) =$ prevalencia, $\lambda = P(X_i = 0 | D = 0) =$ especificidad y $\beta = P(X_i = 1 | D = 1) =$ sensibilidad, y el modelo como está propuesto está indexado por cuatro parámetros que identifican las probabilidades conjuntas en la Tabla **4-3**, se hace necesario reparametrizarlo para poner en evidencia los parámetros de interés.

Para escribir el modelo estadístico asociado a una distribución multinomial que refleja en forma simbólica el proceso realizado para obtener los datos en campo y completar la tabla cruzada **4-3**, y el mismo modelo expresado en la forma requerida para el proceso de inferencia estadística. Entonces:

$$\alpha_1 = P(W_1 = 1) = P(X_i = 1, D = 1) = P(D = 1) \cdot P(X_i = 1 | D = 1) = \tau \cdot \beta \quad (4-11)$$

$$\alpha_2 = P(W_2 = 1) = P(X_i = 1, D = 0) = P(D = 0) \cdot P(X_i = 1 | D = 0) = (1 - \tau) \cdot (1 - \lambda) \quad (4-12)$$

$$\alpha_3 = P(W_3 = 1) = P(X_i = 0, D = 1) = P(D = 1) \cdot P(X_i = 0|D = 1) = \tau \cdot (1 - \beta) \quad (4-13)$$

$$\alpha_4 = P(W_4 = 1) = P(X_i = 0, D = 0) = P(D = 0) \cdot P(X_i = 0|D = 0) = (1 - \tau) \cdot \lambda \quad (4-14)$$

Reescribiendo el modelo en términos de los parámetros de interés, se obtiene:

$$P(W_1 = a, W_2 = b, W_3 = c, W_4 = d|n, \tau, \lambda, \beta) = k \cdot (\tau\beta)^a [(1 - \tau)(1 - \lambda)]^b [\tau(1 - \beta)]^c [(1 - \tau)\lambda]^d \quad (4-15)$$

Con $\tau \in (0, 1)$, $\beta \in (0, 1)$, $\lambda \in (0, 1)$ y $k = \frac{n!}{a! \cdot b! \cdot c! \cdot d!}$ una constante. De esta forma se genera un modelo con tres parámetros y tres grados de libertad, resultando un modelo especificado por el producto de tres distribuciones binomiales independientes.

4.5.2. Estimación de los parámetros de desempeño

Estimación clásica

Asumiendo el procedimiento propio del paradigma clásico de la estadística, se realizó la estimación por máxima verosimilitud para obtener las estimaciones de las sensibilidades y especificidades junto con sus intervalos de confianza considerando un nivel de significancia γ , por tanto, las estimaciones serán:

$$\text{Sensibilidad} = \hat{\beta} = \frac{a}{c + a} \quad (4-16)$$

$$IC(\beta)_{1-\gamma} = \left\{ \hat{\beta} \pm Z_{1-\gamma} \sqrt{\frac{\hat{\beta}(1 - \hat{\beta})}{c + a}} \right\} \quad (4-17)$$

$$\text{Especificidad} = \hat{\lambda} = \frac{d}{b + d} \quad (4-18)$$

$$IC(\lambda)_{1-\gamma} = \left\{ \hat{\lambda} \pm Z_{1-\gamma} \sqrt{\frac{\hat{\lambda}(1 - \hat{\lambda})}{b + d}} \right\} \quad (4-19)$$

Estimación de los parámetros de desempeño mediante el método de estimación por remuestreo Bootstrap

Para la aplicación del método de re-muestreo de Bootstrap en la estimación de los parámetros de desempeño en cada uno de los algoritmos clasificadores, se tomo como base el archivo de datos inicial con los 324 individuos. El algoritmo Bootstrap se aplica de la siguiente forma:

1. Asignar a cada individuo la probabilidad $P(D = 1|B_j)$ correspondiente a la B_j combinación, independientemente si el individuo esta enfermo o no.
2. Clasificar de acuerdo al punto de corte δ a cada uno de los individuos.
3. Extraer una muestra con remplazo de 324 individuos.
4. Realizar los conteos para construir la tabla **4-3** de la siguiente forma: .
 - **a:** Sumar la cantidad de individuos para los cuales se cumple que $P(D = 1|B_j) \geq \delta$ y son clasificados como dengue por el Gold Standard.
 - **b:** Sumar la cantidad de individuos para los cuales se cumple que $P(D = 1|B_j) \geq \delta$ y son clasificados como no dengue por el Gold Standard.
 - **c:** Sumar la cantidad de individuos para los cuales se cumple que $P(D = 1|B_j) < \delta$ y son clasificados como dengue por el Gold Standard.
 - **d:** Sumar la cantidad de individuos para los cuales se cumple que $P(D = 1|B_j) < \delta$ y son clasificados como no dengue por el Gold Standard.
5. Calcular los parámetros de desempeño por máxima verosimilitud tal y como se especifica en 4-16 y 4-18.
6. Repetir de forma independiente M veces los pasos 1, 2 y 3.
7. Calcular el estimador Booststrap como el valor promedio y su intervalo de confianza como los percentiles 2,5 y 97,5.

Los pasos anteriores se realizaron para cada prototipo y el total de repeticiones fue $M = 1000$ veces.

Estimación Bayesiana

Uno de los alcances de este trabajo radica en comparar las estimaciones de los parámetros de desempeño obtenidas por la inferencia clásica y por la inferencia bayesiana. Así pues, dando continuidad, para esta última se van a considerar dos escenarios de estimación, el primero será asumiendo que para cada algoritmo no se tiene información previa sobre su sensibilidad y especificidad y por ello se asignaran distribuciones a priori

no informativas de Laplace, en el otro escenario se considera que hay información externa sobre el posible valor que pueden tomar estos parámetros (sensibilidad y especificidad) y por ello se consideran distribuciones a priori del tipo informativa.

Dado que los parámetros de interés tienen un espacio paramétrico acotado en $(0,1)$ se asumirán como distribuciones a priori $\pi(\theta)$ distribuciones $Beta(\alpha, \beta)$ para cada uno de los parámetros a estimar.

Estimación de parámetros asumiendo distribuciones a priori no informativas: Como distribución a priori no informativa se va a asumir la distribución no informativa de Laplace, esta es $Beta(\alpha = 1, \beta = 1)$. Así pues, las distribuciones posteriores para cada parámetro son:

$$\begin{aligned}\beta|a &\sim Beta(\alpha^* = a + 1, \beta^* = (a + c) - a + 1) \\ \lambda|d &\sim Beta(\alpha^* = d + 1, \beta^* = (b + d) - d + 1)\end{aligned}$$

Donde α^* y β^* denotan los parámetros de la distribución posterior, a es el número de individuos clasificados como positivo con el Gold Standard y el algoritmo, y d es la cantidad de individuos clasificados como dengue negativo por el Gold Standard y el algoritmo.

Adicionalmente, se presentan sus respectivas regiones de Credibilidad:

$$\begin{aligned}RC_\beta &= [Beta(\alpha^* = a + 1, \beta^* = (a + c) - a + 1, \gamma/2); \\ &\quad Beta(\alpha^* = a + 1, \beta^* = (a + c) - a + 1, 1 - \gamma/2)]\end{aligned}$$

$$\begin{aligned}RC_\lambda &= [Beta(\alpha^* = d + 1, \beta^* = (b + d) - d + 1, \gamma/2); \\ &\quad Beta(\alpha^* = d + 1, \beta^* = (b + d) - d + 1, 1 - \gamma/2)]\end{aligned}$$

Donde $Beta(\alpha^*, \beta^*, \gamma)$ denota el cuantil γ de una distribución $Beta(\alpha^*, \beta^*)$ con probabilidad $1 - \gamma$.

Estimación de parámetros asumiendo distribuciones a priori informativas: Puesto que se cuenta con información externa de estimaciones de los parámetros de desempeño de algoritmos que han sido construidos considerando variables similares a las propuestas en este trabajo, se procedió a elicitar los hiperparámetros de la distribución a priori para cada prototipo de acuerdo a la siguiente metodología:

1. Se seleccionaron los documentos Méndez (2017), Caicedo-Borrero et al. (2020) y Quiñones & Suarez (2019) ya que éstos contienen información acerca de estimaciones de los parámetros de desempeño de los algoritmos considerados en este trabajo.

2. Se construyó un vector de valores para cada parámetro de desempeño donde se incluía las estimaciones halladas por cada autor.
3. Para cada vector (sensibilidad y especificidad) se calculó su media y varianza.
4. Se obtuvieron los hiperparámetros de la distribución apriori $Beta(\alpha, \beta)$ usando el método de los momentos, esto es, igualar su media y varianza a los valores calculados en el paso anterior.
5. Una vez calculados los hiperparámetros de la distribución Beta a priori, se procedió a agregar la información de los datos para la construcción de las distribuciones posteriores de acuerdo a la ecuación 3-10.

Los valores elicitados se encuentran en el Anexo A y los valores asociados al procedimiento descrito anteriormente se encuentran en la Tabla **4-4**.

Distribución posterior:

$$\begin{aligned}\beta|a &\sim Beta(\alpha^* = a + \alpha, \beta^* = (a + c) - a + \beta) \\ \lambda|d &\sim Beta(\alpha^* = d + \alpha, \beta^* = (b + d) - d + \beta)\end{aligned}$$

Donde α y β son los hiperparámetros de la distribución Beta apriori informativa (ver Tabla **4-4**), y α^* y β^* son los parámetros de la distribución Beta posterior.

Región de Credibilidad:

$$\begin{aligned}RC(\beta) &= [Beta(\alpha^* = a + \alpha, \beta^* = (a + c) - a + \beta, \gamma/2); \\ &\quad Beta(\alpha^* = a + \alpha, \beta^* = (a + c) - a + \beta, 1 - \gamma/2)]\end{aligned}\tag{4-20}$$

$$\begin{aligned}RC(\lambda) &= [Beta(\alpha^* = d + \alpha, \beta^* = (b + d) - d + \beta, \gamma/2); \\ &\quad Beta(\alpha^* = d + \alpha, \beta^* = (b + d) - d + \beta, 1 - \gamma/2)]\end{aligned}\tag{4-21}$$

Donde $Beta(\alpha^*, \beta^*, \gamma)$ denota el cuantil γ de una distribución $Beta(\alpha^*, \beta^*)$ con una probabilidad $1 - \gamma$.

Prototipo	Parámetro	Media	Varianza	Hiperparámetros	
				α	β
Prototipo I	Sensibilidad	0,675	0,009	15,846	7,635
	Especificidad	0,604	0,027	4,692	3,077
Prototipo II	Sensibilidad	0,642	0,035	3,545	1,981
	Especificidad	0,728	0,021	6,248	2,331
Prototipo III	Sensibilidad	0,553	0,060	1,719	1,391
	Especificidad	0,805	0,011	10,785	2,618
Prototipo IV	Sensibilidad	0,593	0,085	1,091	0,749
	Especificidad	0,898	0,004	18,515	2,092
Prototipo V	Sensibilidad	0,695	0,004	33,722	14,817
	Especificidad	0,695	0,001	126,673	55,590
Prototipo VI	Sensibilidad	0,834	0,002	52,219	10,431
	Especificidad	0,698	0,048	2,351	1,017

Tabla 4-4.: Valores elicitados e hiperparámetros

4.6. Dos pruebas para diagnóstico bajo diseño de cohorte

Al agregar las variables del hemograma a los algoritmos propuestos por Méndez (2017) se pretende evaluar como es su comportamiento conjunto en la clasificación de pacientes con sospecha de dengue. Sin embargo, también es de interés evaluar como se comportan estos algoritmos (Méndez (2017) y hemograma) cuando su aplicación no es en conjunto, es decir, cuando a un paciente se le aplica de forma individual cada algoritmo de Méndez (2017) y los algoritmos del hemograma. Evaluar estas pruebas por separado se fundamenta en el hecho de que en la practica médica se suele considerar varias pruebas que ayuden a confirmar o descartar un diagnóstico, pues se podría llegar a un diagnostico más certero.

Tal y como se mencionó anteriormente, los test diagnósticos a considerar en este modelo serán: Los cuatro algoritmos propuestos en el trabajo de Méndez (2017) y los dos algoritmos asociados al hemograma propuestos en este trabajo (Prototipos V y VI), así pues se tendrán en total 8 pares de algoritmos. Además, para la estimación de los parámetros de desempeño se van a considerar dos escenarios, el primero asumirá que los resultados entre los dos test son independientes, y en el segundo se asumirá dependencia puesto que las dos pruebas son aplicadas al mismo individuo.

4.6.1. Especificación del Modelo

Eventos

M_i = El individuo tiene un resultado positivo en el prototipo i de Méndez (2017), donde $i = 1, \dots, 4$

L_j = El individuo tiene un resultado positivo en el algoritmo j creado a partir de los parámetros del hemograma, donde $j = 1, 2$

GS = El individuo tiene un resultado positivo por el Gold Standard

Variables aleatorias

A cada evento M_i , L_j y GS se le asigna una variable aleatoria que cuantifique el respectivo resultado

$$T_i = \begin{cases} 1 & ; \text{Ocurre } M_i \\ 0 & ; \text{No ocurre } M_i \end{cases} \quad Y_j = \begin{cases} 1 & ; \text{Ocurre } L_j \\ 0 & ; \text{No ocurre } L_j \end{cases} \quad D = \begin{cases} 1 & ; \text{Ocurre GS} \\ 0 & ; \text{No ocurre GS} \end{cases}$$

donde T_i y Y_j se distribuyen $Bernulli(\pi)$.

Los parámetros de desempeño de los distintos algoritmos de clasificación se calcularon teniendo en cuenta la clasificación de la Tabla 4-5.

	GS Positivo			GS Negativo		
	Positivo ($Y_j = 1$)	Negativo ($Y_j = 0$)	Total	Positivo ($Y_j = 1$)	Negativo ($Y_j = 0$)	Total
Positivo ($T_i = 1$)	r_1	r_2	$r_1 + r_2$	r_5	r_6	$r_5 + r_6$
Negativo ($T_i = 0$)	r_3	r_4	$r_3 + r_4$	r_7	r_8	$r_7 + r_8$
Total	$r_1 + r_3$	$r_2 + r_4$	n^+	$r_5 + r_7$	$r_6 + r_8$	n^-

Tabla 4-5.: Arreglo de datos en un diseño de cohorte para validación de dos pruebas para diagnóstico.

Donde:

- r_1 = Cantidad de individuos clasificados como dengue según el Gold Standard, que el prototipo i de Méndez (2017) y el prototipo j creado a partir de los parámetros del hemograma clasificaron como dengue.
- r_2 = Cantidad de individuos clasificados como dengue según el Gold Standard, que el prototipo i de Méndez (2017) clasificó como dengue y el prototipo j creado a partir de los parámetros del hemograma clasificó como no dengue.
- r_3 = Cantidad de individuos clasificados como dengue según el Gold Standard, que el prototipo i de Méndez (2017) clasificó como no dengue y el prototipo j creado a partir de los parámetros del hemograma clasificó como dengue.
- r_4 = Cantidad de individuos clasificados como dengue según el Gold Standard, que el prototipo i de Méndez (2017) y el prototipo j creado a partir de los parámetros del hemograma clasificaron como no dengue.

- r_5 = Cantidad de individuos clasificados como no dengue según el Gold Standard, que el prototipo i de Méndez (2017) y el prototipo j creado a partir de los parámetros del hemograma clasificaron como dengue.
- r_6 = Cantidad de individuos clasificados como no dengue según el Gold Standard, que el prototipo i de Méndez (2017) clasificó como dengue y el prototipo j creado a partir de los parámetros del hemograma clasificó como no dengue.
- r_7 = Cantidad de individuos clasificados como no dengue según el Gold Standard, que el prototipo i de Méndez (2017) clasificó como no dengue y el prototipo j creado a partir de los parámetros del hemograma clasificó como dengue.
- r_8 = Cantidad de individuos clasificados como no dengue según el Gold Standard, que el prototipo i de Méndez (2017) y el prototipo j creado a partir de los parámetros del hemograma clasificaron como no dengue.

Por lo tanto, la función de verosimilitud asociada al modelo en el proceso de inferencia estadística quedaría compuesta por ocho ecuaciones con igual número de incógnitas (parámetros θ) definidas por las probabilidades conjuntas de observar los resultados definidos para las celdas de las Tabla **4-5**. Los ocho parámetros θ son:

$$\theta_1 = P(T = 1, Y = 1, D = 1)$$

$$\theta_2 = P(T = 1, Y = 0, D = 1)$$

$$\theta_3 = P(T = 0, Y = 1, D = 1)$$

$$\theta_4 = P(T = 0, Y = 0, D = 1)$$

$$\theta_5 = P(T = 1, Y = 1, D = 0)$$

$$\theta_6 = P(T = 1, Y = 0, D = 0)$$

$$\theta_7 = P(T = 0, Y = 1, D = 0)$$

$$\theta_8 = P(T = 0, Y = 0, D = 0)$$

Entonces, se tiene el vector de parámetros $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8\}$ y

$$X_1 = \begin{cases} 1; & \text{Ocurre } T = 1, Y = 1, D = 1 \\ 0; & \text{En otro caso} \end{cases} \quad (4-22)$$

$$X_2 = \begin{cases} 1; & \text{Ocurre } T = 1, Y = 0, D = 1 \\ 0; & \text{En otro caso} \end{cases} \quad (4-23)$$

$$X_3 = \begin{cases} 1; & \text{Ocurre } T = 0, Y = 1, D = 1 \\ 0; & \text{En otro caso} \end{cases} \quad (4-24)$$

$$X_4 = \begin{cases} 1; & \text{Ocurre } T = 0, Y = 0, D = 1 \\ 0; & \text{En otro caso} \end{cases} \quad (4-25)$$

$$X_5 = \begin{cases} 1; & \text{Ocurre } T = 1, Y = 1, D = 0 \\ 0; & \text{En otro caso} \end{cases} \quad (4-26)$$

$$X_6 = \begin{cases} 1; & \text{Ocurre } T = 1, Y = 0, D = 0 \\ 0; & \text{En otro caso} \end{cases} \quad (4-27)$$

$$X_7 = \begin{cases} 1; & \text{Ocurre } T = 0, Y = 1, D = 0 \\ 0; & \text{En otro caso} \end{cases} \quad (4-28)$$

$$X_8 = \begin{cases} 1; & \text{Ocurre } T = 0, Y = 0, D = 0 \\ 0; & \text{En otro caso} \end{cases} \quad (4-29)$$

Donde $X_k \sim \text{Bernoulli}(\theta_k)$ con $k = \{1, \dots, 8\}$

Entonces, se tiene que:

$$\begin{aligned} r_1 &= \sum_{m=1}^n X_{1m} & r_2 &= \sum_{m=1}^n X_{2m} & r_3 &= \sum_{m=1}^n X_{3m} ; \\ r_4 &= \sum_{m=1}^n X_{4m} & r_5 &= \sum_{m=1}^n X_{5m} & r_6 &= \sum_{m=1}^n X_{6m} ; \\ r_7 &= \sum_{m=1}^n X_{7m} & r_8 &= \sum_{m=1}^n X_{8m} \end{aligned}$$

Entonces, $Z = (R_1 = r_1, \dots, R_8 = r_8) \sim \text{Multinomial}(n, \Theta)$ con función de masa de probabilidad:

$$P(R_1 = r_1, \dots, R_8 = r_8 | n, \theta_1, \dots, \theta_8) = \frac{n!}{r_1! \dots r_8!} \theta^{r_1} \dots \theta^{r_8} \quad (4-30)$$

Sujeto a:

$$\theta_k \in (0, 1) \quad ; \quad \sum_{k=1}^8 \theta_k = 1 \quad \text{y} \quad r_1 + \dots + r_8 = n$$

Con función de máxima verosimilitud:

$$L(\theta_1, \dots, \theta_8 | r_1, \dots, r_8) \propto \prod_{k=1}^8 \theta_k^{r_k} \quad (4-31)$$

4.6.2. Escenario 1: Estimación de parámetros de desempeño asumiendo independencia condicional entre las dos pruebas

Con el fin de dejarlo expresado en términos de los parámetros de interés (prevalencia, sensibilidades y especificidades) se hace una reparametrización del modelo

- p : Prevalencia
 $p = P(D = 1)$
- S_1 : Sensibilidad del Test 1
 $S_1 = P(T = 1|D = 1)$
- S_2 : Sensibilidad del Test 2
 $S_2 = P(Y = 1|D = 1)$
- E_1 : Especificidad del Test 1
 $E_1 = P(T = 0|D = 0)$
- E_2 : Especificidad del Test 2
 $E_2 = P(Y = 0|D = 0)$

Bajo el supuesto de independencia condicional entre los resultados de clasificación, la reparametrización del modelo se define:

$$\begin{aligned}\theta_1 &= P(D = 1) \cdot P(T = 1, Y = 1|D = 1) = P(D = 1) \cdot P(T = 1|Y = 1) \cdot P(Y = 1|D = 1) \\ &= p \cdot S_1 \cdot S_2\end{aligned}$$

$$\begin{aligned}\theta_2 &= P(D = 1) \cdot P(T = 1, Y = 0|D = 1) = P(D = 1) \cdot P(T = 1|D = 1) \cdot P(Y = 0|D = 1) \\ &= p \cdot S_1 \cdot (1 - S_2)\end{aligned}$$

$$\begin{aligned}\theta_3 &= P(D = 1) \cdot P(T = 0, Y = 1|D = 1) = P(D = 1) \cdot P(T = 0|D = 1) \cdot P(Y = 1|D = 1) \\ &= p \cdot (1 - S_1) \cdot S_2\end{aligned}$$

$$\begin{aligned}\theta_4 &= P(D = 1) \cdot P(T = 0, Y = 0|D = 1) = P(D = 1) \cdot P(T = 0|D = 1) \cdot P(Y = 0|D = 1) \\ &= p \cdot (1 - S_1) \cdot (1 - S_2)\end{aligned}$$

$$\begin{aligned}\theta_5 &= P(D = 0) \cdot P(T = 1, Y = 1|D = 0) = P(D = 0) \cdot P(T = 1|D = 0) \cdot P(Y = 1|D = 0) \\ &= (1 - p) \cdot (1 - E_1) \cdot (1 - E_2)\end{aligned}$$

$$\begin{aligned}\theta_6 &= P(D = 0) \cdot P(T = 1, Y = 0 | D = 0) = P(D = 0) \cdot P(T = 1 | D = 0) \cdot P(Y = 0 | D = 0) \\ &= (1 - p) \cdot (1 - E_1) \cdot E_2\end{aligned}$$

$$\begin{aligned}\theta_7 &= P(D = 0) \cdot P(T = 0, Y = 1 | D = 0) = P(D = 0) \cdot P(T = 0 | D = 0) \cdot P(Y = 1 | D = 0) \\ &= (1 - p) \cdot E_1 \cdot (1 - E_2)\end{aligned}$$

$$\begin{aligned}\theta_8 &= P(D = 0) \cdot P(T = 0, Y = 0 | D = 0) = P(D = 0) \cdot P(T = 0 | D = 0) \cdot P(Y = 0 | D = 0) \\ &= (1 - p) \cdot E_1 \cdot E_2\end{aligned}$$

Después de reparametrizar las probabilidades de éxito asociadas a las diferentes celdas en la Tabla 4-5, se tiene que, la función de verosimilitud toma la siguiente forma:

$$\begin{aligned}L(p, S_1, S_2, E_1, E_2 | r_1, \dots, r_8) &\propto p^{(r_1+r_2+r_3+r_4)} \cdot (1 - p)^{(r_5+r_6+r_7+r_8)} \cdot S_1^{(r_1+r_2)} \cdot (1 - S_1)^{(r_3+r_4)} \\ &\quad \cdot S_2^{(r_1+r_3)} \cdot (1 - S_2)^{(r_2+r_4)} \cdot E_1^{(r_7+r_8)} \cdot E_2^{(r_6+r_8)} \\ &\quad \cdot (1 - E_1)^{(r_5+r_6)} \cdot (1 - E_2)^{(r_5+r_7)}\end{aligned}\tag{4-32}$$

Estimación usando método de máxima verosimilitud

Aplicando el procedimiento de estimación por máxima verosimilitud, se obtienen las siguientes estimaciones puntuales y por regiones considerando un nivel de confianza $1 - \gamma$:

1. Sensibilidad:

$$\hat{S}_1 = \frac{r_1 + r_2}{\sum_{k=1}^4 r_k} = \frac{r_1 + r_2}{n^+}\tag{4-33}$$

$$\hat{S}_2 = \frac{r_1 + r_3}{\sum_{k=1}^4 r_k} = \frac{r_1 + r_3}{n^+}\tag{4-34}$$

2. Especificidad:

$$\hat{E}_1 = \frac{r_7 + r_8}{\sum_{k=5}^8 r_k} = \frac{r_7 + r_8}{n^-}\tag{4-35}$$

$$\hat{E}_2 = \frac{r_6 + r_8}{\sum_{k=5}^8 r_k} = \frac{r_6 + r_8}{n^-}\tag{4-36}$$

3. Intervalos de Confianza:

$$IC(S_l)_{1-\gamma} = \left\{ \hat{S}_l \pm Z_{1-\gamma} \sqrt{\frac{\hat{S}_l(1 - \hat{S}_l)}{n^+}} \right\}; l = 1, 2 \quad (4-37)$$

$$IC(E_l)_{1-\gamma} = \left\{ \hat{E}_l \pm Z_{1-\gamma} \sqrt{\frac{\hat{E}_l(1 - \hat{E}_l)}{n^-}} \right\}; l = 1, 2 \quad (4-38)$$

Estimación de los parámetros de desempeño mediante el método de estimación por remuestreo Bootstrap

Para la aplicación del método de re-muestreo de Bootstrap en la estimación de los parámetros de desempeño en cada uno de los algoritmos clasificadores, se tomo como base el archivo de datos inicial con los 324 individuos. El algoritmo Bootstrap se aplica de la siguiente forma:

1. Asignar a cada individuo la probabilidad $P(D = 1|B_j)$ correspondiente a la B_j combinación, independientemente si el individuo esta enfermo o no.
2. Clasificar de acuerdo al punto de corte δ a cada uno de los individuos.

Nota: Para los algoritmos de Méndez (2017) se utilizaron los puntos de corte encontrados por él (ver Anexo A), y la clasificación de los algoritmos basados en los parámetros del hemograma se realizo de acuerdo a los puntos de corte encontrados en la subsección 4.4.5.

3. Extraer una muestra con remplazo de 324 individuos.
4. Realizar los conteos para construir la tabla **4-5** de la siguiente forma:
 - r_1 = Sumar los individuos clasificados como dengue según el Gold Standard, que el prototipo i de Méndez (2017) y el prototipo j creado a partir de los parámetros del hemograma clasificaron como dengue.
 - r_2 = Sumar los individuos clasificados como dengue según el Gold Standard, que el prototipo i de Méndez (2017) clasificó como dengue y el prototipo j creado a partir de los parámetros del hemograma clasificó como no dengue.
 - r_3 = Sumar los individuos clasificados como dengue según el Gold Standard, que el prototipo i de Méndez (2017) clasificó como no dengue y el prototipo j creado a partir de los parámetros del hemograma clasificó como dengue.
 - r_4 = Sumar los individuos clasificados como dengue según el Gold Standard, que el prototipo i de Méndez (2017) y el prototipo j creado a partir de los parámetros del hemograma clasificaron como no dengue.

- r_5 = Sumar los individuos clasificados como no dengue según el Gold Standard, que el prototipo i de Méndez (2017) y el prototipo j creado a partir de los parámetros del hemograma clasificaron como dengue.
 - r_6 = Sumar los individuos clasificados como no dengue según el Gold Standard, que el prototipo i de Méndez (2017) clasificó como dengue y el prototipo j creado a partir de los parámetros del hemograma clasificó como no dengue.
 - r_7 = Sumar los individuos clasificados como no dengue según el Gold Standard, que el prototipo i de Méndez (2017) clasificó como no dengue y el prototipo j creado a partir de los parámetros del hemograma clasificó como dengue.
 - r_8 = Sumar los individuos clasificados como no dengue según el Gold Standard, que el prototipo i de Méndez (2017) y el prototipo j creado a partir de los parámetros del hemograma clasificaron como no dengue.
5. Se calcularon los parámetros de desempeño por máxima verosimilitud tal y como se especifica en las ecuaciones 4-33 a 4-36.
 6. Repetir de forma independiente M veces los pasos 1, 2 y 3.
 7. Calcular el estimador Booststrap como el valor promedio y su intervalo de confianza como los percentiles 2,5 y 97,5.

Los pasos anteriores se realizaron para cada par de prototipos y el total de repeticiones fue $M = 1000$ veces.

Estimación bayesiana

Dado que los parámetros de interés tienen un espacio paramétrico acotado en $(0,1)$ se asumirán como distribuciones a priori $\pi(\theta)$ distribuciones $Beta(\alpha, \beta)$.

Estimación de parámetros asumiendo distribuciones a priori no informativas: Para la obtención de los hiperparámetros de la distribución a priori, se asumirá la distribución no informativa obtenida usando el método propuesto por Laplace, esta es $Beta(\alpha = 1, \beta = 1)$. Así pues, las distribuciones posteriores para cada parámetro son:

$$S_1|r \sim Beta(\alpha^* = r_1 + r_2 + 1, \beta^* = r_3 + r_4 + 1)$$

$$S_2|r \sim Beta(\alpha^* = r_1 + r_3 + 1, \beta^* = r_2 + r_4 + 1)$$

$$E_1|r \sim Beta(\alpha^* = r_7 + r_8 + 1, \beta^* = r_5 + r_6 + 1)$$

$$E_2|r \sim \text{Beta}(\alpha^* = r_6 + r_8 + 1, \beta^* = r_5 + r_7 + 1)$$

Adicionalmente, se presentan sus respectivas Regiones de Credibilidad considerando una probabilidad de $1 - \gamma$.

$$\begin{aligned} RC(S_1) = & [\text{Beta}(\alpha^* = r_1 + r_2 + 1, \beta^* = r_3 + r_4 + 1, \gamma/2) \\ & ; \text{Beta}(\alpha^* = r_1 + r_2 + 1, \beta^* = r_3 + r_4 + 1, 1 - \gamma/2) \end{aligned} \quad (4-39)$$

$$\begin{aligned} RC(S_2) = & [\text{Beta}(\alpha^* = r_1 + r_3 + 1, \beta^* = r_2 + r_4 + 1, \gamma/2) \\ & ; \text{Beta}(\alpha^* = r_1 + r_3 + 1, \beta^* = r_2 + r_4 + 1, 1 - \gamma/2) \end{aligned} \quad (4-40)$$

$$\begin{aligned} RC(E_1) = & [\text{Beta}(\alpha^* = r_7 + r_8 + 1, \beta^* = r_5 + r_6 + 1, \gamma/2) \\ & ; \text{Beta}(\alpha^* = r_7 + r_8 + 1, \beta^* = r_5 + r_6 + 1, 1 - \gamma/2) \end{aligned} \quad (4-41)$$

$$\begin{aligned} RC(E_2) = & [\text{Beta}(\alpha^* = r_6 + r_8 + 1, \beta^* = r_5 + r_7 + 1, \gamma/2) \\ & ; \text{Beta}(\alpha^* = r_6 + r_8 + 1, \beta^* = r_5 + r_7 + 1, 1 - \gamma/2) \end{aligned} \quad (4-42)$$

Estimación de parámetros asumiendo distribuciones a priori informativas: Para este caso de estimación, el procedimiento para la elicitación de los hiperparámetros de la distribución a priori fue el mismo al usado en la Subsección 4.5.2, por tanto, los hiperparámetros tomados para la construcción de la distribución posterior se encuentran en la Tabla 4-4. Cabe resaltar que para los algoritmos de Méndez (2017) se usaron los mismos valores que para los prototipos I,II,III y IV de este trabajo. A continuación se presentan las distribuciones posteriores de los parámetros de interés con sus respectivas regiones de credibilidad.

Distribución posterior:

$$S_1|r \sim \text{Beta}(\alpha^* = r_1 + r_2 + \alpha, \beta^* = r_3 + r_4 + \beta)$$

$$S_2|r \sim \text{Beta}(\alpha^* = r_1 + r_3 + \alpha, \beta^* = r_2 + r_4 + \beta)$$

$$E_1|r \sim \text{Beta}(\alpha^* = r_7 + r_8 + \alpha, \beta^* = r_5 + r_6 + \beta)$$

$$E_2|r \sim \text{Beta}(\alpha^* = r_6 + r_8 + \alpha, \beta^* = r_5 + r_7 + \beta)$$

Donde α y β son los hiperparámetros de la distribución Beta a priori informativa y se encuentran consignados en la Tabla 4-4.

Adicionalmente, se presentan sus respectivas Regiones de Credibilidad considerando una probabilidad de $1 - \gamma$.

$$RC(S_1) = [Beta(\alpha^* = r_1 + r_2 + \alpha, \beta^* = r_3 + r_4 + \beta, \gamma/2) ; Beta(\alpha^* = r_1 + r_2 + \alpha, \beta^* = r_3 + r_4 + \beta, 1 - \gamma/2)] \quad (4-43)$$

$$RC(S_2) = [Beta(\alpha^* = r_1 + r_3 + \alpha, \beta^* = r_2 + r_4 + \beta, \gamma/2) ; Beta(\alpha^* = r_1 + r_3 + \alpha, \beta^* = r_2 + r_4 + \beta, 1 - \gamma/2)] \quad (4-44)$$

$$RC(E_1) = [Beta(\alpha^* = r_7 + r_8 + \alpha, \beta^* = r_5 + r_6 + \beta, \gamma/2) ; Beta(\alpha^* = r_7 + r_8 + \alpha, \beta^* = r_5 + r_6 + \beta, 1 - \gamma/2)] \quad (4-45)$$

$$RC(E_2) = [Beta(\alpha^* = r_6 + r_8 + \alpha, \beta^* = r_5 + r_7 + \beta, \gamma/2) ; Beta(\alpha^* = r_6 + r_8 + \alpha, \beta^* = r_5 + r_7 + \beta, 1 - \gamma/2)] \quad (4-46)$$

Donde $Beta(\alpha^*, \beta^*, \gamma)$ denota el cuantil γ de una distribución $Beta(\alpha^*, \beta^*)$ con una probabilidad $1 - \gamma$.

4.6.3. Escenario 2: Estimación de parámetros de desempeño asumiendo dependencia condicional entre las dos pruebas para diagnostico en evaluación

Cuando a un individuo se le aplican diferentes pruebas diagnosticas ocurre que el resultado de éstas está correlacionado por el hecho de ser aplicadas al mismo individuo, por tanto es necesario integrar este parámetro de correlación en el modelo. Si solo se tiene en cuenta la respuesta binaria de las pruebas de diagnostico, es posible suponer entonces que la dependencia entre ellas se puede modelar utilizando como parámetro de interés la covarianza entre dos variables aleatorias con Distribución Bernoulli, cuyo efecto es variable según el resultado del patrón oro, es decir, se puede asumir un valor para la covarianza en la población de pacientes enfermos ψ_D y otro en la población de individuos no enfermos ψ_{ND} .

Sean T y Y las variables aleatorias binarias que identifican los resultados de los tests diagnósticos. La covarianza entre T y Y para la población de enfermos se define como:

$$\psi_D = E(T, Y | D = 1) - E(T | D = 1)E(Y | D = 1)$$

y para la población de no enfermos como:

$$\psi_{ND} = E(T, Y|D = 0) - E(T|D = 0)E(Y|D = 0)$$

Donde E denota el valor esperado.

Al desarrollar el procedimiento para ψ_D , se obtiene:

- $E(T|D = 1) = 1 \cdot P(T = 1|D = 1) + 0 \cdot P(T = 0|D = 1) = S_1$
- $E(Y|D = 1) = 1 \cdot P(Y = 1|D = 1) + 0 \cdot P(Y = 0|D = 1) = S_2$
- $E(TY|D = 1) = 1 \cdot P(T = 1, Y = 1|D = 1) + 0 \cdot P(T = 0, Y = 1|D = 1) + 0 \cdot P(T = 1, Y = 0|D = 1) + 0 \cdot P(T = 0, Y = 0|D = 1) = P(T = 1, Y = 1|D = 1)$

Por lo tanto:

$$\psi_D = P(T = 1, Y = 1|D = 1) - S_1 \cdot S_2 \quad (4-47)$$

El mismo procedimiento se repite para el grupo de individuos no enfermos y se obtiene:

$$\psi_{ND} = P(T = 1, Y = 1|D = 0) - ((1 - E_1) \cdot (1 - E_2)) \quad (4-48)$$

Con el fin de dejarlo expresado en términos de los parámetros de interés (prevalencia, sensibilidades, especificidades y covarianzas) y asumiendo dependencia entre los resultados de clasificación, la reparametrización del modelo se define:

$$\theta_1 = P(D = 1) \cdot P(T = 1, Y = 1|D = 1) = p \cdot (S_1 \cdot S_2 + \psi_D)$$

$$\theta_2 = P(D = 1) \cdot P(T = 1, Y = 0|D = 1) = p \cdot [S_1 \cdot (1 - S_2) - \psi_D]$$

$$\theta_3 = P(D = 1) \cdot P(T = 0, Y = 1|D = 1) = p \cdot [(1 - S_1) \cdot S_2 - \psi_D]$$

$$\theta_4 = P(D = 1) \cdot P(T = 0, Y = 0|D = 1) = p \cdot [(1 - S_1) \cdot (1 - S_2) + \psi_D]$$

$$\theta_5 = P(D = 0) \cdot P(T = 1, Y = 1|D = 0) = (1 - p) \cdot [(1 - E_1) \cdot (1 - E_2) + \psi_{ND}]$$

$$\theta_6 = P(D = 0) \cdot P(T = 1, Y = 0|D = 0) = (1 - p) \cdot [(1 - E_1) \cdot E_2 - \psi_{ND}]$$

$$\theta_7 = P(D = 0) \cdot P(T = 0, Y = 1|D = 0) = (1 - p) \cdot [E_1 \cdot (1 - E_2) - \psi_{ND}]$$

$$\theta_8 = P(D = 0) \cdot P(T = 0, Y = 0|D = 0) = (1 - p) \cdot [E_1 \cdot E_2 + \psi_{ND}]$$

La función de verosimilitud asociada al modelo para estimación puede ser escrita como:

$$\begin{aligned}
 L(p, S_1, S_2, E_1, E_2, \psi_D, \psi_{ND} | r_1, \dots, r_8) \propto & p^{(r_1+r_2+r_3+r_4)} \cdot (1-p)^{(r_5+r_6+r_7+r_8)} \cdot [S_1 S_2 + \psi_D]^{r_1} \\
 & \cdot [S_1(1-S_2) - \psi_D]^{r_2} \cdot [(1-S_1)S_2 - \psi_D]^{r_3} \\
 & \cdot [(1-S_1)(1-S_2) + \psi_D]^{r_4} \cdot [(1-E_1)(1-E_2) + \psi_{ND}]^{r_5} \\
 & \cdot [(1-E_1)E_2 - \psi_{ND}]^{r_6} \cdot [E_1(1-E_2) - \psi_{ND}]^{r_7} \\
 & \cdot [E_1 E_2 + \psi_{ND}]^{r_8}
 \end{aligned}
 \tag{4-49}$$

La ecuación 4-49 tiene la forma de una función de verosimilitud de una distribución multinomial. Los parámetros a estimar en este escenario son siete (dos sensibilidades, dos especificidades, dos covarianzas y la prevalencia) sin embargo la Tabla 4-5 solo cuenta con seis grados de libertad lo que implica un problema de falta de identificabilidad en el modelo de estimación.

Desde el punto de vista bayesiano autores como Dendukuri & Joseph (2001) sugieren que una forma de solucionar el problema de falta de identificabilidad es estimar un conjunto de parámetros igual o mayor a la cantidad de grados de libertad faltantes asociándoles distribuciones a priori informativas. Estos mismos autores afirman que este proceder está limitado por el hecho de que las estimaciones obtenidas posteriormente estarán fuertemente influenciado por las a priori utilizadas.

Las estimaciones de este escenario solo se realizaron bajo el enfoque bayesiano asumiendo distribuciones a priori informativas.

Estimación Bayesiana

El procedimiento usado para obtener las estimaciones bajo el enfoque bayesianos es el siguiente:

- Para la prevalencia (p) se asignó una distribución a priori no informativa de Laplace $Beta(\alpha_p = 1, \beta_p = 1)$.
- Para las sensibilidades y especificidades se asignaron distribuciones a priori Beta informativas. Estas serán las mismas de la tabla 4-4
- En cuanto a las covarianzas, se tuvo en cuenta lo siguiente:

En el procedimiento de estimación desarrollado por Dendukuri & Joseph (2001) se considera que la covarianza tiene los siguientes límites.

$$(S_1 - 1)(1 - S_2) \leq \psi_D \leq \min(S_1 S_2) - (S_1 S_2)$$

$$(E_1 - 1)(1 - E_2) \leq \psi_{ND} \leq \min(E_1 E_2) - (E_1 E_2)$$

Aunque los límites inferiores para ψ_D y ψ_{ND} siempre son negativos en este trabajo solo se consideró estimar la dependencia positiva, esto basado en que ambos algoritmos de clasificación apuntan a medir lo mismo, en este caso si el individuo es o no positivo para dengue, por tanto los nuevos límites inferiores fueron fijados como cero.

$$0 \leq \psi_D \leq us$$

$$0 \leq \psi_{ND} \leq uc$$

donde $us = \min(S_1 S_2) - (S_1 S_2)$ y $uc = \min(E_1 E_2) - (E_1 E_2)$

Para los parámetros de covarianza, se considera que estos siguen una distribución beta:

$$\psi_D \sim \text{Beta}(\alpha_{\psi_D}, \beta_{\psi_D})$$

$$\psi_{ND} \sim \text{Beta}(\alpha_{\psi_{ND}}, \beta_{\psi_{ND}})$$

Al combinar la función de verosimilitud de la ecuación 4-49 con las distribuciones previas, se obtiene la siguiente distribución posterior.

$$\begin{aligned} p, S_1, S_2, E_1, E_2 | r_1 \cdots r_8 \propto & L(p, S_1, S_2, E_1, E_2, \psi_D, \psi_{ND} | r_1, \cdots, r_8) \cdot p^{\alpha_p - 1} \cdot (1 - p)^{\beta_p - 1} \\ & \cdot S_1^{\alpha_{S_1} - 1} \cdot (1 - S_1)^{\beta_{S_1} - 1} \cdot S_2^{\alpha_{S_2} - 1} \cdot (1 - S_2)^{\beta_{S_2} - 1} \cdot E_1^{\alpha_{E_1} - 1} \\ & \cdot (1 - E_1)^{\beta_{E_1} - 1} \cdot E_2^{\alpha_{E_2} - 1} \cdot (1 - E_2)^{\beta_{E_2} - 1} \cdot \psi_D^{\alpha_{\psi_D} - 1} \cdot (us - \psi_D)^{\beta_{\psi_D} - 1} \\ & \cdot \psi_{ND}^{\alpha_{\psi_{ND}} - 1} \cdot (uc - \psi_{ND})^{\beta_{\psi_{ND}} - 1} \end{aligned} \quad (4-50)$$

Dada la complejidad del modelo anterior, no es posible obtener las distribuciones marginales para cada parámetro, por ello se utilizan métodos como el algoritmo de Metropolis-Hastings en ambiente Cadenas de Markov y Montecarlo (MCMC) para obtener muestras de la distribución posterior. El vector de parámetros se estimó haciendo uso del Software OpenBugs 3.2.3.

$$U = \begin{cases} 1; & \text{Si ocurre } E^+ \\ 0; & \text{Si ocurre } E^- \end{cases} \quad (4-53)$$

4.7.2. Estimación

Para este escenario, donde se considera un único procedimiento clasificador, se asumió un diseño de cohorte para un test diagnostico (Tabla **4-3**), por esta razón, la estimación de los parámetros de desempeño desde el enfoque clásico, bayesiano y por el método de remuestreo Bootstrap se realizó siguiendo lo expuesto en la sección 4.5.

5. Resultados

En este capítulo se presentan las medidas de desempeño estimadas bajo la inferencia clásica y bayesiana para cada uno de los nuevos algoritmos clasificadores. Además, se consideran los supuestos de independencia y dependencia condicional para la estimación de los parámetros de desempeño de los algoritmos propuestos por Méndez (2017) y los algoritmos V y VI planteados en este trabajo. Por último, se presentan las estimaciones obtenidas de los parámetros de desempeño al considerar un único resultado clasificador después de haber aplicado las pruebas simultáneamente.

El punto de corte usado para clasificar a los individuos fue calculado como el promedio de 1000 puntos de corte que se tenían después de haber estimado un punto de corte para cada matriz de datos simulada, de esta misma manera se obtuvo el área bajo la curva (AUC). Las estimaciones de los parámetros de desempeño se llevaron a cabo considerando tres métodos: clásico, bayesiano y por remuestreo, utilizando los 324 individuos de la base de datos. En el clásico las estimaciones se obtuvieron con los estimadores de máxima verosimilitud, en el enfoque bayesiano las estimaciones se calcularon considerando distribuciones a priori informativas y no informativas para los parámetros, y para el remuestreo el estimador utilizado fue la media de las estimaciones. Este mismo procedimiento fue aplicado para cada uno de los seis prototipos clasificadores.

5.1. Resultados Prototipo I

Este prototipo está formado por las ocho categorías (Tabla 4-2) y los cinco parámetros del hemograma.

Según la Tabla 5-1 para el primer prototipo se encontró un punto de corte (PC) de 0,13, es decir, si a un individuo se le aplica este algoritmo, y su probabilidad predictiva arroja un valor mayor o igual al PC éste será clasificado como dengue positivo. Con respecto a la capacidad predictiva (AUC), se encontró que ésta fue del 78 % (IC 95 %: 77 % – 79 %), esto es, la probabilidad de que este algoritmo pueda distinguir correctamente entre un individuo sano y uno enfermo es del 0,78.

Medida desempeño	Método			
	MV	Boostrap	Bayes Inf	Bayes no Inf
Sensibilidad (Intervalo 95 %)	0,67 (0,48 – 0,86)	0,67 (0,50 – 0,83)	0,67 (0,53 – 0,79)	0,66 (0,49 – 0,82)
Especificidad (Intervalo 95 %)	0,84 (0,79 – 0,88)	0,84 (0,80 – 0,87)	0,83 (0,79 – 0,87)	0,83 (0,79 – 0,87)
Punto de Corte	0,13			
AUC (Intervalo 95 %)	0,78 (0,77 – 0,79)			

Tabla 5-1.: Medidas de desempeño estimadas para el prototipo I

Con respecto a las estimaciones puntuales encontradas, independientemente del método de estimación utilizado se obtuvieron valores iguales o cercanos para cada medida de desempeño, no obstante, solo se evidenció diferencia para los intervalos de confianza (IC) y regiones de credibilidad (RC) en el caso de la sensibilidad, pues la RC correspondiente a la estimación bayesiana informativa resultó ser más angosta (0,53 – 0,79). Cabe resaltar que en todos los casos, los intervalos de la sensibilidad son más amplios en comparación que los intervalos de la especificidad, esto muestra que la sensibilidad es mucho más variable.

5.2. Resultados Prototipo II

Este prototipo está formado por los 12 síntomas (Tabla 4-1) y los cinco parámetros del hemograma.

Medida desempeño	Método			
	MV	Boostrap	Bayes Inf	Bayes no Inf
Sensibilidad (Intervalo 95 %)	0,67 (0,48 – 0,86)	0,67 (0,50 – 0,83)	0,65 (0,48 – 0,82)	0,65 (0,48 – 0,82)
Especificidad (Intervalo 95 %)	0,80 (0,75 – 0,85)	0,80 (0,76 – 0,84)	0,80 (0,76 – 0,84)	0,80 (0,75 – 0,84)
Punto de Corte	0,11			
AUC (Intervalo 95 %)	0,77 (0,76 – 0,77)			

Tabla 5-2.: Medidas de desempeño estimadas para el prototipo II

De acuerdo con los resultado de la Tabla 5-2 un individuo es clasificado como caso positivo para dengue con el prototipo II si la probabilidad predictiva obtenida a partir de la información entregada por el cuadro de signos y síntomas observado en la consulta con el médico es mayor o igual a 0,11. Además, se identificó que su capacidad predictiva (AUC) es

del 77 %. En cuanto a las estimaciones puntuales, se evidencia que la sensibilidad varía entre 0,65 y 0,67, mientras que la especificidad fue la misma en todos los casos (0,8). En todos los métodos de estimación, la especificidad mostró ser más alta que la sensibilidad, esto muestra que este algoritmo tiene mejor desempeño para clasificar correctamente a la población no enferma de dengue.

5.3. Resultados Prototipo III

El prototipo III esta conformado por las ocho categorías de signos y síntomas, y los cinco parámetros del hemograma. Aquí, se asume que la probabilidad $P(D = 1|B_j)$ es una variable aleatoria continua que toma valores entre $[0, 1]$. Desde el punto de vista bayesiano, a esta variable se le puede asignar una distribución a priori que permite modelar su comportamiento, el Teorema de Bayes brinda la posibilidad de combinar esta información sobre el parámetro de interés con la función de verosimilitud asociada a la muestra, con el fin de estimar la probabilidad de tener la enfermedad del dengue en cada una de las posibles combinaciones de categorías de signos y síntomas.

Medida desempeño	Método			
	MV	Boostrap	Bayes Inf	Bayes no Inf
Sensibilidad (Intervalo 95 %)	0,63 (0,43 – 0,82)	0,63 (0,42 – 0,79)	0,62 (0,45 – 0,79)	0,62 (0,43 – 0,79)
Especificidad (Intervalo 95 %)	0,92 (0,89 – 0,95)	0,92 (0,87 – 0,95)	0,92 (0,88 – 0,94)	0,92 (0,88 – 0,96)
Punto de Corte	0,18			
AUC (Intervalo 95 %)	0,85 (0,84 – 0,86)			

Tabla 5-3.: Medidas de desempeño estimadas para el prototipo III

De la tabla 5-3 se infiere que un individuo será clasificado como dengue positivo si su probabilidad predictiva es mayor o igual a 0,18. Además, se identifica que la capacidad clasificatoria cuyo valor fue de (85 %) tuvo un aumento sustancial frente a los dos prototipos evaluados anteriormente.

En relación a las estimaciones de las medidas de desempeño para este prototipo, se evidenció una sensibilidad entre 0,62 y 0,63 donde se sigue observando que al igual que en los prototipos anteriores la sensibilidad presenta intervalos de confianza y credibilidad bastante amplios. Por otro lado, se tiene que la especificidad estimada por todos los métodos fue de 0,92 y sus estimaciones por intervalo son mucho más angostos en comparación con las de la sensibilidad.

5.4. Resultados Prototipo IV

Este prototipo conformado por los 12 signos y síntomas, y las cinco variables del hemograma, al igual que el algoritmo III asume que la probabilidad $P(D = 1|B_j)$ es una variable aleatoria continua y por lo tanto su desarrollo para el calculo de las probabilidades predictivas fue exactamente el mismo que para el prototipo anterior.

La Tabla 5-4 muestra que si a un individuo se le aplicara este algoritmo y su probabilidad fuera igual o mayor a 0,19, éste será clasificado como caso dengue. Además, se obtuvo una capacidad clasificatoria del 92 %. Al estimar los parámetros de desempeño para este algoritmo se puede ver que al igual que en casos anteriores, las estimaciones obtenidas por los diferentes métodos no presentan mucha variación, pues las estimaciones puntuales para la sensibilidad están entre 0,73 y 0,75, y la especificidad en todos los casos fue de 0,91, cabe resaltar que aunque este algoritmo presenta valores relativamente altos en su sensibilidad, esta no deja de ser muy variable, pues se observa intervalos amplios tanto en la estimación clásica y por remuestreo $IC(0,58 - 0,92)$ como en la bayesiana $RC(0,55 - 0,89)$. Este algoritmo muestra algo particular en comparación con los prototipos anteriores, y es que su sensibilidad supera el 70 %.

Medida desempeño	Método			
	MV	Bootstrap	Bayes Inf	Bayes no Inf
Sensibilidad	0,75	0,75	0,74	0,73
(Intervalo 95 %)	(0,58 – 0,92)	(0,58 – 0,92)	(0,55 – 0,89)	(0,55 – 0,88)
Especificidad	0,91	0,91	0,91	0,91
(Intervalo 95 %)	(0,88 – 0,95)	(0,88 – 0,94)	(0,88 – 0,94)	(0,88 – 0,94)
Punto de Corte	0,19			
AUC	0,92			
(Intervalo 95 %)	(0,91 – 0,93)			

Tabla 5-4.: Medidas de desempeño estimadas para el prototipo IV

5.5. Resultados Prototipo V

Según la Tabla 5-5 para el prototipo V, el cual está conformado por las cinco variables del hemograma, se encontró un punto de corte (PC) de 0,14, es decir, si a un individuo se le aplica este algoritmo, y su probabilidad predictiva arroja un valor mayor o igual al PC éste será clasificado como dengue positivo. También se identificó una capacidad predictiva del 73 % (IC 95 %: 72 % – 74 %), esto es, la probabilidad de que este algoritmo pueda distinguir

correctamente entre un individuo sano y uno enfermo es del 0,73.

Medida desempeño	Método			
	MV	Bootstrap	Bayes Inf	Bayes no Inf
Sensibilidad (Intervalo 95 %)	0,58 (0,39 – 0,78)	0,58 (0,38 – 0,79)	0,66 (0,55 – 0,76)	0,58 (0,39 – 0,75)
Especificidad (Intervalo 95 %)	0,82 (0,78 – 0,87)	0,82 (0,78 – 0,87)	0,77 (0,74 – 0,81)	0,82 (0,77 – 0,86)
Punto de Corte	0,14			
AUC (Intervalo 95 %)	0,73 (0,72 – 0,74)			

Tabla 5-5.: Medidas de desempeño estimadas para el prototipo V

Se encontró que para la estimación por máxima verosimilitud, bootstrap y con bayes no informativa, la sensibilidad fue de 0,58 y la especificidad fue de 0,82, sin embargo, al analizar las estimaciones obtenidas por la inferencia bayesiana informativa se puede ver que hubo un aumento en la sensibilidad (0,66), su intervalo de credibilidad fue más reducido y también se identifica una reducción en el valor de la especificidad (0,77) en comparación con las otras estimaciones, esto debido a que la distribución a priori informativa contribuyó considerablemente en las estimaciones.

5.6. Resultados Prototipo VI

El prototipo VI esta conformado por las cinco variables del hemograma, además, asume que la probabilidad $P(D = 1|B_j)$ es una variable aleatoria continua, por ende, al igual que los demás algoritmos que hacen este mismo supuesto, la estimación de la probabilidad de tener la enfermedad del dengue en cada una de las posibles combinaciones de las variables del hemograma, se realiza desde el enfoque bayesiano, el cual permite combinar la información previa con la información de la muestra.

De acuerdo con los resultado de la Tabla **5-6** un individuo es clasificado como dengue positivo con el prototipo VI si su probabilidad predictiva es mayor o igual a 0,15. Además, se identificó que su capacidad predictiva o AUC es del 83 %.

Las estimaciones puntuales encontradas por los tres enfoques obtuvieron valores iguales para la especificidad 0,8, mientras que para la sensibilidad fue más alta en el caso de la estimación bayesiana informativa, donde su región de credibilidad también fue la más angosta $RC(0,74 - 0,90)$. En comparación con los algoritmos anteriores, este prototipo VI ofrece más

balance entre sensibilidad y especificidad, además, aunque su capacidad clasificatoria (AUC) no ha sido la más alta, se considera aceptable.

Medida desempeño	Método			
	MV	Bootstrap	Bayes Inf	Bayes no Inf
Sensibilidad (Intervalo 95 %)	0,79 (0,63 – 0,95)	0,79 (0,63 – 0,96)	0,82 (0,74 – 0,90)	0,76 (0,60 – 0,90)
Especificidad (Intervalo 95 %)	0,80 (0,75 – 0,84)	0,80 (0,76 – 0,84)	0,80 (0,75 – 0,84)	0,80 (0,75 – 0,84)
Punto de Corte	0,15			
AUC (Intervalo 95 %)	0,83 (0,83 – 0,84)			

Tabla 5-6.: Medidas de desempeño estimadas para el prototipo VI

5.7. Comparación entre los prototipos clasificadores

Los valores predictivos permiten darse una idea acerca de la eficacia que tiene una prueba diagnóstica, pues se conocen como la probabilidad de tener (Valor predictivo positivo) o no la enfermedad (Valor predictivo negativo) una vez conocido el resultado de la prueba diagnóstica. Cabe resaltar que estas probabilidades dependen de la prevalencia de la enfermedad en la población.

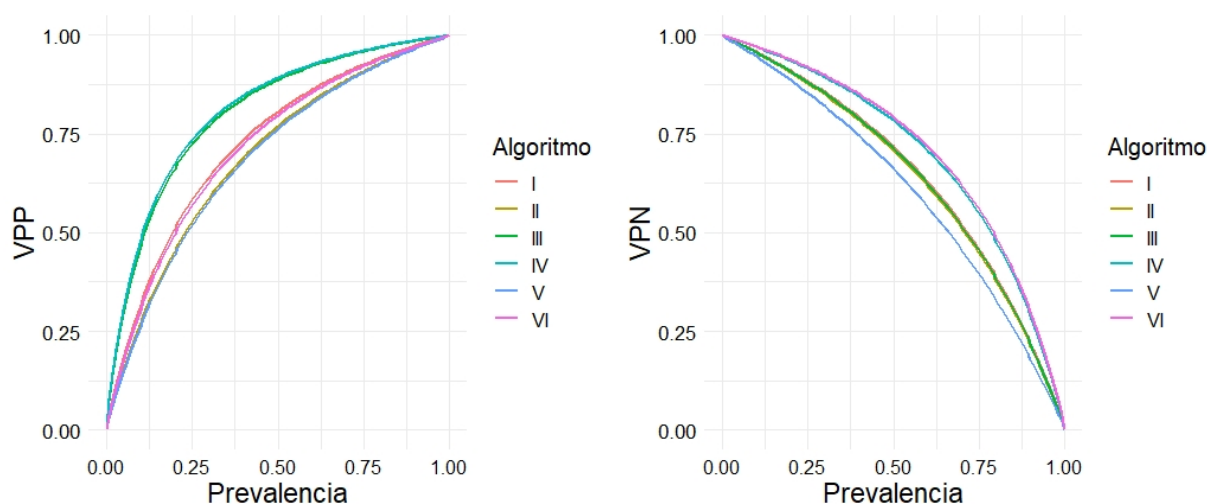


Figura 5-1.: Valores Predictivos

En la Figura 5-1 se presentan las curvas del valor predictivo positivo (VPP) y el valor predictivo negativo (VPN) para cada prototipo clasificador considerando diferentes valores de prevalencia de dengue. En gráfico del lado izquierdo, se puede observar que al aumentar la prevalencia, el VPP aumenta, es decir, que a mayor prevalencia de enfermedad en la población, es más probable que el paciente esté enfermo de dengue dado que tuvo un resultado positivo con el prototipo clasificador. Los prototipos que mostraron mejor desempeño en cuanto al VPP son el III y IV, los cuales están compuestos por los 12 signos y síntomas, y los de desempeño más bajo fueron los prototipos II y V. Cabe resaltar que en valores de la prevalencia cercanos a cero ó a uno, los VPP de los algoritmos I, II, V y VI tienen un comportamiento similar.

En cuanto al valor predictivo negativo, éste tiene un comportamiento inverso al VPP, pues al aumentar la prevalencia de la enfermedad, éste disminuye, es decir, la probabilidad de que un paciente no presente dengue (esté sano) dado que el prototipo dio negativo disminuye. Los prototipos que muestran mejor desempeño son: los prototipos IV y VI, y el que presenta el desempeño más bajo es el prototipo V. Por otro lado, se puede observar que los prototipos I, II y III presentaron casi la misma curva. En general, el prototipo que muestra un mejor desempeño según su VPP y VPN es el IV, aquel que considera los 12 signos y síntomas, y las variables del hemograma.

En la Figura 5-2 se muestra las curvas de valores predictivos positivo y negativo en un mismo gráfico para cada uno de los algoritmos, allí se puede ver el valor de prevalencia donde estas dos curvas se intersectan y se maximizan ambas probabilidades (VPP y VPN).

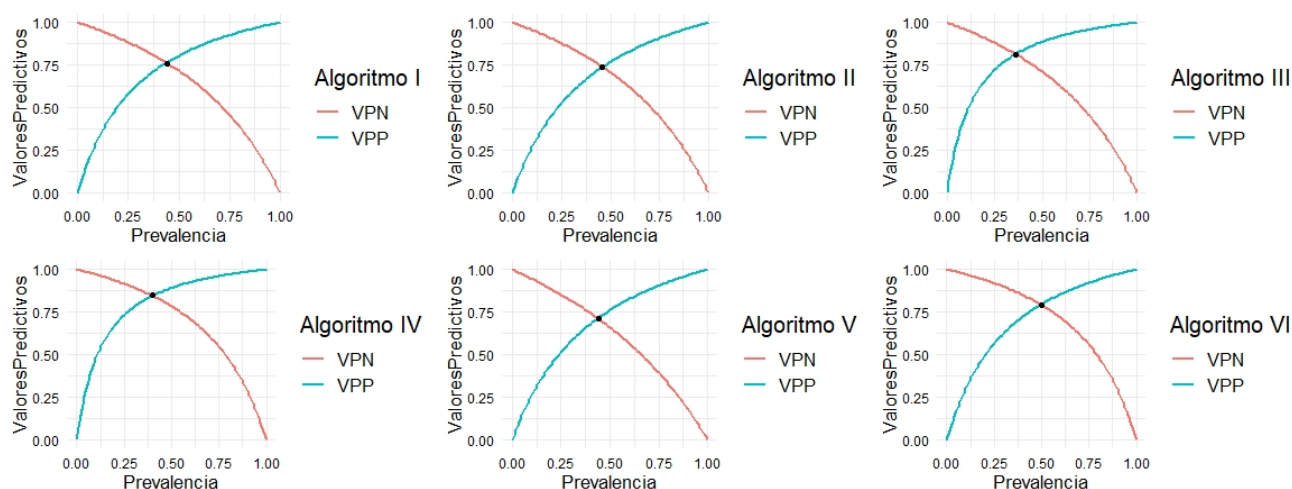


Figura 5-2.: Valores Predictivos para cada Algoritmo de Clasificación bajo diferentes prevelencias

En el algoritmo III se encontró el valor de prevalencia más bajo (0,36) donde se maximizan los valores predictivos (0,81), lo cual indica que en escenarios donde la probabilidad de que un individuo tenga dengue en la población sea baja este algoritmo de clasificación presenta una mejor calidad diagnostica a comparación de los demás algoritmos, mientras que, el valor de prevalencia más alto (0,50) que maximiza los valores predictivos fue encontrado con el algoritmo VI. Ver Tabla 5-7.

	Alg I	Alg II	Alg III	Alg IV	Alg V	Alg VI
Prevalencia	0,44	0,46	0,36	0,40	0,44	0,50
Valor Predictivo	0,76	0,74	0,81	0,85	0,71	0,79

Tabla 5-7.: Valores de prevalencia donde se maximizan los Valores Predictivos para cada Algoritmo

5.8. Estimación de parámetros de desempeño asumiendo independencia condicional entre las dos pruebas

Los test diagnósticos a considerar en este escenario serán: Los cuatro algoritmos propuestos en el trabajo de Méndez (2017) y los dos algoritmos asociados al hemograma propuestos en este trabajo (Prototipos V y VI), el objetivo es indagar acerca del comportamiento de los parámetros de desempeño de estos algoritmos cuando se aplican uno por uno de manera independiente, y ver más adelante si al realizar la aplicación en paralelo como resultado final se produce una mejora en cuanto a los parámetros de desempeño.

5.8.1. Estimación parámetros de desempeño

En la Tabla 5-8 se puede ver que no hay mayor diferencia entre las estimaciones obtenidas por los diferentes métodos, sin embargo hay que resaltar que las regiones de credibilidad de las estimaciones bayesianas son más angostas que los intervalos de confianza obtenidos por la estimación clásica y remuestreo. En cuanto a la sensibilidad de los algoritmos, es posible ver que aquel que tiene la sensibilidad más alta es el prototipo VI, el cual considera los cinco parámetros del hemograma y asume que la probabilidad $P(D = 1|B_j)$ es una variable continua, mientras que el prototipo III de Méndez, cuyos insumos son las ocho categorías de signos y síntomas presentó la sensibilidad más baja. En general, todos los prototipos presentan especificidades altas, y es de resaltar que los prototipos III y IV muestran tener especificidades casi de 1, es decir que estos algoritmos son capaces de clasificar correctamente a todos los individuos sanos de la muestra y podrían estar generando una alteración en la tasa de falsos negativos. El prototipo que mostró mayor balance entre sensibilidad y especificidad fue el prototipo VI.

	Estimación clásica		Estimación Bayesiana				Estimación Remuestreo	
	Sensibilidad (IC 95 %)	Especificidad (IC 95 %)	No Informativa		Informativa		Sensibilidad (IC 95 %)	Especificidad (IC 95 %)
			Sensibilidad (RC 95 %)	Especificidad (RC 95 %)	Sensibilidad (RC 95 %)	Especificidad (RC 95 %)		
Prototipo I	0,63	0,82	0,61	0,82	0,65	0,82	0,62	0,82
Méndez	(0,43 – 0,82)	(0,78 – 0,87)	(0,43 – 0,67)	(0,77 – 0,83)	(0,51 – 0,69)	(0,77 – 0,83)	(0,42 – 0,79)	(0,78 – 0,87)
Prototipo II	0,58	0,77	0,57	0,76	0,58	0,77	0,58	0,77
Méndez	(0,39 – 0,78)	(0,72 – 0,82)	(0,39 – 0,64)	(0,71 – 0,78)	(0,39 – 0,64)	(0,72 – 0,78)	(0,37 – 0,79)	(0,72 – 0,82)
Prototipo III	0,21	0,99	0,23	0,99	0,29	0,99	0,21	0,99
Méndez	(0,05 – 0,37)	(0,98 – 1,00)	(0,09 – 0,28)	(0,98 – 0,99)	(0,14 – 0,34)	(0,98 – 0,99)	(0,04 – 0,37)	(0,98 – 0,99)
Prototipo IV	0,46	0,99	0,46	0,99	0,46	0,99	0,46	0,99
Méndez	(0,26 – 0,66)	(0,98 – 1,00)	(0,27 – 0,52)	(0,98 – 0,99)	(0,27 – 0,53)	(0,97 – 0,99)	(0,26 – 0,66)	(0,99 – 1,00)
Prototipo V	0,58	0,82	0,57	0,82	0,66	0,77	0,58	0,82
	(0,39 – 0,78)	(0,78 – 0,87)	(0,38 – 0,64)	(0,77 – 0,83)	(0,54 – 0,70)	(0,73 – 0,78)	(0,38 – 0,79)	(0,78 – 0,87)
Prototipo VI	0,79	0,80	0,76	0,79	0,82	0,79	0,79	0,80
	(0,63 – 0,95)	(0,75 – 0,84)	(0,59 – 0,82)	(0,75 – 0,81)	(0,73 – 0,85)	(0,75 – 0,81)	(0,63 – 0,96)	(0,76 – 0,84)

Tabla 5-8.: Medidas de desempeño estimadas asumiendo independencia entre los resultados de las dos pruebas

5.9. Estimación de parámetros de desempeño asumiendo dependencia condicional entre las dos pruebas

En este escenario se van a considerar los mismos algoritmos del escenario anterior, bajo la salvedad de que los resultados no se pueden asumir independientes, pues existe un nivel de dependencia entre los resultados por el hecho de ser obtenidas de un mismo individuo. La estimación de los parámetros de desempeño, debido a la complejidad de este modelo, solo se realizarán bajo un enfoque bayesiano donde se consideran distribuciones a priori informativas para los parámetros.

5.9.1. Estimación Bayesiana

En la Tabla 5-9 se encuentran las estimaciones de la covarianza entre las pruebas diagnósticas tanto para la población de enfermos como de no enfermos, allí se puede ver que todos los valores estimados son cercanos a cero y por tanto, se podría inferir que la covarianza entre las pruebas es casi nula.

		Covarianza			
		ψ_D - Enfermos		ψ_{ND} - Sanos	
		Estimación Puntual	RC(95 %)	Estimación Puntual	RC(95 %)
Prototipo I Méndez	Prototipo V	0,042	(0,0002 – 0,064)	0,017	(0,0005 – 0,02)
	Prototipo VI	0,035	(0,0002 – 0,054)	0,008	(0,00003 – 0,012)
Prototipo II Méndez	Prototipo V	0,019	(0,000037 – 0,029)	0,024	(0,0018 – 0,032)
	Prototipo VI	0,016	(0,00003 – 0,024)	0,009	(0,00003 – 0,015)
Prototipo III Méndez	Prototipo V	0,023	(0,00001 – 0,036)	0,004	(0,00001 – 0,006)
	Prototipo VI	0,032	(0,0007 – 0,046)	0,003	(0,000008 – 0,004)
Prototipo IV Méndez	Prototipo V	0,028	(0,00002 – 0,04)	0,002	(0,000003 – 0,003)
	Prototipo VI	0,019	(0,00004 – 0,029)	0,002	(0,000003 – 0,003)

Tabla 5-9.: Estimación de los parámetros de dependencia entre los algoritmos de clasificación (covarianzas)

A pesar de que se considera un parámetro de dependencia entre cada par de pruebas al realizar las estimaciones de los parámetros de desempeño los resultados no difieren en mayor medida a los obtenidos cuando se consideró independencia, esto se debe a que la covarianza estimada aporta muy poco pues sus valores son cercanos a cero. Debido a lo anterior, se resalta de la tabla **5-10** que la sensibilidad en la mayoría de los algoritmos obtuvo una capacidad clasificatoria inferior a 0,7 y sus regiones de credibilidad son amplias a comparación de la especificidad, también se observó que los prototipos III y IV presentaron un mejor rendimiento para la correcta clasificación de los individuos sanos (especificidad) alcanzando valores superiores al 0,9, y al igual que en el caso anterior, el prototipo VI alcanza una óptima clasificación tanto para individuos enfermos (0,82) como sanos (0,79).

	Estimación Bayesiana	
	Sensibilidad RC (95 %)	Especificidad RC (95 %)
Prototipo I Méndez	0,66 (0,52 – 0,71)	0,81 (0,77 – 0,83)
Prototipo II Méndez	0,58 (0,39 – 0,65)	0,76 (0,72 – 0,78)
Prototipo III Méndez	0,29 (0,15 – 0,34)	0,99 (0,98 – 1,00)
Prototipo IV Méndez	0,47 (0,28 – 0,53)	0,98 (0,97 – 0,99)
Prototipo V	0,66 (0,55 – 0,70)	0,77 (0,73 – 0,79)
Prototipo VI	0,82 (0,74 – 0,85)	0,79 (0,75 – 0,81)

Tabla 5-10.: Estimación bayesiana - Medidas de desempeño estimadas asumiendo dependencia entre los resultados de las dos pruebas

5.10. Estimación de los parámetros de desempeño después de obtener un único procedimiento clasificador

Para este escenario se propuso ejecutar una nueva clasificación teniendo en cuenta los eventos planteados en la sección 4.7, esto con el fin de indagar si hay una mejora en cuanto a la capacidad clasificatoria al aplicar los prototipos de Méndez (2017) y los algoritmos creados a partir de las variables del hemograma (Protipos V y VI) en paralelo como clasificador final.

Al aplicar las pruebas en paralelo, se obtuvo un total de ocho nuevos algoritmos, producto de combinar cada uno de los cuatro algoritmos de Méndez (2017), con cada algoritmo construido a partir de las variables del hemograma (en total 2).

5.10.1. Estimación de parámetros de desempeño

En la Tabla **5-11** se presentan los resultados de las estimaciones de los parámetros de desempeño junto con sus intervalos de confianza o regiones de credibilidad obtenidos por los diferentes métodos de estimación. Se resalta que las estimaciones por método bayesiano no varían mucho en cuanto a asumir distribuciones previas informativas o no informativas, sin embargo, las regiones de credibilidad de las estimaciones informativas son un poco más angostas en algunos casos.

De las sensibilidades estimadas se encontró que la mayoría están por encima del 0,75 lo que indica que las combinaciones de los algoritmos que generan esas sensibilidades altas presentan las mejores capacidades clasificatorias para detectar a los verdaderos enfermos, es decir, a los verdaderos casos de dengue, además, se identificó que las estimaciones de los intervalos de confianza varían en un rango amplio en general. Por otro lado, la especificidad en las combinaciones que implican a los prototipos I y II de Méndez (2017) corresponden a valores inferiores a 0,7 mientras que en los prototipos III y IV alcanzan valores superiores a 0,8, lo cual permite identificar que estos últimos brindan una mejor capacidad para detectar a los individuos sanos y sus intervalos de confianza resultaron ser angostos, lo que los hace más precisos.

También, se identifica que el prototipo II de Méndez combinado con el prototipo VI presenta valores de sensibilidad más altos, y los prototipos de III y IV Méndez junto con los prototipos creados a partir de los parámetros del hemograma (Prototipos V y VI) obtuvieron los mejores resultados para la especificidad (ceranos a 0,8), sin embargo, se considera que en términos de un balance clasificatorio el prototipo IV de Méndez y el prototipo VI presentan los mejores resultados, pues se encontró que tanto su sensibilidad como especificidad es mayor

o igual a 0,80. En resumen, se demostró que la aplicación de las pruebas en paralelo con el fin de obtener un único procedimiento clasificador mejora considerablemente el rendimiento para clasificar a los individuos con dengue.

		Estimación clásica		Estimación Bayesiana				Estimación Remuestreo	
				No informativa		Informativa			
		Sensibilidad (IC 95 %)	Especificidad (IC 95 %)	Sensibilidad (RC 95 %)	Especificidad (RC 95 %)	Sensibilidad (RC 95 %)	Especificidad (RC 95 %)	Sensibilidad (IC 95 %)	Especificidad (IC 95 %)
Prototipo I Méndez	Prototipo V	0,79 (0,63 – 0,95)	0,69 (0,64 – 0,75)	0,78 (0,59 – 0,91)	0,69 (0,64 – 0,74)	0,73 (0,60 – 0,85)	0,69 (0,64 – 0,74)	0,80 (0,62 – 0,95)	0,69 (0,64 – 0,75)
	Prototipo VI	0,87 (0,74 – 1,00)	0,67 (0,61 – 0,71)	0,84 (0,68 – 0,94)	0,67 (0,60 – 0,71)	0,78 (0,64 – 0,88)	0,66 (0,61 – 0,72)	0,87 (0,74 – 1,00)	0,67 (0,61 – 0,72)
Prototipo II Méndez	Prototipo V	0,83 (0,68 – 0,98)	0,66 (0,60 – 0,71)	0,81 (0,64 – 0,93)	0,66 (0,60 – 0,71)	0,80 (0,65 – 0,93)	0,66 (0,61 – 0,71)	0,83 (0,68 – 0,98)	0,66 (0,60 – 0,71)
	Prototipo VI	0,92 (0,80 – 1,00)	0,63 (0,57 – 0,68)	0,85 (0,73 – 0,98)	0,63 (0,57 – 0,68)	0,87 (0,74 – 0,97)	0,63 (0,58 – 0,69)	0,92 (0,81 – 1,00)	0,63 (0,57 – 0,68)
Prototipo III Méndez	Prototipo V	0,67 (0,48 – 0,86)	0,82 (0,78 – 0,86)	0,66 (0,48 – 0,82)	0,82 (0,76 – 0,86)	0,66 (0,49 – 0,82)	0,82 (0,78 – 0,86)	0,66 (0,48 – 0,85)	0,8 (0,78 – 0,87)
	Prototipo VI	0,79 (0,63 – 0,95)	0,80 (0,75 – 0,85)	0,77 (0,59 – 0,90)	0,80 (0,75 – 0,84)	0,77 (0,60 – 0,89)	0,80 (0,75 – 0,84)	0,79 (0,63 – 0,95)	0,79 (0,75 – 0,84)
Prototipo IV Méndez	Prototipo V	0,71 (0,53 – 0,89)	0,82 (0,78 – 0,87)	0,67 (0,50 – 0,84)	0,82 (0,78 – 0,86)	0,70 (0,53 – 0,85)	0,83 (0,78 – 0,87)	0,71 (0,53 – 0,89)	0,82 (0,78 – 0,87)
	Prototipo VI	0,83 (0,68 – 0,98)	0,80 (0,75 – 0,85)	0,81 (0,63 – 0,93)	0,80 (0,75 – 0,84)	0,81 (0,65 – 0,93)	0,81 (0,76 – 0,85)	0,84 (0,68 – 0,98)	0,80 (0,76 – 0,85)

Tabla 5-11.: Medidas de desempeño estimadas asumiendo un único proceso clasificador

6. Conclusiones y recomendaciones

En este capítulo se presentan las conclusiones y recomendaciones más relevantes a las que se llegaron durante la realización de este trabajo de grado.

El algoritmo de simulación planteado en la sección 4.5.2 permitió construir archivos de datos con un volumen considerado, lo cual resultó ser una herramienta importante a la hora de evaluar el comportamiento de los algoritmos a través de la convergencia de sus parámetros de desempeño. Cabe resaltar que esta metodología de simulación puede ser aplicada en cualquier caso donde se requiera ampliar un archivo de datos y se pueda asignar una distribución a la variable de interés, en este caso la probabilidad de tener dengue dado la presencia de alguna categoría de signos y síntomas.

Al realizar la estimación de los parámetros de desempeño por los diferentes métodos: máxima verosimilitud, estimación bayesiana y remuestreo bootstrap, se esperaba que éstas difirieran entre sí, pues con métodos como el remuestreo y estimación bayesiana se esperaba mitigar el problema de la muestra pequeña, sin embargo se pudo observar que las estimaciones fueron consistentes por los tres métodos utilizados, lo cual conlleva a inferir que el tamaño de muestra no está afectando las estimaciones, por tanto, con el fin de conseguir estimaciones un poco más precisas de los parámetros de desempeño se recomienda para la estimación bayesiana utilizar distribuciones a priori informativas que hayan sido elicítadas directamente sobre expertos en el diagnóstico de dengue.

Al comparar el área bajo la curva (AUC) de los algoritmos propuestos en este trabajo con los algoritmos contruidos por Méndez (2017) se evidenció un aumento en la capacidad clasificatoria, además, la comparación de los parámetros de desempeño permitió identificar que la sensibilidad no presenta cambios significativos, mientras que la especificidad para todos los algoritmos aumentó considerablemente. Por lo anterior, es posible concluir que la adición de los cinco parámetros del hemograma a los algoritmos de Méndez (2017) mejoran la probabilidad de detectar a un individuo como sano cuando éste realmente está sano. Se recomienda proponer un estudio mediante el cual se pueda identificar cual o cuales de las variables del hemograma consideradas en este trabajo brindan a los algoritmos una mejor capacidad de discriminación para la clasificación del dengue, es decir, encontrar cual de las variables adicionadas maximiza los parámetros de desempeño.

Al analizar los valores predictivos para cada algoritmo bajo distintos valores de prevalencia, fue posible identificar que el algoritmo construido a partir de las ocho categorías y las variables del hemograma (Prototipo III), y el algoritmo que considera los 12 signos y síntomas junto con las cinco variables del hemograma (Prototipo IV), mostraron mejor desempeño en cuanto a su VPP, pues ante los distintos valores de prevalencia estos algoritmos identificaron en mayor proporción a aquellos pacientes que poseen la enfermedad del dengue. Al estimar los parámetros de desempeño de estos algoritmos se encontró que el algoritmo III alcanzó una sensibilidad de 0,63 y una especificidad de 0,92, mientras que el algoritmo IV tuvo una sensibilidad más alta (0,75) y una especificidad de 0,91. Con respecto al VPN se encontró que los algoritmos de mejor desempeño fueron el IV y el VI, donde se encontró que para el algoritmo VI el cual considera los cinco parámetros del hemograma se estimó una sensibilidad de 0,82 y una especificidad de 0,80. En general, el prototipo que muestra un mejor desempeño según su VPP y VPN es el IV.

En cuanto a la eficiencia del test diagnóstico se identificó para cada algoritmo bajo que valor de prevalencia funciona mejor, de esta manera, se concluyó que en zonas donde la prevalencia está entre 0,3 – 0,4 resulta oportuno evaluar a cada paciente ingresado a consulta con el algoritmo III o IV, si la prevalencia es mayor a 0,4 se puede evaluar al paciente con el algoritmos I, II ó V, mientras que zonas donde la prevalencia alcanza valores cercanos a 0,5 se considera prudente usar el algoritmo VI. Cabe resaltar que para la elección de los algoritmos en cada nivel de prevalencia se debe tener en cuenta la información con la que se cuenta en el centro hospitalario, sea solamente información de signos y síntomas ó adicionalmente información del hemograma.

Para las estimaciones de los parámetros de desempeño bajo los supuestos de independencia y dependencia condicional se esperaba encontrar mayores diferencias, sin embargo, este comportamiento no se evidenció debido a que las covarianzas son cercanas a cero y por ende no se encontraron estimaciones menos sesgadas bajo el supuesto de dependencia condicional, por este motivo autores como Tovar & Achcar (2012) recomiendan el uso de modelos probabilísticos que caractericen estructuras de dependencia débiles como es el caso de las funciones cópula. Esos mismos autores reportaron que las dependencias entre las pruebas para diagnóstico aunque tienden a ser débiles y no lineales tienen un efecto sobre las estimaciones de los parámetros de desempeño que no logra ser capturado con los modelos que asumen estructuras de dependencia lineales, tipo covarianza. En cuanto a la sensibilidad de los algoritmos, es posible ver que aquel que tiene la sensibilidad más alta es el prototipo VI (0,79) y el que tiene la sensibilidad más baja es el prototipo III (0,21) de Méndez. Por otro lado, se resalta que los prototipos III y IV muestran tener especificidades casi de uno lo cual podría presentar un efecto sobre la tasa de falsos positivos una alteración en la tasa de falsos negativos. Por último, el prototipo VI mostró mayor balance entre sensibilidad (0,79) y especificidad (0,80).

Al obtener un único procedimiento clasificador producto de la aplicación de las pruebas en paralelo, se pudo identificar que hubo una buena clasificación de los pacientes que están verdaderamente enfermos, pues en la mayoría de los nuevos clasificadores la sensibilidad supero la especificidad. En general, se encontró que la combinación de los algoritmos de Méndez con el algoritmo VI, el cual incluye únicamente los parámetros del hemograma y cuyas probabilidades fueron calculadas asumiendo la forma de la fórmula de Bayes para variables continuas, presentaron los mejores resultados para los parámetros de desempeño, ya que tanto la sensibilidad y especificidad alcanzaron estimaciones superiores a 0,65, siendo la combinación del prototipo IV de Méndez y el prototipo VI el de mejor balance clasificatorio (Sensibilidad = 0,83 y Especificidad = 0,80).

La metodología propuesta en el presente trabajo de grado no esta limitada únicamente a la construcción de algoritmos para diagnosticar pacientes con dengue, por el contrario es susceptible de ser implementada a cualquier tipo de enfermedad para la cual, el médico tratante cuente con información de los signos y síntomas del paciente en forma binaria y al menos una prueba de laboratorio cuyo resultado puede ser cuantitativo o binario.

A. Anexos:

1. Valores a Elicitar:

8 Categoría - Eventos		8 Categoría - Variable Aleatoria		12 Signos y Sintomas - Eventos		12 Signos y Sintomas - Variable Aleatoria		5 Parametros del Hemograma - Eventos		5 Parametros del Hemograma - Variable Aleatoria	
Sen	Espe	Sen	Espe	Sen	Espe	Sen	Espe	Sen	Espe	Sen	Espe
0.666	0.59	0.722	0.603	0.672	0.677	0.757	0.816	0.642	0.656	0.858	0.52
0.665	0.592	0.72	0.646	0.663	0.694	0.74	0.84	0.76	0.677	0.884	0.496
0.642	0.656	0.858	0.52	0.76	0.724	0.813	0.882	0.635	0.724	0.813	0.882
0.635	0.677	0.738	0.884	0.742	0.742	0.779	0.894	0.742	0.723	0.779	0.894
0.66	0.73	0.72	0.72	0.69	0.84	0.78	0.86				
0.63	0.79	0.69	0.77	0.51	0.87	0.6	0.89				
0.64	0.58	0.17	0.9	0.07	0.93	0.02	0.98				
0.57	0.68	0.56	0.68	0.1	0.9	0.01	0.98				
0.7	0.72	0.22	0.99	0.44	0.95	0.41	0.99				
0.59	0.81	0.64	0.82	0.48	0.91	0.42	0.99				
0.87	0.24	0.54	0.65	0.76	0.724	0.757	0.816				
0.63	0.63	0.63	0.61	0.745	0.695	0.813	0.882				
0.9	0.38	0.66	0.9			0.809	0.86				
0.65	0.38	0.74	0.87								
		0.722	0.603								
		0.858	0.52								
		0.719	0.695								

Tabla A-1.: Valores para elicitar

2. Puntos de Corte Méndez (2017)

Prototipos Méndez	Punto de Corte δ
I	0,096
II	0,101
III	0,650
IV	0,54

Tabla A-2.: Puntos de corte encontrados por Méndez (2017) para cada prototipo

3. Código R

En el siguiente enlace podrá consultar el código de R usado para el desarrollo de este Trabajo de Grado: <https://drive.google.com/drive/folders/1cNow-P48ER9KwP4v4CwqJMcy6s0rxqPr?usp=sharing>

Bibliografía

- AACC (2020), 'Complete blood count (cbc)', url<https://labtestsonline.org/>. Accedido 13-11-2020.
- Altman, D. G. & Bland, J. M. (1994), 'Diagnostic tests. 1: Sensitivity and specificity.', *BMJ: British Medical Journal* **308**(6943), 1552.
- Bernardo, J. M. & Smith, A. F. (2009), *Bayesian theory*, Vol. 405, John Wiley & Sons.
- Bernatsky, S., Lix, L., Hanly, J., Hudson, M., Badley, E., Peschken, C., Pineau, C., Clarke, A., Fortin, P., Smith, M. et al. (2011), 'Surveillance of systemic autoimmune rheumatic diseases using administrative data', *Rheumatology international* **31**(4), 549–554.
- Böhning, D. & Patilea, V. (2008), 'A capture–recapture approach for screening using two diagnostic tests with availability of disease status for the test positives only', *Journal of the American Statistical Association* **103**(481), 212–221.
- Box, G. E. & Tiao, G. C. (2011), *Bayesian inference in statistical analysis*, Vol. 40, John Wiley & Sons.
- Bravo-Grau, S. & Cruz, J. P. (2015), 'Estudios de exactitud diagnóstica: Herramientas para su interpretación', *Revista chilena de radiología* **21**(4), 158–164.
- Caicedo-Borrero, D. M., Tovar, R., Méndez, A., Parra, B., Bonelo, A., Celis, J., Villegas, L., Collazos, C. & Osorio, L. (2020), 'Development and performance of dengue diagnostic clinical algorithms in colombia', *The American Journal of Tropical Medicine and Hygiene* p. tpmd190722.
- Caicedo, D. M., Méndez, A. C., Tovar, J. R. & Osorio, L. (2019), 'Desarrollo de algoritmos clínicos para el diagnóstico del dengue en colombia', *Biomédica* **39**(1), 170–185.
- Calvo, E. P., Coronel-Ruiz, C., Velazco, S., Velandia-Romero, M. & Castellanos, J. E. (2016), 'Diagnóstico diferencial de dengue y chikungunya en pacientes pediátricos', *Biomédica* **36**(2), 35–43.
- Castellanos, J. E. & Coronel-Ruiz, C. (2014), 'Dengue disease diagnosis: A puzzle to be solved', *Revista de la Facultad de Medicina* **62**(4), 617–629.

- Cerda, J. & Cifuentes, L. (2012), 'Using roc curves in clinical investigation: theoretical and practical issues', *Revista chilena de infectología: organo oficial de la Sociedad Chilena de Infectología* **29**(2), 138–141.
- Dawid, A. P. (1979), 'Conditional independence in statistical theory', *Journal of the Royal Statistical Society: Series B (Methodological)* **41**(1), 1–15.
- de Ullibarri Galparsoro, L. & Fernández, P. (1998), 'Curvas roc', *Atención Primaria en la Red* **5**(4), 229–35.
- Dendukuri, N. & Joseph, L. (2001), 'Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests', *Biometrics* **57**(1), 158–167.
- Díaz Novás, J., Gallego Machado, B. & León González, A. (2006), 'El diagnóstico médico: bases y procedimientos', *Revista Cubana de Medicina General Integral* **22**(1), 0–0.
- Donis, J. H. (2012), 'Evaluación de la validez y confiabilidad de una prueba diagnóstica', *Avances en biomedicina* **1**(2), 73–81.
- Escrig-Sos, J., Martínez-Ramos, D. & Miralles-Tena, J. M. (2006), 'Pruebas diagnósticas: nociones básicas para su correcta interpretación y uso', *Cirugía Española* **79**(5), 267–273.
- Fajardo Gutiérrez, A. (2017), 'Medición en epidemiología: prevalencia, incidencia, riesgo, medidas de impacto', *Revista Alergia México* **64**(1), 109–120.
- Fox, J. (2002), 'Bootstrapping regression models appendix to an r and s-plus companion to applied regression'.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013), *Bayesian data analysis*, CRC press.
- Georgiadis, M. P., Johnson, W. O., Gardner, I. A. & Singh, R. (2003), 'Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **52**(1), 63–76.
- Gonçalves, L., Subtil, A., Oliveira, M. R. & Bermudez, P. (2014), 'Roc curve estimation: An overview', *REVSTAT-Statistical Journal* **12**(1), 1–20.
- Guzmán, M. G., García, G. & Kourí, G. (2006), 'El dengue y el dengue hemorrágico: prioridades de investigación', *Revista panamericana de salud pública* **19**, 204–215.
- Hajian-Tilaki, K. (2013), 'Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation', *Caspian journal of internal medicine* **4**(2), 627.
- Hanley, J. A. & McNeil, B. J. (1982), 'The meaning and use of the area under a receiver operating characteristic (roc) curve.', *Radiology* **143**(1), 29–36.

- Huerta Aragonés, J., Cela de Julián, E. & AEPap (2018), 'Hematología práctica: interpretación del hemograma y de las pruebas de coagulación', *AEPap, editor. Curso de Actualización Pediatría* pp. 507–6.
- Instituto Nacional de Salud (2019), 'Boletín epidemiológico semanal. semana epidemiológica no. 16 de 2019'.
URL: <https://www.ins.gov.co/buscador-eventos/BoletinEpidemiologico/2019%20Bolet%20C3%20ADn%20epidemiol%C3%B3gico%20semana%2016.pdf>
- Jones, G., Johnson, W. O., Hanson, T. E. & Christensen, R. (2010), 'Identifiability of models for multiple diagnostic testing in the absence of a gold standard', *Biometrics* **66**(3), 855–863.
- Knottnerus, J. A., Van Weel, C. & Muris, J. (2002), *The evidence base of clinical diagnosis*, BMJ books London.
- Lynch, J. G. (1982), 'On the external validity of experiments in consumer research', *Journal of consumer Research* **9**(3), 225–239.
- Martínez Camblor, P. (2007), 'Comparación de pruebas diagnósticas desde la curva roc', *Revista Colombiana de Estadística* **30**(2), 163–176.
- Méndez, A. C. (2017), Propuesta metodológica para construir reglas de clasificación: caso de aplicación dengue. Tesis de pregrado, Universidad del Valle.
- Minsalud (2021), Minsalud invita a no bajar la guardia contra el dengue, Boletín de Prensa 821, Ministerio de Salud de Colombia.
- Mogollón, G., Sefair, C., Upegui, D. & Tovar, J. R. (2014), 'Colangiopancreatografía magnética: valor diagnóstico para detectar coledocolitiasis en pacientes con pancreatitis aguda leve', *Revista Cubana de Cirugía* **53**(1), 41–51.
- Padilla, J. C., Rojas, D. P. & Sáenz, R. (2012), 'Dengue en colombia: epidemiología de la reemergencia a la hiperendemia'.
- Pan-ngum, W., Blacksell, S. D., Lubell, Y., Pukrittayakamee, S., Bailey, M. S., de Silva, H. J., Lalloo, D. G., Day, N. P., White, L. J. & Limmathurotsakul, D. (2013), 'Estimating the true accuracy of diagnostic tests for dengue infection using bayesian latent class models', *PloS one* **8**(1).
- Pepe, M. S. et al. (2003), *The statistical evaluation of medical tests for classification and prediction*, Medicine.
- Quiñones, K. J. & Suarez, L. F. (2019), Validación externa de prototipos clasificadores en un caso aplicado: Dengue. Tesis de pregrado, Universidad del Valle.

- Salech, F., Mery, V., Larrondo, F. & Rada, G. (2008), 'Estudios que evalúan un test diagnóstico: interpretando sus resultados', *Revista médica de Chile* **136**(9), 1208–1208.
- Solanas, A. & Olivera, V. (1992), 'Bootstrap: fundamentos e introducción a sus aplicaciones', *Anuario de psicología/The UB Journal of psychology* (55), 143–154.
- Swets, J. A. (1979), 'Roc analysis applied to the evaluation of medical imaging techniques.', *Investigative radiology* **14**(2), 109–121.
- Tovar Cuevas, J. R. (2015), 'Inferencia bayesiana e investigación en salud: un caso de aplicación en diagnóstico clínico', *Revista Médica de Risaralda* **21**(1).
- Tovar, J. R. & Achcar, J. A. (2011), 'Indexes to measure dependence between clinical diagnostic tests: A comparative study', *Revista Colombiana de Estadística* **34**(3), 433–450.
- Tovar, J. R. & Achcar, J. A. (2012), 'Two dependent diagnostic tests: Use of copula functions in the estimation of the prevalence and performance test parameters', *Revista Colombiana de Estadística* **35**(3), 331–347.
- Vacek, P. M. (1985), 'The effect of conditional dependence on the evaluation of diagnostic tests', *Biometrics* pp. 959–968.
- Velandia, M. & Castellanos, J. E. (2011), 'Virus del dengue: estructura y ciclo viral', *Infectio* **15**(1).
- Villamil Gómez, W. E., González Camargo, O., Rodriguez Ayubi, J., Zapata Serpa, D. & Rodriguez Morales, A. J. (2016), 'Dengue, chikungunya and zika co-infection in a patient from colombia', *Journal of infection and public health* **9**(5), 684–686.
- World Health Organization (2020), 'Dengue and severe dengue'.
URL: <https://www.who.int/es/news-room/fact-sheets/detail/dengue-and-severe-dengue>
- World Health Organization, Special Programme for Research, Training in Tropical Diseases, World Health Organization. Department of Control of Neglected Tropical Diseases & World Health Organization. Epidemic and Pandemic Alert (2009), *Dengue: guidelines for diagnosis, treatment, prevention and control*, World Health Organization.
- Zhang, B., Salieb-Beugelaar, G. B., Nigo, M. M., Weidmann, M. & Hunziker, P. (2015), 'Diagnosing dengue virus infection: rapid tests and the role of micro/nanotechnologies', *Nanomedicine: Nanotechnology, Biology and Medicine* **11**(7), 1745–1761.
- Zhang, H., Li, W., Wang, J., Peng, H., Che, X., Chen, X. & Zhou, Y. (2014), '2 ns1-based tests supplying a diagnostic utility for confirming dengue 3 infection: a meta-analysis q1', *International Journal of Infectious Diseases* **30**, e1–e10.